ASME
SETTING THE STANDARD
1880 2005

Editor
ROBERT M. McMEEKING
Assistant to the Editor
LIZ MONTANA

APPLIED MECHANICS DIVISION
Executive Committee
(Chair) W.-K. LIU
T. N. FARRIS
K. RAVI-CHANDAR
D. J. INMAN
Z. SUO
Associate Editors
E. ARRUDA (2007)
H. ESPINOSA (2007)
H. GAO (2006)
S. GOVINDJEE (2006)
D. A. KOURIS (2005)
K. M. LIECHTI (2006)
A. M. MANIATTY (2007)
I. MEZIC (2006)
M. P. MIGNOLET (2006)
S. MUKHERJEE (2006)
O. O'REILLY (2007)
K. RAVI-CHANDAR (2006)
N. SRI NAMACHCHIVAYA (2006)
Z. SUO (2006)
T. E. TEZDUYAR (2006)
N. TRIANTAFYLLIDIS (2006)
B. A. YOUNIS (2006)

PUBLICATIONS DIRECTORATE
Chair, ARTHUR G. ERDMAN

OFFICERS OF THE ASME
President, RICHARD E. FEIGEL
Executive Director, V. R. CARTER
Treasurer, T. PESTORIUS

PUBLISHING STAFF
Managing Director, Publishing
PHILIP DI VIETRO
Production Coordinator
JUDITH SIERANT
Production Assistant
MARISOL ANDINO

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. **STATEMENT from By-Laws.** The Society shall not be responsible for statements or opinions advanced in papers or . . . printed in its publications (B7.1, Para. 3).
COPYRIGHT © 2005 by The American Society of Mechanical Engineers. For authorization to photocopy material for internal or personal use under those circumstances not falling within the fair use provisions of the Copyright Act, contact the Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, tel: 978-750-8400, www.copyright.com. Request for special permission or bulk copying should be addressed to Reprints/Permission Department. Canadian Goods & Services Tax Registration #126148048.

Journal of Applied Mechanics

Published Bimonthly by ASME

VOLUME 72 • NUMBER 5 • SEPTEMBER 2005

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.

Journal of Applied Mechanics **Volume 72, Number 5** **SEPTEMBER 2005**

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

Nano-Scale Effects in the Sliding and Rolling of a Cylinder on a Substrate

Ö. T. Sari

G. G. Adams
e-mail: adams@coe.neu.edu
Fellow ASME

S. Müftü
Mem. ASME

Mechanical and Industrial Engineering Department,
Northeastern University,
Boston, MA 02115

The behavior of a nano-scale cylindrical body (e.g., a fiber), lying on a substrate and acted upon by a combination of normal and tangential forces, is the subject of this investigation. As the scale decreases to the nano level, adhesion becomes an important issue in this contact problem. Thus, this investigation treats the two-dimensional plane strain elastic deformation of both the cylinder and the substrate during a rolling/sliding motion, including the effect of adhesion using the Maugis model. For the initiation of sliding, the Mindlin approach is used, whereas for rolling, the Carter approach is utilized. Each case is modified for nano-scale effects by including the effect of adhesion on the contact area and by using the adhesion theory of friction for the friction stress. Analytical results are given for the normal and tangential loading problems, including the initiation of sliding and rolling in terms of dimensionless quantities representing adhesion, cylinder size, and applied forces. [DOI: 10.1115/1.1831291]

Introduction

At the macro-scale, adhesion between contacting bodies has a negligible effect on surface interactions. However, as the contact size decreases, adhesion between the bodies becomes significant, especially for clean surfaces and lightly loaded systems. As technology advances, the scale of some devices decrease and adhesion becomes an important issue, especially in such applications as information storage devices and microelectromechanical systems.

In contacting bodies, atoms of materials are separated by an equilibrium distance z_0 of a few angstroms. At that separation, van der Waals forces are dominant over electrostatic forces [1]. The derivative of the Lennard-Jones potential expresses the surface pressure in terms of the separation z between two parallel surfaces

$$p_a(z)=\frac{8w}{3z_0}\left[\left(\frac{z_0}{z}\right)^9-\left(\frac{z_0}{z}\right)^3\right] \tag{1}$$

where w is the work of adhesion $w=\gamma_1+\gamma_2-\gamma_{12}$. Here, the surface energies of the contacting bodies are γ_1 and γ_2, and the interface energy of the two surfaces is γ_{12}. The Lennard-Jones potential includes attractive forces, which act over a long range, and repulsive forces, which act over a short range. If the separation is less than the equilibrium spacing, the repulsive (i.e., contact) forces are dominant, whereas if the separation exceeds the equilibrium spacing, there will be a net attractive force (i.e., the adhesion force). The Lennard-Jones pressure vanishes at the equilibrium spacing.

Numerous studies have been conducted on the adherence of spherical bodies. Bradley used the attractive force between two molecules and, by integrating through the whole bodies, found the cohesive force between two rigid spheres of radii R_1 and R_2 [2]. The pull-off force F required to separate two rigid spherical bodies was found to be

$$F=2\pi wR \tag{2}$$

where $R=R_1R_2/(R_1+R_2)$ is the equivalent radius of curvature.

Johnson, Kendall, and Roberts (JKR) presented a theory on the adherence of deformable elastic bodies [3]. In the JKR approximation, the adhesion outside the contact region is assumed to be zero. The balance among the elastic, surface, and potential energies was used, and the resulting stresses at the edges of the contact were infinite. The contact area is larger than the contact area predicted by Hertz, and the pull-off force was found to be

$$F=(3/2)\pi wR \tag{3}$$

A few years later, Derjaguin, Muller, and Toporov (DMT) presented another theory on the adhesion of deformable elastic bodies [4]. In the DMT theory, the adhesion force is taken into account outside the contact area, but the form of the stress distribution in the contact is assumed to be unaffected. The pull-off force was found to be the same as the Bradley relation. It is seen that for both the JKR and DMT theories, the pull-off force is independent of the elastic properties of the materials.

Although the JKR and DMT theories seem to be competitive, it was shown by Tabor [5] that these two theories represent the extremes of a parameter μ, which is the ratio of elastic deformation to the range of adhesive forces, i.e.,

$$\mu=\left(\frac{Rw^2}{E^2z_0^3}\right)^{1/3} \tag{4}$$

where E is the composite Young's modulus. Thus, bodies in which the elastic deformation is large compared to the range of surface forces are in the JKR regime, whereas the DMT regime corresponds to elastic deformations, which are much less than the range of surface forces. Greenwood constructed an adhesion map that covers these regimes [6].

Maugis presented a model for the transition between the JKR and DMT theories [7]. Similar to the Dugdale model of a crack, the Maugis model assumes a constant tensile surface stress σ_0 in regions where the surfaces are separated by a distance less than h, where h is obtained from the work of adhesion through the relation $w=\sigma_0 h$. When this work of adhesion is set equal to that for the Lennard-Jones potential, the maximum separation distance for adhesion h (measured from the equilibrium position) is found to be $0.97z_0$, where σ_0 is taken as the theoretical strength. Baney and Hui [8] used the Maugis model to investigate the two-dimensional problem of the adhesion of two circular cylinders.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, September 3, 2003; final revision, February 5, 2004. Associate Editor: Z. Suo. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

Results are given for the contact and adhesion regions as functions of the applied normal force and also for the pull-off force.

In this paper, the adhesion of contacting cylinders, or equivalently a cylinder in contact with a half-space, at the nano-scale is considered. If the cylinder is subjected to a combined tangential and normal loading, it may remain at rest, roll, slide, or undergo a complex motion depending on the magnitudes and the application points of the loading. The elastic behavior of the cylindrical body and half-space with adhesion are investigated using the plane strain theory of elasticity. A similar problem was treated in the thesis by Sari [9].

This nano-scale sliding and rolling analysis differs from the corresponding macro-scale problem in two important ways. First, due to the small scale of the contact area, adhesion becomes important. The Maugis model is used to approximate the adhesive stress outside the contact region in a manner similar to Baney and Hui. Second, in the macro scale, Coulomb friction, which states that frictional force is proportional to the normal load, is considered valid. Due to surface roughness in the contact of two nominally flat surfaces, the bodies touch at a discrete number of contacts, so that the real contact area becomes much smaller than the apparent contact area. If a statistical distribution for the asperity summits is assumed, it turns out that the real area of contact is approximately proportional to the normal force. This result is consistent with the adhesion theory of friction. According to the adhesion theory of friction, contacting asperities form strong adhesive junctions. In order to have a relative motion between the contacting bodies, these junctions need to be broken, which occurs when the interfacial shear stress reaches the shear strength of the weaker material. Therefore, a constant shear stress occurs in the real area of contact during the relative motion of contacting bodies. Thus, a constant frictional shear stress along with a near proportionality between normal load and real contact area, gives a friction force approximately proportional to the normal load, i.e., Coulomb friction. At the nano scale, however, in which the contact radius is in the order of 10 nm or less, it seems more reasonable to assume a single real contact area; i.e., constant shear stress during sliding.

The atomic force microscopy (Carpick et al. [10]) and surface force apparatus (SFA) (Homola et al. [11]) experiments as well as the modeling work of Hurtado and Kim [12,13] show that the friction stress is scale dependent. The friction stress has a constant high value ($\approx G^*/43$) in the nanoregion (contact radii less than about 15 nm) and a constant but lower value ($\approx G^*/1286$) in the micro region (contact radii greater than about 40 μm), where $G^*=2G_1G_2/(G_1+G_2)$ and G_1 and G_2 are the shear moduli of each contacting body. A transition region connects these two regions. Because the focus of our investigation is the nano region, the friction stress is assumed to be constant rather than scale dependent. The work of adhesion (w) for solids typically varies from about 50 mJ/m^2 to as high as a few J/m^2 for ultra-clean metal-to-metal contacts (e.g., Israelachvili [14]). The actual adhesion energy is affected by contaminants, but for the case of clean copper-on-copper contact $w\approx 4$ J/m^2 and its elastic modulus is $E_{\mathrm{Cu}}\approx 100$ GPa, leading to $\sigma_0/E\approx 0.2$, where E is the composite modulus defined later.

Adhesion of cylindrical bodies on a substrate is encountered in nano-wires, carbon nano-tubes and nano-fibers, and in different fields such as microbiology, microelectronics, and micro- and nanoelectromechanical systems devices. Determination of the forces necessary to roll or slide a cylindrical body on the substrate are important quantities to know in these applications. In some cases it is important to manipulate these single fibers to form a structure, whereas in other instances the sliding and rolling motions are important in contamination removal processes.

Theory and Discussion of the Results

The contact of a cylindrical body of radius R with a flat surface is investigated. The results are equally valid for the contact of two cylinders by using the equivalent radius of curvature previously defined. Linear plane strain elasticity is used throughout, which implies that the forces are given per unit length.

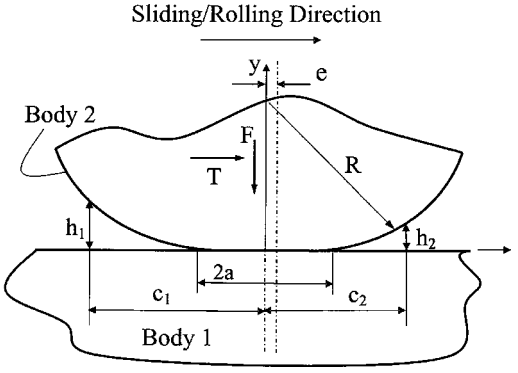


Fig. 1 Contact of a cylinder with a half-space under normal and tangential loading

According to plane strain linear elasticity (e.g., Barber [15]), the derivative of the surface normal displacements can be written in terms of normal and shear stresses as

$$\frac{d(u_y^{(1)}-u_y^{(2)})}{dx}=\frac{A}{4\pi}\int_{-c}^{c}\frac{p_y(\xi)d\xi}{x-\xi}-\frac{B}{4}p_x(x) \tag{5}$$

Similarly, the relation between the derivative of the relative displacements of the bodies in the tangential direction to the boundary stresses is given by

$$\frac{d(u_x^{(1)}-u_x^{(2)})}{dx}=-\frac{A}{4\pi}\int_{-a}^{a}\frac{p_x(\xi)d\xi}{x-\xi}-\frac{B}{4}p_y(x) \tag{6}$$

In Eqs. (5) and (6) p_x is the tangential traction in the x-direction and the contact pressure p_y is considered positive in compression. The material parameters A and B are given by

$$A=\frac{4(1-\nu_1)}{G_1}+\frac{4(1-\nu_2)}{G_2}=\frac{8}{E},\quad B=\frac{2-4\nu_1}{G_1}-\frac{2-4\nu_2}{G_2},$$
$$\frac{1}{E}=\frac{1-\nu_1^2}{E_1}+\frac{1-\nu_2^2}{E_2},\quad E_i=2G_i(1+\nu_i) \tag{7}$$

where G_1, G_2 are the shear moduli, ν_1, ν_2 are the Poisson ratios, E_1, E_2 are the moduli of elasticity of bodies "1" and "2", respectively, and E is the composite modulus. Equation (5) can be simplified for either identical materials ($B=0$) or for the frictionless case [$p_x(x)=0$]. Even if the materials are not identical, the effect of the constant B is usually small [16] and is often neglected. Thus, normal/shear stresses do not produce relative tangential/normal displacements.

Normal Loading. Consider normal loading in which a normal load F is applied to a cylinder with the tangential force T equal to zero. This problem has been solved by Baney and Hui [8], but their analysis is summarized here because the results of the normal loading problem determine the contact region used in the sliding and rolling analyses. There is a central contact zone ($-a<x<a$) surrounded by two adhesion zones ($a<|x|<c$) in which the separated surfaces are under a constant tensile stress, as described by the Maugis adhesion model [7]. This configuration is shown in Fig. 1 where, by symmetry, $c_1=c_2=c$, $h_1=h_2=h$, and $e=0$. The tensile adhesive stress is effective up to a separation h, beyond which it vanishes.

The geometric relation for the deformations in the normal direction ($u_y^{(1)}$ and $u_y^{(2)}$ for bodies 1 and 2, respectively) at the contact interface is

$$u_y^{(1)} - u_y^{(2)} = -\delta_0 + \frac{x^2}{2R}, \quad -a < x < a \tag{8}$$

in the contact zone, where δ_0 is the maximum cylinder penetration, which occurs at the center of the contact zone.

By proceeding as Baney and Hui [8], Eqs. (5) and (8) are combined using the Maugis condition in the adhesion zones

$$p_y(x) = -\sigma_0, \quad -c < x < -a, \quad a < x < c \tag{9}$$

with $B=0$. The solution is given by the superposition of the Hertz solution, the solution for an exterior crack (Tada et al. [17], p. 110), and the homogeneous solution of Eq. (5). That superposition of solutions is also subject to the conditions that the stress is bounded at both ends ($x = \pm a$). The result is [8]

$$p_y(x) = \frac{E}{2R}\sqrt{a^2 - x^2} - \frac{2\sigma_0}{\pi}\tan^{-1}\sqrt{\frac{c^2 - a^2}{a^2 - x^2}}, \quad -a < x < a \tag{10}$$

Because the contact half-width (a) and the adhesion half-width (c) are both unknown, two extra equations are necessary. Using force equilibrium in the y-direction, the total applied load F can be related to the contact half-width (a), i.e.,

$$\int_{-a}^{a} p_y(\xi)\, d\xi - 2(c-a)\sigma_0 = F \tag{11}$$

resulting in

$$F = \frac{\pi E a^2}{4R} - 2\sigma_0\sqrt{c^2 - a^2} \tag{12}$$

The geometrical relation for the relative separation of the two bodies at $x=c$ and $x=a$ is

$$(u_y^{(1)} - u_y^{(2)})_a^c = \frac{c^2 - a^2}{2R} - h \tag{13}$$

which can be used along with Eq. (5), and leads to

$$\frac{am}{2R}\sqrt{m^2-1} + \left[\frac{4\sigma_0}{\pi E}\sqrt{m^2-1} - \frac{a}{2R}\right]\ln(m + \sqrt{m^2-1}) - \frac{4\sigma_0 m}{\pi E}\ln(m) = \frac{h}{a}, \quad m = \frac{c}{a} \tag{14}$$

This solution, with adhesion included, shows that the pressure distribution is compressive in the central part of the contact zone but tensile at the ends of the contact zone.

As determined by Baney and Hui [8], the solution (12) and (14) can be written using the dimensionless quantities

$$\bar{F} = \frac{F}{(\pi E w^2 R)^{1/3}}, \quad \bar{a} = \frac{a}{2(R^2 w/\pi E)^{1/3}}, \quad \lambda = \frac{4\sigma_0}{(\pi^2 E^2 w/R)^{1/3}} \tag{15}$$

as

$$\bar{F} = \bar{a}^2 - \lambda\bar{a}\sqrt{m^2-1} \tag{16}$$

and

$$\frac{1}{2}\lambda\bar{a}^2[m\sqrt{m^2-1} - \ln(m + \sqrt{m^2-1})] + \frac{1}{2}\lambda^2\bar{a}[\sqrt{m^2-1}\ln(m + \sqrt{m^2-1}) - m\ln(m)] = 1 \tag{17}$$

Their parameter λ is similar to that defined by Maugis.

It is noted that Eqs. (16) and (17) form a pair of coupled nonlinear equations with $\bar{F}$ known and $\bar{a}$ and m unknown. However, it is more convenient to solve these equations by slowly varying m from just above unity to a large number. Equation (17) is then solved as a quadratic equation for $\bar{a}$ with only one positive root.

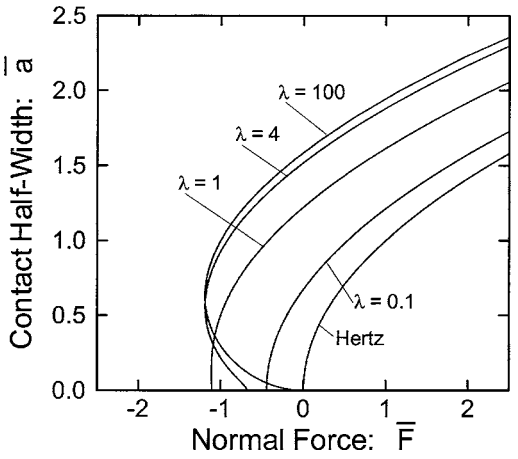


Fig. 2 The variation of the dimensionless contact half-width ($\bar{a}$) with the dimensionless normal load ($\bar{F}$) for various values of λ

The corresponding value of $\bar{F}$ is determined from Eq. (16). The results for the dimensionless contact half-width ($\bar{a}$) versus the dimensionless normal force ($\bar{F}$) for various values of λ are shown in Fig. 2. As discussed in [8], large λ correspond to the JKR regime, small λ approach Hertz contact, and when λ is of order unity the results can be approximated by the DMT theory. For nonzero λ there is a pull-off force which, for a sufficiently large value of λ, occurs at a nonzero contact radius.

Initiation of Sliding. Tangential forces can be transmitted by friction in contacting bodies. Consider a cylinder in contact with a half-space, compressed by a normal force F, and acted upon by a tangential force T (Fig. 1). The problem is also solved for the uncoupled case; i.e., $B=0$ in Eq. (5). Thus, the contact area remains in a state of stick during the application of the normal load and the contact area remains constant during the application of the tangential force. With the normal load constant, the tangential force is gradually increased in order to initiate sliding.

Mindlin studied the initiation of macro-scale sliding of a cylinder using Coulomb friction without adhesion [18]. He first assumed that stick prevailed everywhere during the tangential loading phase. When that problem is solved, however, a singularity exists for the shear stress distribution at the edges $x = |a|$, whereas the normal stress is bounded. Hence, the frictional inequality must be violated; so it was deduced that slip will occur at the edges of the contact zone. Thus, according to Mindlin's theory there is a central stick zone ($|x| < d$) surrounded by slip zones symmetrically located in both the leading and trailing edges.

In nano-scale contacts, adhesion between the two bodies will affect the contact width. This effect in cylindrical contacts is described in the previous section. With respect to tangential loading, in the macro scale, Coulomb friction is used, whereas at the nano scale, the friction stress in the slip zone is assumed to be constant; i.e., the adhesion theory of friction, as discussed previously.

According to the plane strain linear elasticity formulation, the relation between the relative deformations of the bodies in the tangential direction to the boundary stresses is given by Eq. (6). In the slip regions

$$p_x(x) = \tau_0, \quad d < |x| < a \tag{18}$$

whereas in the stick region

$$\frac{d(u_x^{(1)} - u_x^{(2)})}{dx} = 0, \quad -d < x < d \tag{19}$$

Prior to the initiation of global sliding motion, the horizontal shift (i.e., $u_x^{(1)}-u_x^{(2)}$) of the bodies in the stick zone ($-d<x<d$) will be constant.

The solution of Eq. (6) subject to Eqs. (18) and (19) can be found by superposition of the solution for an external crack loaded by a symmetric constant shear stress in a finite region near the crack tip (Tada et al. [[17], p. 110])

$$\tau_{xy}=-\frac{2\tau_0}{\pi}\left\{\sqrt{\frac{a^2-d^2}{d^2-x^2}}-\tan^{-1}\sqrt{\frac{a^2-d^2}{d^2-x^2}}\right\},\quad -d<x<d \tag{20}$$

and the homogeneous solution of Eq. (6) given by, e.g., Erdélyi [19]

$$\tau_{xy}=\frac{D_0}{\sqrt{d^2-x^2}}\quad -d<x<d \tag{21}$$

where D_0 is determined from the condition that τ_{xy} is bounded at $x=|d|$. Thus, the shear stress in the stick region becomes

$$\tau_{xy}=\frac{2\tau_0}{\pi}\tan^{-1}\sqrt{\frac{a^2-d^2}{d^2-x^2}}\quad -d<x<d \tag{22}$$

Force balance in the horizontal direction can be used to find the relation between the half-length of the stick zone (d) and the applied shear force (T). This procedure gives

$$T=2(a-d)\tau_0+\int_{-d}^{d}\tau(\xi)d\xi \tag{23}$$

The integral in Eq. (23) has been evaluated by Baney and Hui in [8] and gives

$$T=2\tau_0\sqrt{a^2-d^2} \tag{24}$$

As the tangential force reaches the critical value of $2a\tau_0$, a state of complete slip (i.e., global sliding) occurs with $d=0$.

Pure Sliding. If the contacting cylinder has a relative sliding motion with respect to the plane it is in contact with, there need not be symmetry, due to the nonlinear nature of adhesion. In this case, the origin of the coordinate system will be chosen in the center of the contact region. An eccentricity e will indicate the value of x corresponding to the apex of the undeformed cylinder. The leading adhesion zone will be a strip ($a<x<c_2$) and the trailing adhesion zone will be another strip ($-c_1<x<-a$), as shown in Fig. 1.

The geometrical relation between the deformations of the bodies in the y-direction inside the contact is

$$u_y^{(1)}-u_y^{(2)}=-\delta_0+\frac{(x-e)^2}{2R} \tag{25}$$

The same elasticity formulation used for the symmetric normal contact can be used here; i.e., Eq. (5) subject to Eq. (25) in the contact region $-a<x<a$. The solution can be found by the superposition of four problems. The first problem is the solution for a constant tensile stress in the leading edge ($a<x<c_2$) and the second is for a constant tensile stress in the trailing edge ($-c_1<x<-a$). These solutions, which correspond to an external crack, are in Tada et al. [[17], p. 107]. The mode-I stress intensity factors at the leading and trailing edges are

$$K_I(a)=-\sigma_0\sqrt{\frac{a}{\pi}}\{(1+m_2)\sqrt{m_2^2-1}+(1-m_1)\sqrt{m_1^2-1}\} \tag{26}$$

$$K_I(-a)=-\sigma_0\sqrt{\frac{a}{\pi}}\{(1+m_1)\sqrt{m_1^2-1}+(1-m_2)\sqrt{m_2^2-1}\}$$

where $m_1=c_1/a$ and $m_2=c_2/a$. The third problem corresponds to the solution of Eqs. (5) and (25) without adhesion; i.e., a Hertz-type solution with an eccentricity that can be found using [19]

$$p_Y(x)=\frac{E}{2R}\left(\sqrt{a^2-x^2}-e\sqrt{\frac{a-x}{a+x}}\right),\quad -a<x<a \tag{27}$$

and the fourth is the homogeneous solution of Eq. (5); i.e.,

$$p_Y(x)=D/\sqrt{a^2-x^2},\quad -a<x<a \tag{28}$$

The sum of these four solutions must be such that the normal stress is bounded at the ends at $x=|a|$. Recall that the model-I stress intensity factor at $x=a$ is defined by

$$K_I(a)=\lim_{x\to a}\sqrt{2\pi(a-x)}\tau_{yy} \tag{29}$$

The condition that the normal stress is bounded at $x=a$ becomes

$$D=-\frac{\sigma_0 a}{\pi}\{(1+m_2)\sqrt{m_2^2-1}+(1-m_1)\sqrt{m_1^2-1}\} \tag{30}$$

whereas the requirement that the normal stress is bounded at $x=-a$, becomes

$$e=\frac{\sigma_0 R}{E\pi}\{(1+m_1)\sqrt{m_1^2-1}+(1-m_2)\sqrt{m_2^2-1}\}+\frac{DR}{Ea} \tag{31}$$

The geometrical relations for the separation of the two bodies can be used that is in the same form as Eq. (13), but with different integration limits due to nonsymmetry.

When the cylinder is in a state of steady sliding, the adhesion effect in the trailing edge is assumed to be larger than in the leading edge. This assumption is considered valid because adhesion between surfaces is affected by surface contamination. Since the surface will be partially cleaned due to the sliding motion of the contacting surfaces, adhesion in the trailing edge is expected to be larger than in the leading edge. This effect is accounted for by taking the adhesion separation distance in the trailing edge (h_1) larger than in the leading edge (h_2), whereas σ_0 is assumed unchanged. Due to the difference between h_1 and h_2, the adhesion width values will not be the same in the leading and the trailing edges. Since there are two more unknowns, the separation equations must be written for both the leading and trailing edges. At the leading edge this procedure gives

$$(u_y^{(1)}-u_y^{(2)})_a^{c_2}=\frac{(c_2-e)^2-(a-e)^2}{2R}-h_2 \tag{32}$$

and at the trailing edge it yields

$$(u_y^{(1)}-u_y^{(2)})_{-a}^{-c_1}=\frac{(-c_1-e)^2-(-a-e)^2}{2R}-h_1 \tag{33}$$

Superposition of the four solutions yields

$$\frac{2\sigma_0 a m_2}{\pi E h_1}(m_2^2-1)+\frac{2\sigma_0 a}{\pi E h_1}\left\{(m_1+m_2)\cosh^{-1}\left|\frac{1+m_1m_2}{m_1+m_2}\right|-m_1\sqrt{m_1^2-1}\sqrt{m_2^2-1}\right\}-\frac{a^2}{2Rh_1}\{(m_2-2e/a)\sqrt{m_2^2-1}-(1-2e/a)\ln(m_2+\sqrt{m_2^2-1})\}+\frac{2D}{Eh_1}\ln(m_2+\sqrt{m_2^2-1})=-\frac{h_2}{h_1} \tag{34}$$

from the leading edge condition and

$$\frac{2\sigma_0 a m_1}{\pi E h_1}(m_1^2-1)+\frac{2\sigma_0 a}{\pi E h_1}\left\{(m_1+m_2)\cosh^{-1}\left|\frac{1+m_1m_2}{m_1+m_2}\right|-m_2\sqrt{m_1^2-1}\sqrt{m_2^2-1}\right\}-\frac{a^2}{2Rh_1}\{(m_1+2e/a)\sqrt{m_1^2-1}-(1-2e/a)\ln(m_1+\sqrt{m_1^2-1})\}+\frac{2D}{Eh_1}\ln(m_1+\sqrt{m_1^2-1})=-1 \tag{35}$$

from the trailing edge condition.

The applied normal force can be found from force equilibrium in the y-direction, resulting in

$$F=\pi D+\frac{\pi E a^2}{4R}-\frac{\pi E a e}{2R} \tag{36}$$

A resultant moment will act on the upper body due to the asymmetric normal stress distribution. If moment equilibrium is written with respect to the center of contact, the resultant moment (clockwise direction acting on the half-space taken to be positive) is found and is given by

$$M=\frac{\pi E a^2 e}{4R} \tag{37}$$

The solution of Eqs. (34)–(37) can be facilitated with the nondimensional quantities (Eq. (15)). It can readily be shown that

$$\frac{2\sigma_0 a}{\pi E h_1}=\frac{\lambda^2\bar{a}}{4},\quad \frac{2D}{Eh_1}=\frac{\pi\bar{D}\lambda^2\bar{a}}{4},\quad \bar{D}=\frac{D}{\sigma_0 a},\quad \frac{a^2}{Rh_1}=\lambda\bar{a}^2,\quad \bar{e}=\frac{e}{a},\quad \bar{M}=\frac{M}{2wR} \tag{38}$$

in which $w=\sigma_0 h_1$. By using Eqs. (38), the nondimensional force and moment become

$$\bar{F}=\frac{\pi}{2}\lambda\bar{D}\bar{a}+\bar{a}^2-2\bar{e}\bar{a}^2,\quad \bar{M}=\bar{e}\bar{a}^3 \tag{39}$$

Thus Eqs. (34) and (35), along with the explicit expressions in Eqs. (30) and (31) and the nondimensional quantities in Eq. (38), represent a pair of equations which, for specified λ and h_1/h_2, can be solved for m_1 and m_2. Finally, the nondimensional force and moment are found from Eq. (39).

The results for the dimensionless contact half-width ($\bar{a}$) versus the dimensionless normal force ($\bar{F}$) for various values of λ are shown in Figs. 3 and 4 for $h_1/h_2=2$ and $h_1/h_2=5$, respectively. As h_1/h_2 increases, the force required to produce a given contact width increases. This effect is greater for large λ (JKR region, due to elastic deformation) than it is for moderate λ (DMT regime, limited elastic deformation) or small λ (Hertz regime, small adhesion). Figures 5 and 6 show the dimensionless adhesion half-width difference (m_1-m_2) versus dimensionless contact half-width ($\bar{a}$) for different values of λ with $h_1/h_2=2$ and $h_1/h_2=5$, respectively. This measure of the asymmetry of the adhesion zones becomes large for small values of the dimensionless contact radius ($\bar{a}$). It is also much greater for small λ than for large λ. This result may appear counterintuitive. However, the contact half-width is normalized by a quantity that includes the cube-root of the work of adhesion, whereas λ varies as the two-thirds power of w. In addition, large λ corresponds to greater elastic deformation, which is better capable of accommodating the asymmetry in the work of adhesion. The results for the dimensionless average adhesion length $(m_1+m_2)/2$ versus dimensionless contact radius are shown for various values of λ in Figs. 7 and 8 for $h_1/h_2=2$ and $h_1/h_2=5$, respectively. Finally, the dimensionless moment ($\bar{M}$) versus the dimensionless normal force ($\bar{F}$) is shown for various values of λ in Figs. 9 and 10 for $h_1/h_2=2$ and $h_1/h_2=5$, respectively. Note that as the normal force approaches the pull-off force, the moment approaches a finite value, even in the small λ regime where $\bar{a}$ vanishes at pull-off. This result is the combined effect of the asymmetry in the work of adhesion along with the small elastic deformation.

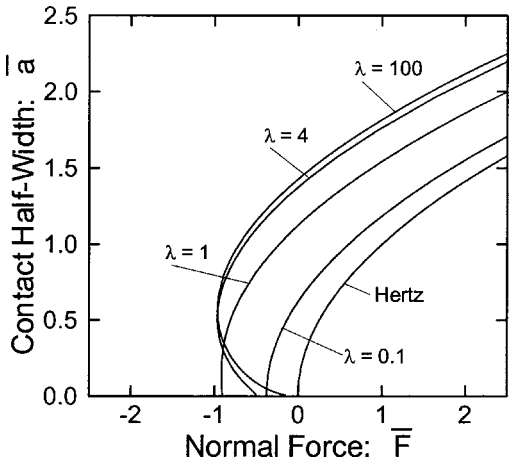


Fig. 3 The variation of the dimensionless contact half-width ($\bar{a}$) with the dimensionless normal load ($\bar{F}$) for various values of λ during sliding with $h_1/h_2=2$

Rolling. The problem of steady-state rolling of an elastic cylinder on an elastic half-space (or equivalently one cylinder rolling on another) with Coulomb friction was solved by Carter [20]. According to Carter's solution, the leading edge of the contact zone ($d<x<a$) is in a state of stick and the trailing edge ($-a<x<d$) is in a state of slip. In Carter's approach, $x=x'-Vt$ is a coordinate moving to the right with the speed of the contact region, and x' represents the stationary coordinate system. As in the application to a locomotive wheel [20], the upper body is the driving cylinder. During this rolling motion, the linear velocity of the center of the cylinder is slightly less than ωR, where ω is the angular velocity. Similarly, for two rollers, the angular velocities of the rollers will not be inversely proportional to their radii. Due to slip, the driving roller linear velocity will have a somewhat greater magnitude than that of the driven roller. The creep velocity represents this velocity difference.

At the nano scale the shear stress (friction stress) in the slip

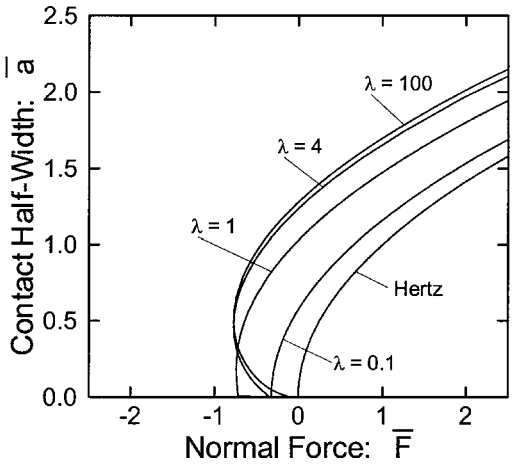


Fig. 4 The variation of the dimensionless contact half-width ($\bar{a}$) with the dimensionless normal load ($\bar{F}$) for various values of λ during sliding with $h_1/h_2=5$

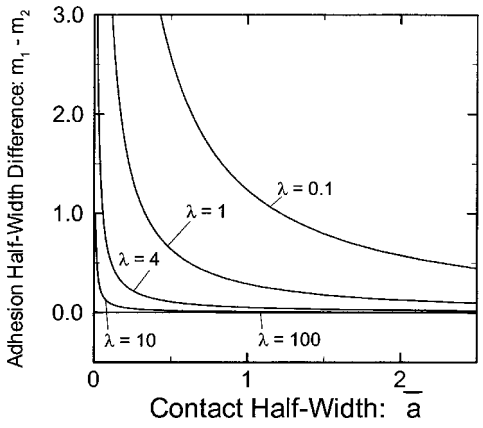


Fig. 5 The difference between the dimensionless adhesion half-widths of the trailing and leading edges (m_1-m_2) versus dimensionless contact half-width ($\bar{a}$) for various values of λ during sliding with $h_1/h_2=2$

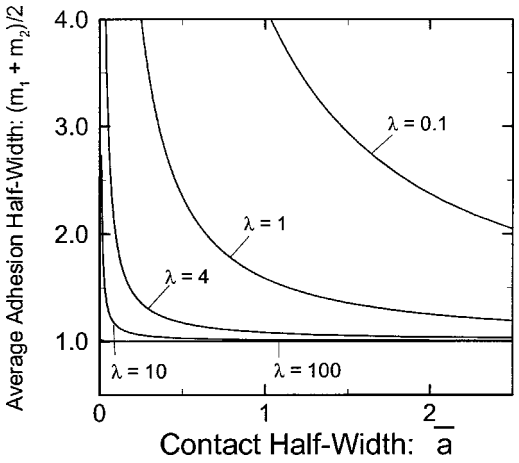


Fig. 7 The average dimensionless adhesion half-width of the trailing and leading edges (m_1+m_2)/2 versus dimensionless contact half-width ($\bar{a}$) for various values of λ during sliding with $h_1/h_2=2$

zone is assumed to be constant, as previously discussed. As with sliding, adhesion affects the relation between the normal force and contact width.

The tangential relative displacement (shift) between the bodies is expressed as

$$s(x,t)=u_x^{(2)}(x,0,t)-u_x^{(1)}(x,0,t)+C(t) \tag{40}$$

where C represents the rigid body motion of the upper body relative to the lower body. In the stick zone, the time derivative of the shift in the moving coordinate system is zero. The stick condition can be written as

$$\dot{s}(x,t)=V\frac{d}{dx}(u_{x2}-u_{x1})+\dot{C}=0, \quad d<x<a \tag{41}$$

where $\dot{C}(t)$ is the constant rigid body slip (or creep) velocity. Furthermore the shear stress is constant in the slip region, i.e.,

$$p_x(x)=-\tau_0, \quad -a<x<d \tag{42}$$

The solution of Eq. (6), subject to Eqs. (41) and (42), can be found by superposition of the solution for a crack external to the stick region and loaded in shear on one side, i.e., the slip zone (Tada et al. [[17], p. 107]), the solution of Eq. (6) due to the constant creep velocity, and the homogeneous solution of Eq. (6). The mode-II stress intensity factors for the external crack problem at the ends of the stick zone are [17]

$$K_{\mathrm{II}}(d)=\frac{\tau_0}{\sqrt{\pi}}\left\{2\sqrt{a\left(\frac{a+d}{a-d}\right)}+\frac{1}{2}\sqrt{2(a-d)}\cosh^{-1}\left(\frac{3a+d}{a-d}\right)\right\} \tag{43}$$

$$K_{\mathrm{II}}(a)=\frac{\tau_0}{\sqrt{\pi}}\left\{2\sqrt{a\left(\frac{a+d}{a-d}\right)}-\frac{1}{2}\sqrt{2(a-d)}\cosh^{-1}\left(\frac{3a+d}{a-d}\right)\right\}$$

whereas the solution to the creep velocity and the homogeneous solution are

$$p_x(x)=\frac{\dot{C}E}{2V}\sqrt{\frac{a-x}{x-d}}+\frac{D_1}{\sqrt{(a-x)(x-d)}}, \quad d<x<a \tag{44}$$

The requirement that the solution be bounded at $x=a$ and $x=d$ leads to

$$D_1=-\frac{\tau_0}{\pi}\left\{\sqrt{2a(a+d)}-\frac{1}{2}(a-d)\cosh^{-1}\left(\frac{3a+d}{a-d}\right)\right\} \tag{45}$$

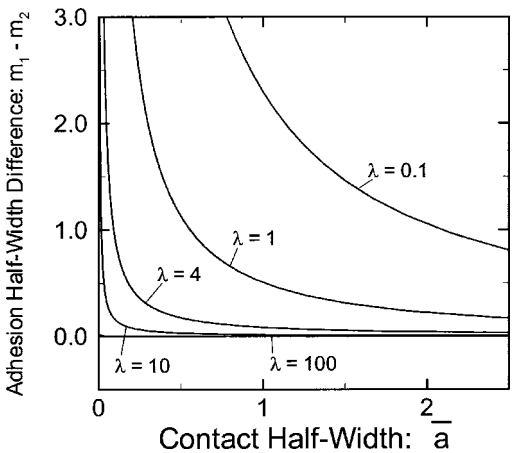


Fig. 6 The difference between the dimensionless adhesion half-widths of the trailing and leading edges (m_1-m_2) versus dimensionless contact half-width ($\bar{a}$) for various values of λ during sliding with $h_1/h_2=5$

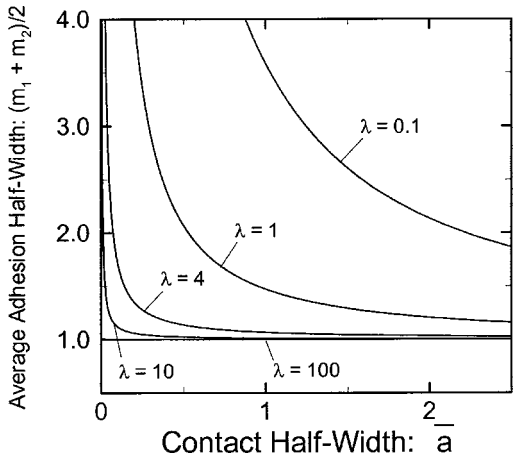


Fig. 8 The average dimensionless adhesion half-width of the trailing and leading edges (m_1+m_2)/2 versus dimensionless contact half-width ($\bar{a}$) for various values of λ during sliding with $h_1/h_2=5$

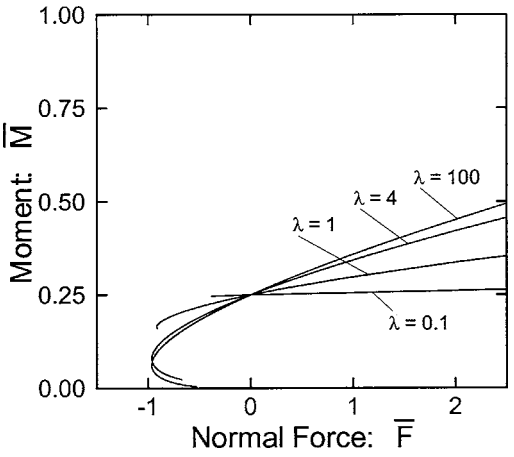


Fig. 9 The dimensionless resultant moment ($\bar{M}$) versus dimensionless contact half-width ($\bar{a}$) for various values of λ during sliding with $h_1/h_2=2$

and

$$\frac{\dot{C}}{V} = -\frac{2\tau_0}{\pi E}\cosh^{-1}\left(\frac{3a+d}{a-d}\right) \tag{46}$$

If force equilibrium is written in the x-direction, the applied shear force can be related to the contact width parameter (d) by

$$T = \int_{-a}^{a} p_x(x)dx = -\tau_0\sqrt{2a(d+a)} \tag{47}$$

Equation (47) gives the variation of the applied tangential force with the extent of the slip zone. The greater the traction force, the larger is the slip zone. As $d \to a$ the rolling motion approaches complete slip. Equation (46) gives the dimensionless creep velocity, which is linear in τ_0/E and varies nonlinearly with the slip zone parameter (d/a). As the traction force increases, d increases and hence the magnitude of the creep velocity increases logarithmically according to Eq. (46). The meaning of the negative creep velocity is that ωR for the driving wheel is greater than the velocity of the contact zone.

Conclusions

As the scale of contacting bodies decreases, adhesion effects become significant, especially for smooth surfaces and lightly loaded systems. The behavior of a cylinder (e.g., a fiber), which adheres to a substrate and is subjected to a combination of normal and tangential forces, is the subject of this investigation. In the presence of a tangential load, the cylinder might slide, roll, or undergo a complex motion. This paper treats the two-dimensional elastic contact problem of the cylinder on the substrate during a rolling/sliding motion and includes the effect of adhesion using the Baney and Hui version of the Maugis–Dugdale model.

Mindlin's classic investigation of the initiation of sliding of a cylinder on a half-plane used Coulomb friction. As the scale decreases, two effects should be included. First, adhesion increases the contact width, especially under light normal loading. Second, according to the adhesion theory of friction for a single real area of contact, the shear stress can be assumed to be constant in the slip regions. These effects are included for the uncoupled case in which normal/shear loading does not produce relative tangential/normal displacements. During the initiation of sliding, there is a central stick zone surrounded by slip regions in the leading and trailing edges. As the applied tangential load increases, the lengths of the slip zones increase until, at a certain value of the tangential force, there is complete slip.

During steady sliding, the abrasive action of the shear stress can be expected to partially clean the surface, resulting in different leading and trailing edge adhesive properties. This effect is included in the model of steady nano-scale sliding.

Carter investigated the rolling of a cylinder on a substrate at the macro-scale. Similarly to the sliding case, as the dimensions approach the nano-scale, adhesion under normal loading and constant shear stress in the slip regions under tangential loading are assumed. The variations of the creep velocity and the length of the stick zone with the applied shear force are determined. As the traction force increases, the stick zone length decreases and the creep velocity increases eventually leading to pure slip with rotation.

Future work should include the analysis of the transition from sliding to rolling and applications to technological problems. It should also address the analysis of the three-dimensional problem (spherical case) for sliding and rolling motions.

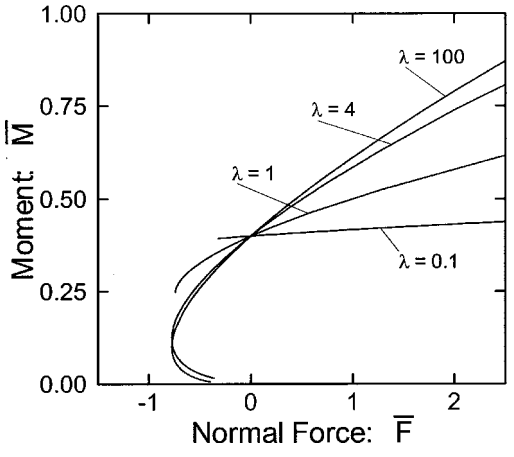


Fig. 10 The dimensionless resultant moment ($\bar{M}$) versus dimensionless contact half-width ($\bar{a}$) for various values of λ during sliding with $h_1/h_2=5$

References

[1] Rimai, D. S., and Quesnel, D. J., 2001, *Fundamentals of Particle Adhesion* (Polymer Surfaces and Interfaces Series), Global Press.

[2] Bradley, R. S., 1932, "The Cohesive Force Between Solid Surfaces and the Surface Energy of Solids," Philos. Mag., **13**, pp. 853–862.

[3] Johnson, K. L., Kendall, K., and Roberts, A. D., 1971, "Surface Energy and the Contact of Elastic Solids," Proc. R. Soc. London, Ser. A, **324**, pp. 3101–3130.

[4] Derjaguin, B. V., Muller, V. M., and Toporov, Y. P., 1975, "Effect of Contact Deformations on the Adhesion of Particles," J. Colloid Interface Sci., **67**, pp. 314–326.

[5] Tabor, D., 1976, "Surface Forces and Surface Interactions," J. Colloid Interface Sci., **58**, pp. 1–13.

[6] Greenwood, J. A., 1997, "Adhesion of Elastic Spheres," Proc. R. Soc. London, Ser. A, **453**, pp. 1277–1297.

[7] Maugis, D., 1992, "Adhesion of Spheres: the JKR-DMT Transition Using a Dugdale Model," J. Colloid Interface Sci., **150**, pp. 243–269.

[8] Baney, J. M., and Hui, C.-Y., 1997, "A Cohesive Zone Model for the Adhesion of Cylinders," J. Adhes. Sci. Technol., **11**, pp. 393–406.

[9] Sari, O. T., 2003, "Nano-Scale Effects in Adherence, Sliding and Rolling of a Cylinder on a Substrate," MS thesis, Mechanical Engineering Dept., Northeastern University, Boston.

[10] Carpick, R. W., Agrait, N., Ogletree, D. F., and Salmeron, M., 1996, "Measurement of Interfacial Shear (Friction) With an Ultrahigh Vacuum Atomic Force Microscope," J. Vac. Sci. Technol. B, **14**, pp. 1289–1295.

[11] Homola, A. M., Israelachvili, J. N., McGuiggan, P. M., and Gee, M. L., 1990, "Fundamental Experimental Studies in Tribology: The Transition From 'Interfacial' Friction of Undamaged Molecularly Smooth Surfaces to 'Normal' Friction With Wear," Wear, **136**, pp. 65–83.

[12] Hurtado, J. A., and Kim, K.-S., 1999, ''Scale Effects in Friction of Single Asperity Contacts: Part I; From Concurrent Slip to Single-Dislocation-Assisted Slip,'' Proc. R. Soc. London, Ser. A, **455**, pp. 3363–3384.
[13] Hurtado, J. A., and Kim, K.-S., 1999, ''Scale Effects in Friction in Single Asperity Contacts: Part II; Multiple-Dislocation-Cooperated Slip,'' Proc. R. Soc. London, Ser. A, **455**, pp. 3385–3400.
[14] Israelachvili, J., 1992, *Intermolecular and Surface Forces*, Academic Press, New York.
[15] Barber, J. R., 2002, *Elasticity*, 2nd ed., Kluwer Academic, Dordrecht, The Netherlands.
[16] Johnson, K. L., 1985, *Contact Mechanics*, Cambridge University Press, Cambridge, UK.
[17] Tada, H., Paris, P. C., and Irwin, G. R., 2000, *The Stress Analysis of Cracks Handbook*, 3rd ed., ASME, New York.
[18] Mindlin, R. D., 1949, ''Compliance of Elastic Bodies in Contact,'' J. Appl. Mech., **17**, pp. 259–268.
[19] Erdélyi, A., 1954, *Tables of Integral Transforms*, Bateman Manuscript Project, McGraw-Hill, New York.
[20] Carter, F. W., 1926, ''On the Action of a Locomotive Driving Wheel,'' Proc. R. Soc. London, Ser. A, **112**, pp. 151–157.

Veering Phenomena in Systems With Gyroscopic Coupling

Stefano Vidoli
e-mail: stefano.vidoli@uniroma1.it

Fabrizio Vestroni
e-mail: vestroni@uniroma1.it

Dipartimento di Ingegneria Strutturale e Geotechnica,
Università degli Studi di Roma "La Sapienza,"
Via Eudossiana 18,
00184 Roma, Italia

The sharp divergence of two root-loci for a critical value of the parameters is called veering. Veering phenomena are interesting since they involve relevant energetic exchanges between the eigenmodes and strongly affect the undamped forced response of the system. A straightforward perturbation approach has already been used in the literature to analyze the dependence of the eigensprectrum on a system parameter and formulate a veering criterion. This perturbation approach and other ideas are generalized to the study of veering in discrete and continuous systems with gyroscopic operators of internal coupling and the results applied to a real electromechanical interaction.
[DOI: 10.1115/1.1940666]

1 Introduction

A linear conservative system is considered endowed with all the possible kinds of coupling between its degrees of freedom: elastic, gyroscopic, and inertial. The associated eigenvalue problem is parametrized by a set of coordinates on the space of system parameters $\mathcal{P}$; thus the dependence on parameters of the eigenvalues can be geometrically described by N hypersurfaces in the space $\mathcal{P}\times\mathbb{R}$. However, for practical reasons, only some sections of these hypersurfaces are usually plotted. Here the attention is focused on the two-dimensional sections where the eigenvalues are plotted as set of curves in $\hat{\mathcal{P}}\times\mathbb{R}$, $\hat{\mathcal{P}}$ being a one-dimensional section of $\mathcal{P}$. The special regions examined are the *crossing* and *veering* regions; in the latter case, as the considered coordinate on $\hat{\mathcal{P}}$ increases, two or more curves tend to cross one another, veer away, and then diverge. Veering regions are particularly important since, as will be shown, they are associated with conditions of maximum coupling between the eigenmodes.

One of the first occurrences of the word veering in the literature is due to Leissa [1] and relates to the natural frequencies of a clamped rectangular membrane. Several other authors analytically found or experimentally observed veering phenomena: Doll and Mote [2] dealing with vessels under pressure, Ramaswamy and Marcus [3] in molecular physics, Nair and Durvasula [4] in the study of plate vibrations, Triantafyllou [5], Russel and Lardner [6], Behbahani and Perkins [7], Cheng and Perkins [8] in the field of cable dynamics. The relations between eigenvalue veering and mode localization have been investigated in particular within the field of periodic systems perturbed by a small disorder parameter: contributions have been given by Pierre [9], Triantafyllou [10], and Natsiavas [11].

A precise veering criterion is found in Perkins and Mote [12]; they used a straightforward perturbation technique to study both the case of discrete and continuous systems. However, since at least two parameters are involved in producing the veering, this technique fails in the vicinity of the avoided crossing. To overcome this drawback Pierre [13] used a modified perturbation approach, while Happawana et al. [14] and Natsiavas [11] proposed two similar techniques of singular perturbation. Finally geometric approaches to the problem have been given by Arnold [15] and Triantafyllou [10].

In order to prove an optimality criteria for an electromechanical continuous system some hints of Arnold are used here and the direct perturbation technique [12] is extended to the case of gyroscopic coupling-operators. Indeed we only need a tool for veering detection in systems with multiple degrees of freedom, as, for instance, the electromechanical system presented in Sec. 5; we do not need a more refined perturbation expansion since the local (near veering) behavior of both eigenvalues and eigenvectors can be analytically described through a change of coordinates of the general analytical solution for the case of two degrees of freedom (dof) systems. Hence Sec. 2 is devoted to the exact solution of a 2-dof system for all possible kinds of linear coupling: elastic, gyroscopic, and inertial. The forced response of the system is also analytically computed and the veering regions measured. A classification of veering is introduced and the relation of this phenomenon with the mode localization is enlightened. Based on these results, a strategy for the efficient exchange of energy between structural modes—provided they are both coupled with almost another one electric mode—is proposed.

2 Geometric Description of Veering

A geometric description of the veering phenomenon can be given, following the approach outlined in [15], in terms of ellipsoids associated to symmetric positive definite eigenproblems. Some results due to Meirovitch [16] are also needed in order to draw the basic eigenproblem into a positive definite, symmetric form.

The equations governing the dynamics of the considered linear conservative system with N degrees of freedom are written as

$$M\ddot{q}+R\dot{q}+Lq=f(t),\quad q(t),f(t)\in\mathbb{R}^N. \tag{1}$$

A superimposed dot means the derivative with respect to time t, M, and L are positive definite, symmetric matrices on $\mathbb{R}^N(\mathrm{PSym}_N)$, while R is a skew-symmetric matrix on $\mathbb{R}^N$ (Skw_N) representing the presence of gyroscopic coupling. These matrices are supposed to depend smoothly on a set of real parameters $\mathcal{P}\subset\mathbb{R}^{N(3N-1)/2}$. A simple calculation shows that the maximal dimension of $\mathcal{P}$, i.e., the total number of independent parameters in Eq. (1) is $N(3N-1)/2$; this equals the total number of independent components in one skew-symmetric and two symmetric matrices minus N possible choices to make dimensionless the Lagrangian parameters.

Due to the presence of the gyroscopic addend, the eigenvalue problem associated to Eq. (1):

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, June 20, 2003; final revision, February 14, 2004. Associate Editor: M. P. Mignolet. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

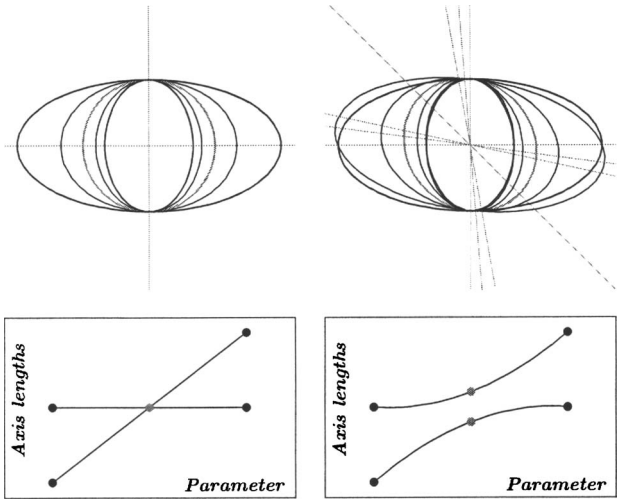


Fig. 1 Sequences of ellipses starting and ending in the same points of $\mathcal{E}$. The major axes are depicted as thin gray lines.

$$(L+\lambda_h^2 M+\lambda_h R)u_h=0, \quad q(t)=\sum_h u_h \exp(\lambda_h t), \tag{2}$$

is not symmetric. Thus, apparently, the related properties, as the eigenvectors orthogonality, do not hold. However in [16] a general procedure is introduced to lead the eigenvalue problem (2) to a symmetric form of size $2N$, namely:

$$(S-\omega_h^2 T)y_h=0, \qquad z_h=\frac{1}{\omega_h}T^{-1}Gy_h,$$
$$T,S\in \mathrm{PSym}_{2N}, \qquad G\in \mathrm{Skw}_{2N}, \tag{3}$$

where S and T are two positive definite, symmetric matrices, while the original eigenvalues and eigenvectors of problem (2) are recovered by

$$\{\lambda_h u_h, u_h\}=y_h+iz_h, \quad \lambda_h=i\omega_h. \tag{4}$$

Hence, to every conservative linear system (1), is associated the symmetric eigenvalue problem $(3)_1$ of size $2N$. Since each symmetric eigenvalue problem on $\mathbb{R}^{2N}$ is described by an ellipsoid in $\mathbb{R}^{2N}$, to find double crossing points, i.e., points in the space of parameters $\hat{\mathcal{P}}$ where two eigenvalues coincide, is equivalent to finding points where the associated ellipsoid has two axes of equal lengths or in other words is an ellipsoid of revolution.

Two useful circumstances are recalled:

(1) the set of ellipsoids on $\mathbb{R}^{2N}$ is a manifold $\mathcal{E}$ of dimension $N(2N+1)$;
(2) the set of ellipsoids of revolution, with only two equal axes, is a finite union $\mathcal{E}_R$ of smooth manifolds in $\mathcal{E}$ with codimension 2 or higher; a proof is given in [15].

An immediate consequence is that $\mathcal{E}_R$ does not divide the space $\mathcal{E}$: thus two arbitrary points in $\mathcal{E}$ can be connected through a continuous curve of ellipsoids with no equal axes. This also means that two arbitrary sets of eigenvalues can always be continuously connected through continuous curves of eigenvalues that do not cross. In Fig. 1 two different sequences of ellipses (sampling of continuous parametrized paths in $\mathcal{E}$) are shown beginning and ending in the same ellipses; namely an ellipse with an horizontal major axis is transformed continuously into an ellipse with a vertical major axis, the length of which equals the initial minor axis. Above them, the axis lengths, meaning the eigenvalues, are correspondingly drawn as functions of the increasing parameter.

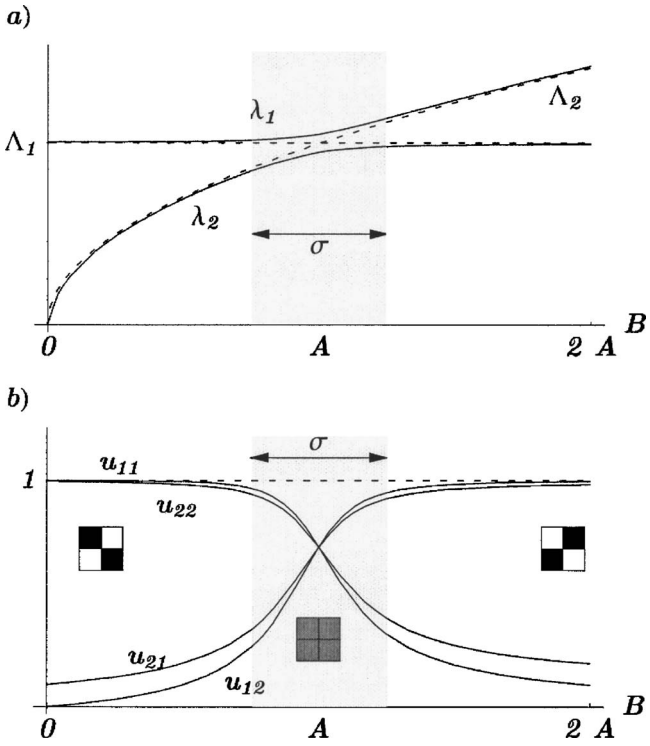


Fig. 2 Veering and crossing geometry: (a) eigenvalues as functions of *B*; (b) eigenvectors components as functions of *B*

In the first case the sequence passes through a circle, i.e., an ellipse of revolution; in the second there is no ellipse of revolution whatsoever. Note that, in order to interchange the axes, the only way to avoid the circle is to rotate the ellipse and as a consequence to incline the major and minor axes. Hence the only way to avoid a crossing point is for the eigenvectors to rotate continuously; in this process certainly they reach an inclination of 45° with respect to the horizontal, i.e., a condition of maximum coupling between the initially vertical and horizontal components.

Veering is exactly the described process: in avoiding a crossing point a condition of maximum coupling is necessarily attained between the two involved components q_1 and q_2. In the following sections the parameters governing both the possibility of veering and its rate (how fast the major and minor axis interchanges) are examined in detail.

3 Exact Solution for 2-dof Systems

In this section some aspects of the free and forced oscillations of systems under veering conditions are recalled and analytically described. To this end a 2-dof system is considered for which the eigenvalue problem (2) is described by the real-valued 2×2 matrices

$$L=\begin{pmatrix} A & C_L \\ C_L & B \end{pmatrix}, \quad R=\begin{pmatrix} 0 & C_R \\ -C_R & 0 \end{pmatrix}, \quad M=\begin{pmatrix} 1 & C_M \\ C_M & 1 \end{pmatrix}, \tag{5}$$

and $u_h=(u_{h1},u_{h2})^{\mathsf{T}}$. An arbitrary kind of coupling between the two degrees of freedom is considered through the coupling parameters C_r $(r=L,R,M)$; when these last vanish, the eigenvalues of the systems are obviously

$$\Lambda_1 := \lambda_1|_{C_r=0}=\pm i\sqrt{A}, \quad \Lambda_2 := \lambda_2|_{C_r=0}=\pm i\sqrt{B}, \tag{6}$$

and the associated eigenvectors are $U_1=(1,0)^{\mathsf{T}}$ and $U_2=(0,1)^{\mathsf{T}}$. The two eigenvalues, corresponding to the plus sign, are plotted as dashed curves in Fig. 2(*a*) as functions of the system parameter B; due to the vanishing coupling there is a crossing at $B=A$. However, when at least one of C_r is different from zero, the crossing is

avoided and the two curves sharply veer and diverge; thus the out-of-diagonal coefficients C_r control whether veering can take place or not. The analytical expression for the eigenvalues of the system (1) and (5) follows:

$$\lambda_{1,2} = \pm i \sqrt{\frac{D \mp \sqrt{D^2 - 4(C_M^2 - 1)(C_L^2 - AB)}}{2(C_M^2 - 1)}},$$
$$D := 2C_L C_M - A - C_R^2 - B. \tag{7}$$

The veering phenomenon is shown by the solid curves in Fig. 2(*a*): the eigenvalues loci approach, but never cross, and then diverge, the minimum distance being attained at $B=A$. The region of distortion of $\lambda_{1,2}$ from $\Lambda_{1,2}$ is sketched with the gray color and its length is called σ.

Also the associated eigenvectors components vary as the parameter B increases. Referring to Fig. 2(*b*), out of the veering region the eigenvectors are almost near to the uncoupled case but, when passing through the veering, there is a permutation of the eigenvectors order. However this permutation is achieved smoothly, as the smooth transition of the axes of the associated ellipsoid during veering, and in the veering region all the eigenvector components have non-vanishing comparable values, i.e., there is no pure eigenvector.

The picture presented suggests a consideration of the relation between veering and localization, see [9–11]. If one, respectively, calls *lumped* and *spread* an eigenvector with only one relevant component and a mode with all its components of comparable nonvanishing value, the aforementioned behavior of the eigenvectors leads to localization and globalization phenomena. In the veering region, indeed, the modes result to be spread, while away from veering, or in the unperturbed situation, they tend to be lumped.

This mathematical circumstance, due to the presence of small coupling terms, is reflected in opposite evidence from a physical point of view: when dealing with disordered systems, the parameters C_i physically mean the disorder levels, the unperturbed situation, to start with, has typically spatially global modes and, as a consequence, the actual modes become local in the veering region. On the contrary, when the parameters C_i physically represent coupling constants, the unperturbed situation has typically spatially localized modes; this means that the actual modes become global in the veering region.

The length σ of the veering or distortion region represents crucial information in many applications; it can be analytically determined solving

$$\frac{\lambda_h - \Lambda_h}{\Lambda_h} = \xi, \quad h = 1,2; \tag{8}$$

where ξ is a given small number introduced to measure the frequency resolution between the coupled (λ_h) and uncoupled (Λ_h) eigensolutions. We do not compute the length σ in the general case when all the coupling parameters are present, but the values σ_r ($r=L,R,M$), i.e., the lengths of the distortion regions when C_r is the only nonvanishing coupling parameter. Thus, for instance, σ_R represents the length of the distortion region when $C_R \neq 0$ and $C_L = C_M = 0$. For the three cases the following dimensionless expressions are obtained:

$$\frac{\sigma_L}{2A} = \frac{C_L^2}{A^2 \xi} - \xi, \quad \frac{\sigma_R}{2A} = \frac{C_R^2(1+\xi)}{A\xi} - \xi, \quad \frac{\sigma_M}{2A} = \frac{C_M^2(1+\xi)^2}{\xi} - \xi. \tag{9}$$

These results are compared assuming $\eta := C_L^2/A^2 = C_R^2/A = C_M^2$; it is easily seen that, under this assumption, one gets $\sigma_L < \sigma_R < \sigma_M$. Thus the inertial coupling—measured by C_M—involves the largest veering region, while the stiffness coupling—measured by C_L—involves the smallest one. The lengths of the veering regions are also affine functions of the dimensionless coupling parameter η; hence, if η is small, one can mistakenly interpret a veering for a crossing depending on frequency resolution. Certainly in many physical circumstances the length of the distortion region can be very small and, from an experimental point of view, hard to detect; as matter of fact outside the distortion region the difference of the coupled from the uncoupled case vanishes and, observing the eigensolution only, the distinction of a veering from a crossing can be very difficult.

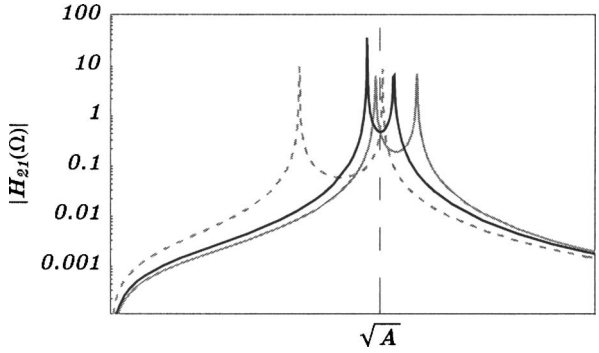


Fig. 3 Frequency-response function $H_{21}(\Omega)$ for different values of the ratio B/A (black $B/A=11$; gray $B/A=1.25$; dashed $B/A=0.75$)

A solution to this experimental difficulty can be found by observing the forced oscillations near the suspected veering conditions. The attention should be focused on the frequency response function:

$$H_{21}(\Omega) = \frac{\tilde{u}_2(\Omega)}{\tilde{f}_1(\Omega)}, \quad \text{where } \tilde{g}(\Omega) := \int_0^\infty g(t)\exp(-i\Omega t)dt, \tag{10}$$

obtained measuring the second component output of the same dynamical system loaded only in its first component, $f(t) = (f_1(t),0)^T$, and, in particular, on its behavior near the veering condition, i.e., when both $\Omega^2 \to A$ and $B \to A$. The frequency-response function $H_{21}(\Omega)$, indeed, identically vanishes when all the coupling parameters C_r are zero, but is sensibly different from zero when at least one coupling parameter is present. Figure 3 shows the graph of $|H_{21}(\Omega)|$ for three different values of the parameters ratio B/A, being $C_L = C_M = 0$ and $C_R^2 = A/100$. The intensity and the bandwidth of the frequency response $H_{21}(\Omega)$ is sensibly higher in veering conditions; actually its minimum value in this case is given by

$$\|H_{21}(\sqrt{A})\|_{B\to A} = \frac{C_R}{\sqrt{A}(C_R^2 + 4A\delta_1\delta_2)}, \tag{11}$$

where δ_1 and δ_2 are the percentages of critical damping for the two degrees of freedom, respectively. Measurements of Eq. (11) allow one to estimate the value of the coupling constant C_r once an estimate of the damping coefficients by standard techniques is obtained. The two close peaks of the frequency response function are responsible for the well-known beating phenomenon in the free and forced oscillations of the system; this phenomenon is observable only in the transient response of slightly damped systems. Similar situations arise when an arbitrary combination of the coupling parameters is present; it is worth noticing that, in case of negligible damping coefficients, the maximum amplitude reached by the second component and the time needed to achieve this level of oscillation are proportional to the inverse of the coupling parameters C_r. Hence forcing one component in veering conditions can result in dangerous vibrations of the other coupled component; on the other hand, the same effect is actually exploited to standard control techniques, see for instance [17,18].

4 Continuous Systems: Direct Perturbation of the Eigensystem

The results of the previous section apply to systems with two degrees of freedom; however, also when an infinite number of eigenvalue curves are involved as in the case of continuous systems, the veering of two eigenvalue curves is locally equivalent to a 2-dof veering: all the properties stated in the previous section still hold locally (i.e., in neighborhood of the veering region). Hence, when dealing with continuous systems, it is sufficient to detect the veering between two eigencurves and then to map the local behavior of these two curves into the 2-dof general solution through a suitable change of coordinates. To this aim one is led to project the involved spatial differential operators onto the eigenmodes obtained with vanishing coupling parameters and use a direct perturbation of the coupling parameters, as given by Perkins and Mote [12], to ascertain the presence of veering phenomena. The method introduced in [12] is here specialized to the case of continuous systems with gyroscopic operators of internal coupling; in the following section, the results are applied to the case of an electromechanical system with internal resonance.

Let $\mathcal{H}$ be a subspace of $L^2(\mathcal{D})$, the space of $\mathbb{R}^N$-valued square-integrable functions defined on a domain $\mathcal{D}$ and verifying suitable smoothness conditions on its boundary $\partial\mathcal{D}$; let u_h, $U_h \in \mathcal{H}$. This space is endowed with the usual L^2 inner product, i.e. $\langle u_h, u_k\rangle_{\mathcal{D}} := \int_{\mathcal{D}} u_h u_k$.

The following eigenvalue problem, function of a real perturbation parameter ϵ, is considered: find $u_h \in \mathcal{H}$ and $\lambda_h \in \mathbb{C}$ such that

$$\langle L(\epsilon)(u_h), u_k\rangle_{\mathcal{D}} + \langle \Xi(\epsilon)(u_h), (u_k)\rangle_{\partial\mathcal{D}} = \langle \lambda_h^2 M(\epsilon)(u_h) + \lambda_h R(\epsilon)(u_h), u_k\rangle_{\mathcal{D}} \tag{12}$$

for every $u_k \in \mathcal{H}$. Here

$$\begin{array}{llllll} M,L: & \mathbb{R}\to\mathcal{LS}_{\mathcal{D}} & R: & \mathbb{R}\to\mathcal{L}_{\mathcal{D}} & \Xi: & \mathbb{R}\to\mathcal{L}_{\partial\mathcal{D}} \\ & \epsilon\mapsto M(\epsilon), L(\epsilon) & & \epsilon\mapsto R(\epsilon) & & \epsilon\mapsto \Xi(\epsilon) \end{array} \tag{13}$$

are differentiable functions from the real axis $\mathbb{R}$ to the spaces $\mathcal{LS}_{\mathcal{D}}$ (linear self-adjoint operators on $\mathcal{H}$), $\mathcal{L}_{\mathcal{D}}$ (linear operators on $\mathcal{H}$), and $\mathcal{L}_{\partial\mathcal{D}}$ (linear operators on ∂D). Moreover let $R(0)=0$. As a consequence, the Taylor series of the involved operators around $\epsilon=0$ can be calculated getting

$$\begin{aligned} L(\epsilon) &= L(0) + \epsilon L'(0) + \epsilon^2 L''(0)/2 =: L_0 + \epsilon L_1 + \epsilon^2 L_2, \\ M(\epsilon) &= M(0) + \epsilon M'(0) + \epsilon^2 M''(0)/2 =: M_0 + \epsilon M_1 + \epsilon^2 M_2, \\ R(\epsilon) &= \epsilon R'(0) + \epsilon^2 R''(0)/2 =: \epsilon R_1 + \epsilon^2 R_2, \\ \Xi(\epsilon) &= \Xi(0) + \epsilon \Xi'(0) + \epsilon^2 \Xi''(0)/2 =: \Xi_0 + \epsilon \Xi_1 + \epsilon^2 \Xi_2. \end{aligned} \tag{14}$$

The unperturbed $\epsilon=0$ eigenvalue problem is then

$$\langle L_0(U_h), U_k\rangle_{\mathcal{D}} + \langle \Xi_0(U_h), U_k\rangle_{\partial\mathcal{D}} = \Lambda_h^2 \langle M_0(U_h), U_k\rangle_{\mathcal{D}}. \tag{15}$$

It follows that the unperturbed eigenfunctions satisfy the boundary conditions $\Xi_0(U_h)=0$, on $\partial\mathcal{D}$, and, since L_0 and M_0 are self-adjoint, can be orthogonalized according to $\langle M_0(U_h), U_k\rangle_{\mathcal{D}} = \delta_{hk}$. Moreover, due to the representation theorem, the set of the unperturbed eigenfunctions form a complete basis for $L^2(\mathcal{D})$; as a consequence the ansatz $u_h = c_{hk} U_k$ can be usefully considered; here a summation over the repeated index is understood.

In order to find the analytical relation between the eigenvalue solution and the perturbation parameter, λ_h and c_{hk} are expanded in series of ϵ

$$\begin{aligned} \lambda_h &= \lambda_{0h} + \epsilon\lambda_{1h} + \epsilon^2\lambda_{2h} + \cdots, \\ c_{hk} &= c_{0hk} + \epsilon c_{1hk} + \epsilon^2 c_{2hk} + \cdots. \end{aligned} \tag{16}$$

Substituting the considered ansatz into Eq. (12), a chain of algebraic problems arises. The solution at order 0 is $\lambda_{0h}=\Lambda_h$, and $c_{0hk}=\delta_{hk}$; the solution at order ϵ is

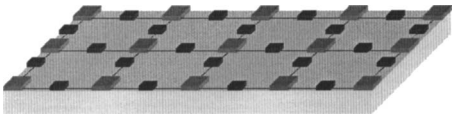


Fig. 4 Assembled plate and electric network

$$\lambda_{1h} = \frac{L_{1hh} - \Lambda_h^2 M_{1hh} - \Lambda_h R_{1hh}}{2\Lambda_h}, \qquad c_{1hk} = \begin{cases} 0, \text{ for } h = k, \\ \dfrac{\Lambda_h^2 M_{1hk} + \Lambda_h R_{1hk} - L_{1hk}}{\Lambda_k^2 - \Lambda_h^2}, \text{ for } h \neq k, \end{cases} \tag{17}$$

where

$$L_{1hk} := \langle L_1(U_h), U_k\rangle_{\mathcal{D}} + \langle \Xi_1(U_h), U_k\rangle_{\partial\mathcal{D}}, \qquad M_{1hk} := \langle M_1(U_h), U_k\rangle_{\mathcal{D}}, \quad R_{1hk} := \langle R_1(U_h), U_k\rangle_{\mathcal{D}}. \tag{18}$$

Finally, the solution at order ϵ^2 is

$$\lambda_{2h} = \frac{1}{2\Lambda_h}\left[\frac{(L_{1hh} + 3\Lambda_h^2 M_{1hh} + \Lambda_h R_{1hh})(\Lambda_h^2 M_{1hh} + \Lambda_h R_{1hh} - L_{1hh})}{4\Lambda_h^2} + L_{2hh} + \Lambda_h^2 M_{2hh} - \Lambda_h R_{2hh} + \sum_{h\neq k} \frac{(\Lambda_h^2 M_{1hk} + \Lambda_h R_{1hk} - L_{1hk})(\Lambda_h^2 M_{1kh} + \Lambda_h R_{1kh} - L_{1kh})}{\Lambda_h^2 - \Lambda_k^2} \right], \tag{19}$$

and

$$c_{2hk} = \begin{cases} 0 \text{ for } h = k, \\ \dfrac{1}{\Lambda_k^2 - \Lambda_h^2}\left[\Lambda_h^2 M_{2hk} + \Lambda_h R_{2hk} - L_{2hk} + \sum_{j\neq h} \times \dfrac{(L_{1jk} - \Lambda_h^2 M_{1jk} - \Lambda_h R_{1jk})(\Lambda_h^2 M_{1hj} + \Lambda_h R_{1hj} - L_{1hj})}{\Lambda_h^2 - \Lambda_j^2} + (L_{1hh} - \Lambda_h^2 M_{1hh} - \Lambda_h R_{1hh}) \times \left(M_{1hk} + \dfrac{R_{1hk}}{2\Lambda_h} + \dfrac{\Lambda_h^2 M_{1hk} + \Lambda_h R_{1hk} - L_{1hk}}{\Lambda_k^2 - \Lambda_h^2} \right) \right] \\ \text{for } h \neq k, \end{cases} \tag{20}$$

where

$$L_{2hk} := \langle L_2(U_h), U_k\rangle_{\mathcal{D}} + \langle \Xi_2(U_h), U_k\rangle_{\partial\mathcal{D}}, \qquad M_{2hk} := \langle M_2(U_h), U_k\rangle_{\mathcal{D}}, \quad R_{2hk} := \langle R_2(U_h), U_k\rangle_{\mathcal{D}}. \tag{21}$$

The interaction between two eigenmodes and the effects of the perturbation are described by the matrix:

$$\tilde{c}_{hk} := c_{hk} - \delta_{hk} = \epsilon c_{1hk} + \epsilon^2 c_{2hk}. \tag{22}$$

Recalling the assumed ansatz, $u_h = c_{hk} U_k$, a nonvanishing $\tilde{c}_{12}$ means a nonvanishing projection of the perturbed first mode on the second unperturbed mode. This implies an energetic exchange, proportional to the perturbation, in the time evolution of the U_1 and U_2 components.

5 An Electromechanical System With Gyroscopic Coupling

The results obtained in Sec. 4 are used to derive the qualitative behavior of an electromechanical system under two different boundary conditions. The system, presented in [19], is a Kirchhoff plate coupled with a set of distributed piezoelectric actuators interconnected through an electric transmission net. Figure 4 shows

Table 1 Plate modes labeling

	$h=1,2$	3, 4	5, 6	7, 8	9, 10	11, 12	13, 14	15, 16	17, 18
i_h	1	1	2	2	1	3	2	3	3
j_h	1	2	1	2	3	1	3	2	3

a sketch of the assembled structure: the gray and black boxes respectively represent the piezoelectric patches and the electric impedances among them.

The plate deformation is described through the transverse displacement v of the plate mid-plane, while the electric state in the transmission net through the time-primitive ϕ of the electric potential.

The efficiency of this system is based on a veering condition; indeed the maximum exchange of energy needed to achieve an efficient control action is obtained tuning the electric eigenfrequencies to the values of the mechanical ones (see Sec. 2). The system free oscillations are governed by the following set of linear partial differential equations:

$$\alpha\Delta\Delta v + \ddot{v} - \gamma\Delta\dot{\phi} = 0,$$

$$-\beta\Delta\phi + \ddot{\phi} + \gamma\Delta\dot{v} + \delta\dot{\phi} + \delta\gamma\Delta v = 0, \tag{23}$$

where α, β, γ, and δ are dimensionless parameters, Δ is the Laplacian operator, and $\Delta\Delta$ the double Laplacian operator related to the Kirchhoff model of the plate. The piezoelectric effect induces a gyroscopic coupling between the standard Kirchhoff plate equation and the standard Laplace equation for the electric net.

In this case $\mathcal{D}$ is the rectangular domain $[0,1]\times[0,d]$ and the functions u and U, defined in Sec. 4, are $\mathbb{R}^2$-valued square integrable functions on $\mathcal{D}$. Moreover the coupling γ and the electric dissipation δ are supposed to be of the same order of magnitude setting $\gamma=\epsilon$ and $\delta=\epsilon\varphi$. In view of Eq. (12) the following differential operators $L(\epsilon)$, $M(\epsilon)$ and $R(\epsilon)$ are obtained:

$$L(\epsilon) := \begin{pmatrix} \alpha\Delta\Delta & 0 \\ \epsilon^2\varphi\Delta & -\beta\Delta \end{pmatrix}, \quad M(\epsilon) := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad R(\epsilon) := \epsilon\begin{pmatrix} 0 & -\Delta \\ \Delta & \varphi \end{pmatrix}. \tag{24}$$

5.1 Simply Supported Boundary Conditions. The plate is considered to be simply supported and electrically grounded at the boundary so that

$$\Xi(\epsilon) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \dfrac{\partial^2}{\partial x^2} & 0 \end{pmatrix} \quad \text{for } y = 0, d;$$

$$\Xi(\epsilon) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \dfrac{\partial^2}{\partial y^2} & 0 \end{pmatrix} \quad \text{for } x = 0, 1. \tag{25}$$

The order 0 solution is easily found as follows

$$U_h = e_0 \sin(i_h\pi x)\sin(j_h\pi y/d), \quad \Lambda_h^{(m)} = \pi^4\sqrt{\alpha}\left(i_h^2 + \frac{j_h^2}{d^2}\right)^2,$$

$$h = 1,3,5,\ldots,$$

$$U_{h+1} = e_1 \sin(i_h\pi x)\sin(j_h\pi y/d), \Lambda_h^{(e)} = \pi^2\sqrt{\beta}\left(i_h^2 + \frac{j_h^2}{d^2}\right), \tag{26}$$

where $e_0=\{1,0\}$ and $e_1=\{0,1\}$ form the canonical basis in $\mathbb{R}_2$; e_0 and e_1, respectively, mean a totally mechanical and electric nature of the eigenfunction. Hence for these boundary conditions (25) the mechanical and electrical eigen-functions share the same spatial modal shapes at order 0. The modes are labeled according to Table 1.

The odd values of the index h are therefore related to mechanical modes while the even are related to electrical modes.

Since Ξ and M do not depend on the perturbation parameter, the matrices Ξ_1, M_1, Ξ_2 and M_2 vanish; moreover L_1 and R_2 vanish because $L'(0)=0$ and $R''(0)=0$. Hence the coupling is governed by the operators R_1 and L_2.

Now consider two arbitrary modes, say h and k, of the system (23).

(a) If the modes are both mechanical (h and k are both odd), then there is no interaction since both R_{1hk} and L_{2hk} vanish.

(b) If the modes are both electrical, then $R_{1hk}=\rho\delta_{hk}$, $\rho=d\varphi/4$, and from Eqs. (17), (19), and (20)

$$\lambda_h = \Lambda_h - \epsilon\frac{\rho}{2} + \epsilon^2\frac{\rho_2}{4\Lambda_h}, \quad c_{hk} = \delta_{hk}; \tag{27}$$

thus the two eigenvectors remain unchanged, while the eigenvalues are modified by the presence of the dissipation.

(c) If one mode is mechanical and one electrical, there is an interaction only if $k=h+1$ or in other words if the modes share the same spatial modal forms; otherwise the coupling between the two modes is again vanishing. Set $h\to m$ and $k=h+1\to e$; then

$$R_{1me} = \langle R_1(U_m), U_e\rangle_{\mathcal{D}} = -\frac{\pi^2 d}{4}\left(i_h^2 + \frac{j_h^2}{d^2}\right), \quad R_{1em} = -R_{1me}, \tag{28}$$

$$L_{2me} = \langle L_2(U_m), U_e\rangle_{\mathcal{D}} = \frac{\pi^2 d\varphi}{4}\left(i_h^2 + \frac{j_h^2}{d^2}\right), \quad L_{2em} = 0. \tag{29}$$

Substituting these results in Eqs. (17) and (20), the dependence of the eigenvectors on the coupling parameter ϵ is derived up to the second order:

$$\tilde{c}_{me} = \frac{1}{\Lambda_e^2 - \Lambda_m^2}(\epsilon\Lambda_m R_{1me} - \epsilon^2 L_{2me}), \quad \tilde{c}_{em} = \frac{1}{\Lambda_e^2 - \Lambda_m^2}(\epsilon\Lambda_e R_{1me}). \tag{30}$$

This proves that the two eigenvectors are rotated by the perturbation, especially when the $\Lambda_e^2-\Lambda_m^2$ is small. This rotation with respect to the unperturbed eigensolutions signifies that the perturbed eigenvectors have both electric and mechanical components; for important physical implications of this fact the reader can refer to [19,20].

The singularity for $\Lambda_e\to\Lambda_m$ is due to the use of a straightforward perturbation and can be avoided using more complex techniques, such as a modified version of this method as in [9] or a matched asymptotic expansion as in [11]. However, since the interest is only devoted in detecting the veering, this complexity has been intentionally avoided; moreover the local behavior of eigen-

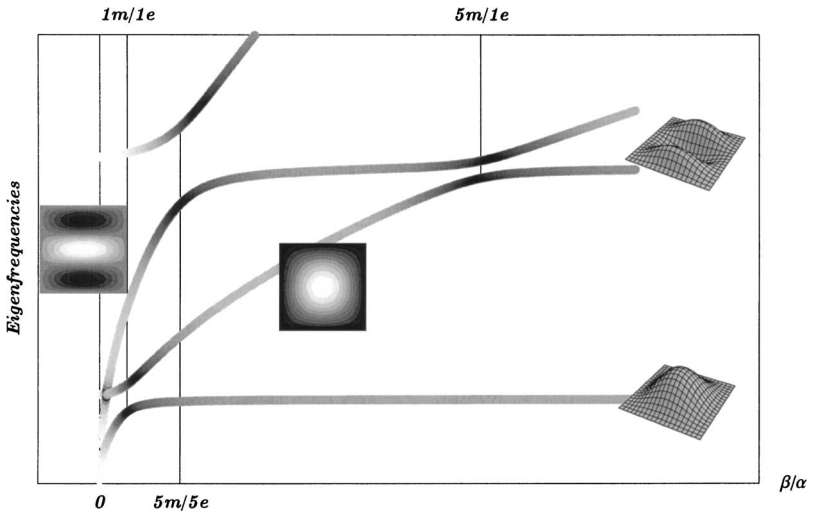


Fig. 5 The eigenvalues for h=1, 5 in the clamped case of boundary conditions. Darker regions indicate veering phenomena.

frequencies and eigenvectors is locally equivalent to the 2-dof case which can be analytically described as in Sec. 2.

5.2 Clamped Boundary Conditions. The plate is considered to be clamped on all sizes and electrically grounded at the boundary so that

$$\Xi(\epsilon)=\begin{pmatrix}1 & 0\\ 0 & 1\\ \dfrac{\partial}{\partial x} & 0\end{pmatrix} \quad \text{for } y=0,d;$$
$$\Xi(\epsilon)=\begin{pmatrix}1 & 0\\ 0 & 1\\ \dfrac{\partial}{\partial y} & 0\end{pmatrix} \quad \text{for } x=0,1. \tag{31}$$

The zero-order solution is now given by

$$U_h=e_0 f_h(x,y), \quad \Lambda_h^{(m)}=\sqrt{\alpha}\varsigma_h,$$
$$h=1,3,5,\ldots,$$
$$U_{h+1}=e_1\sin(i_h\pi x)\sin(j_h\pi y/d), \quad \Lambda_h^{(e)}=\pi^2\sqrt{\beta}\left(i_h^2+\frac{j_h^2}{d^2}\right), \tag{32}$$

where $f_h(x,y)$ are the eigenfunctions of the double Laplacian on a plate with clamped sizes and ς_h are the associated eigenvalues.

If pairs of purely electrical or purely mechanical modes are considered then, as before, the zero-order eigenvectors are not modified [cases (a) and (b) in Sec. 5.1]. On the contrary, if the two considered modes (h and k) are mechanical and electrical, the coupling does not vanish only if $k=h+1$ as before. Indeed, since the electric and mechanical eigenfunctions do not have the same spatial modal shapes, the matrix $\tilde{c}_{hk}$ has many nonvanishing entries for each row. This means that the mode to mode interaction is not anymore one-to-one. Thus each mechanical or electrical mode has a nonvanishing coupling with several other modes and each branch of the root locus presents several veerings with the other branches. In Fig. 5 the eigenvalues curves are represented relative to two sets of strongly coupled modes namely the first and fifth modes of the clamped plate—displayed through three-dimensional plots—and the first and fifth modes of the electrically grounded net—displayed through contourplots.

The gray-level intensity of the curves in Fig. 5 is proportional to the ratio between the out-of-diagonal and in-diagonal terms in the eigenvector matrix: a veering is hence highlighted by a darker region. Three evident veerings are present in Fig. 5 as the parameter β/α is varied: between the first mechanical and electrical modes (1m/1e), the fifth mechanical and electrical modes (5m/5e), but also between the fifth mechanical and the first electric modes (5m/1e). A fourth veering is hardly seen between the curves 1m and 5e.

Since in veering conditions an initial amount of mechanical energy transforms back and forth into its electric form, a suitable time-dependent tuning of the frequency ratio β/α allows for energetic exchanges not only between mechanical and electric modes but also between two arbitrary mechanical modes provided they are coupled with the same set of electric modes. This fact does not contradict the aforementioned assertion on the coupling of purely mechanical modes, valid only for a fixed value of the frequency ratio.

Suppose, indeed, that the initial energy is given only in the first mechanical mode; if the frequency ratio β/α allows veering with the first electric mode (1m/1e), after a certain time the energy is stored only in this last mode. Now shift the frequency ratio β/α to the value 5m/1e; the energy flows in the fifth mechanical mode; an additional shift into a region with no veering can confine the energy into the fifth mechanical modes. It is worth noticing that, from a practical point of view, the frequency shifts are easily achieved by an electric tuning of a simple potentiometer.

6 Concluding Remarks

The main geometrical properties and dynamical implications of the veering between two eigenvalue curves have been studied considering an arbitrary elastic, inertial and gyroscopic coupling. Through the use of the ellipsoid associated to positive definite symmetric eigenproblems it is shown that in order to avoid a crossing point a condition of maximum coupling must be achieved.

The case of two degrees of freedom systems is then analytically solved; as a matter of fact the veering of two eigenvalue branches is locally equivalent to the 2-dof case, also when an infinite number of degrees of freedom is involved, as in continuous systems. Depending on the amplitude of the veering region, it is sometimes difficult to distinguish a veering from a crossing using only the eigensolution in a discrete number of parameter values. Mean-

while, this experimental difficulty can be solved observing the forced undamped response of the system since in this case its response, in the presence of veering, sensibly differs from the uncoupled case (crossing).

Veering governs the coupling between the involved modes: when these modes are spatially global, localization occurs in the veering region; on the contrary when these modes are spatially localized in the veering region global motion arises.

The direct perturbation technique [12] has been applied to veering detection in an electromechanical system governed by some linear PDEs with a gyroscopic coupling operator. Considering the coupling coefficient as perturbative parameter and using, as functional basis, the uncoupled eigenmodes, the presence of veering has been ascertained and the main dynamical features of the system caught. It is shown how a multiple veering allows for energetic exchanges between uncoupled modes, achieved through a simple time-dependent tuning of a resistance. This phenomenon could have useful applications in several disciplines.

Acknowledgment

This work has been partially supported by MURST-COFIN 01-02 (FENDIS project, http://www.disg.uniromal.it/fendis/).

References

[1] Leissa, A. W., 1974, "On a Curve Veering Abberation," J. Appl. Math. Phys. (ZAMP), **25**, pp. 99–111.
[2] Doll, R. W., and Mote, C. D., 1976, "On the Dynamics Analysis of Curved and Twisted Cylinders Transporting Fluids," J. Pressure Vessel Technol., **98**, pp. 143–149.
[3] Ramaswamy, R., and Marcus, R. A., 1981, "Perturbative Examination of Avoided Crossings," J. Chem. Phys., **72**, pp. 1379–1384.
[4] Nair, P. S., and Durvasula, S., 1973, "On Quasi Degeneracies in Plate Vibration Problems," Int. J. Mech. Sci., **15**, pp. 975–986.
[5] Triantafyllou, M. S., 1984, "The Dynamics of Taut Inclined Cables," Q. J. Mech. Appl. Math., **37**, pp. 421–440.
[6] Russel, J. C., and Lardner T. J., 1998, "Experimental Determination of Frequencies and Tension for Elastic Cables," J. Eng. Mech., **124**, pp. 1067–1072.
[7] Behbahani-Nejad, M., and Perkins, N. C., 1996, "Freely Propagating Waves in Elastic Cables," J. Sound Vib., **196**, pp. 189–202.
[8] Cheng, S. P., and Perkins, N. C., 1992, "Closed Form Vibration Analysis of Sagged Cable/mass Suspensions," J. Appl. Mech., **59**, pp. 923–928.
[9] Pierre, C., 1988, "Mode Localization and Eigenvalues Loci Veering Phenomena in Disordered Structures," J. Sound Vib., **126** (3), pp. 485–502.
[10] Triantafyllou, M. S., and Triantafyllou, G. S., 1991, "Frequency Coalescence and Mode Localization Phenomena: A Geometric Theory," J. Sound Vib., **150** (3), pp. 485–500.
[11] Natsiavas, S., 1993, "Mode Localization and Frequency Veering in a Nonconservative System With Dissimilar Components," J. Sound Vib., **165**, pp. 137–147.
[12] Perkins, N. C., and Mote, C. D., 1986, "Comments on Curve Veering in Eigenvalues Problems," J. Sound Vib., **106**, pp. 451–463.
[13] Pierre, C., and Dowell, E. H., 1987, "Localization of Vibrations by Structural Irregularity," J. Sound Vib., **114**, pp. 549–564.
[14] Happawana, G. S., Bajaj, A. K., and Nwokah, O. D. I., 1993, "A Singular Perturbation Analysis for Eigenvalue Veering and Modal Sensitivity in Perturbed Linear Periodic Systems," J. Sound Vib., **160**, pp. 225–242.
[15] Arnold, V. I., 1989, *Mathematical Method of Classical Mechanics*, 2nd ed., Springer-Verlag, New York.
[16] Meirovitch, L., 1974, "A New Method of Solution of the Eigenvalue Problem for Gyroscopic Systems," AIAA J. **12** (10), pp. 1337–1342.
[17] den Hartog, J. P., 1934, *Mechanical Vibrations*, 4th ed., McGraw-Hill, New York.
[18] Lacarbonara, W., and Vestroni, F., 2002, "Feasibility of a Vibration Absorber Based on Hysteresis," Proceedings of the Third World Conference on Structural Control, Como, Italy.
[19] Vidoli, S., and dell'Isola, F., 2001, "Vibration Control in Plates by Uniformly Distributed Actuators Interconnected via Electric Networks," Eur. J. Mech. A/Solids, **20**, pp. 435–456.
[20] dell'Isola, F., Vestroni, F., and Vidoli, S., 2002, "A Class of Electromechanical Systems: Linear and Nonlinear Mechanics," J. Theoretical Appl. Mech., **40** (1), pp. 47–71.

Scaling Laws From Statistical Data and Dimensional Analysis

Patricio F. Mendez
Department of Metallurgical and Materials Engineering,
Colorado School of Mines,
Golden, CO 80401
e-mail: pmendez@mines.edu

Fernando Ordóñez
Industrial and Systems Engineering,
University of Southern California,
Los Angeles, CA 90089
e-mail: fordon@usc.edu

Scaling laws provide a simple yet meaningful representation of the dominant factors of complex engineering systems, and thus are well suited to guide engineering design. Current methods to obtain useful models of complex engineering systems are typically ad hoc, tedious, and time consuming. Here, we present an algorithm that obtains a scaling law in the form of a power law from experimental data (including simulated experiments). The proposed algorithm integrates dimensional analysis into the backward elimination procedure of multivariate linear regressions. In addition to the scaling laws, the algorithm returns a set of dimensionless groups ranked by relevance. We apply the algorithm to three examples, in each obtaining the scaling law that describes the system with minimal user input. [DOI: 10.1115/1.1943434]

1 Introduction

In engineering design, we are constantly faced with the need to describe the behavior of complex engineered systems for which there is no closed-form solution or exhaustive analysis. This usually leads to a tedious, time consuming, and detailed study of the engineering process in question, delaying the overall process of design and limiting the total number of possibilities that can be investigated. For example, the design of a welding procedure involves so many parameters that it can seldom be predicted reliably; therefore, extensive experimentation must take place in order to determine an ideal process setup. A set of simple and intuitive design laws based only on the most relevant parameters would be of enormous help in this case. Scaling laws in the form of power laws, which we will simply call scaling laws, are particularly well suited for this purpose.

Scaling laws are ubiquitous in engineering. In fact, they have been used to explain the behavior of many physical, biological (e.g., [1,2]), psychophysical [3], geophysical (e.g., [4,5]), Internet traffic [6], and even economic systems [7]. A broad sample of problems that can be described with such scaling laws is presented in [8]. Segel [9] provides a good overview of simplification and scaling. Some reasons for the wide applicability of power law models in engineering are: (i) the combination of units has the form of a power law, (ii) the expressions of many physical phenomena have the form of power laws as noted above, and (iii) many empirical regressions of engineering data in log-log plots tend to give a straight line, which corresponds to a power law.

Scaling laws are of enormous utility during the early stages of design, when the configuration of a system and the choice of materials are still uncertain. In this case, they provide quick estimations of the feasibility of a design, help determine optimal sizes, and contribute to decisions about configuration and materials. These laws are also useful for control systems and for decision-making algorithms, predicting the behavior of a system much faster than computationally intensive models such as finite element analysis or computational fluid mechanics. The design and interpretation of physical models of reduced size, such as reduced-scale aircraft in wind tunnels, are based on scaling laws. When experimental databases or numerical models exist, scaling laws can be used to generalize and extrapolate the results obtained. For existing machines, scaling laws are useful for setup and tuning operations.

Our work aims at facilitating the process of engineering design by providing a computational tool that derives the best power law from experimental data. We propose the algorithm SLAW (Scaling LAWs). This algorithm combines a linear regression model of the experimental data with physical considerations of the process, namely, that the units of the resulting model match the units of the dependent variable. We look for the power law model that minimizes the prediction error only among models that have the correct units. The output of the algorithm is a physically meaningful and simple power law, representing the process and a set of dimensionless groups ordered by their relevance to the problem. The user input in selecting the simple model, and the ability to correct it further using the dimensionless groups, provide the means to construct a model that achieves the desired balance between accuracy and simplicity. An early version of this algorithm was presented in [10]. SLAW grew from that version by incorporating: a rounded model output, where all coefficients are fractions typical in the equations in physics and engineering; user input to select the best scaling law; and ranked dimensionless groups that explain the residual error. In this work, we call "model" any scaling law that captures the main behavioral trends of a process as a function of the parameters of the problem; this should not be confused with other ways of representing the problem, such as finite element models.

The SLAW algorithm differs from classical dimensional analysis in that it selects the scaling law with the smallest predictive error out of all the dimensionally correct models. There are computational implementations of dimensional analysis, such as that described by Kasprzak et al. [11], which are able to construct dimensionally correct models and check the completeness of the set of variables. SLAW differs from these approaches in the systematic search for an expression involving the smallest acceptable number of variables.

SLAW also differs from other statistical simplifications, such as the principal directions of the matrix of correlation, in that these other approaches, besides not necessarily providing the correct units, reduce the mathematical complexity of the problem but still consider all physical parameters, regardless of their importance. Previous works that combines linear regressions and dimensional analysis, such as landmark work by Vignaux and Scott [12], Vignaux [13], Li and Lee [14], and Dovi et al. [15], all use dimensionless groups determined a priori. In contrast, SLAW automatically generates the ranked dimensionless groups.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, November 14, 2003; final revision, November 26, 2004. Associate Editor: K. M. Liechti. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

The automatic determination of formulas, such as scaling laws, from data has been an active area of research in the artificial intelligence community. Important examples of this research are the algorithm BACON due to pioneering work by Bradshaw et al. [16], algorithms ABACUS [3,17] and COPER [17], and recent work by Washio and Motoda [3,18]. Important differences that the algorithm proposed here has with both BACON and ABACUS are that these latter algorithms produce only scaling laws with integer exponents and require that the data vary one variable at a time. The algorithm COPER and the work by Washio and Motoda differ from the current work in that they do not explicitly construct a scaling law that is the simplest with respect to some criterion, and the dimensionless groups obtained are not necessarily ranked by relevance to the model.

In Sec. 2, we describe the assumptions made on the physical process and the main ideas in the methodology. In Sec. 3, we present the algorithm SLAW. We apply this methodology to three different examples: a pendulum, ceramic-to-metal joining, and the "punch test" (a standard test used to determine mechanical properties of materials). We describe the examples and present the results obtained with our algorithm in Sec. 4. In Sec. 5, we present the conclusions of this work.

2 Methodology

2.1 Scope of Methodology. To illustrate this discussion, consider the problem of joining a ceramic cylinder to a metallic cylinder, as pictured in Fig. 1.

One of the quantities of interest in this problem is the volumetric strain energy in the ceramic, which can be expressed as

$$u(X,Z) = Y(X)u^*(X,Z) \tag{1}$$

Here, $u(X,Z)$ is the volumetric strain energy at each point in the ceramic and depends on the *problem parameters* X and other variables Z. In this example, problem parameters X can include the radius of the cylinders r, the yield strength of the metal σ_Y, and the elastic modulus of the materials E_c and E_m, while other variables Z typically represent the space and time coordinates of a point in the cylinders. The quantity Y is a characteristic value that only depends on problem parameters and has the same units as u. The function u^* is a dimensionless function that shows the volumetric strain energy variations relative to the characteristic value.

Our objective in this paper is to obtain a simple yet meaningful expression for Y from experimental data, as the characteristic value captures important trends of the quantity of interest. For example, in Sec. 4 we show that the characteristic value of the volumetric strain energy in ceramic-to-metal joining as a function of the parameters is $Y=\sigma_Y^2 r^3/E_c$. To use a nomenclature that is consistent with linear regression literature, we will also refer to the characteristic value Y as the *dependent variable* and to the parameters X as the *independent variables*.

Dimensional analysis states that the characteristic value Y can be decomposed in one power law expression with the same units of u and a function f of m dimensionless groups $\Pi_1,\ldots,\Pi_m$, which also have a power law expression as a function of the parameters. If we assume that there are n parameters for the problem $X_1,\ldots,X_n$, this means that

$$Y = a_0 \prod_{j=1}^{n} X_j^{a_{0j}} f(\Pi_1,\ldots,\Pi_m) \tag{2}$$

with $\Pi_i = a_i \Pi_{j=1}^{n} X_j^{a_{ij}}$. For simplicity, we will assume that the function f can be approximated by a power law. This is reasonable, given that most of the behavior of the dependent variable is typically captured by the power law expression and f shows small, smooth, and monotonic variations within a regime. Therefore, we can write Eq. (2) as

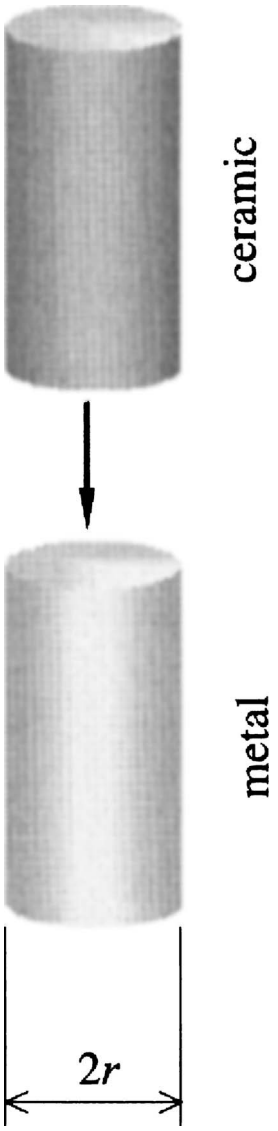


Fig. 1 Geometry of the ceramic and metal parts to be joined

$$Y = a_0 \prod_{j=1}^{n} X_j^{a_{0j}} \prod_{i=1}^{m} \left(a_j \prod_{j=1}^{n} X_j^{a_{ij}} \right) \tag{3}$$

$$= a \prod_{j=1}^{n} X_j^{\sum_{i=0}^{m} a_{ij}} \tag{4}$$

with $a = \Pi_{i=0}^{m} a_i$. Equation (4) is a valid model for the problem under two additional assumptions on the system, which are standard assumptions in dimensional analysis: (i) We assume that at least all parameters that determine the problem are considered. This assumption is also necessary in the analysis of regressions. Omitting a relevant parameter can result in ignoring a dominant effect and, thus, missing the correct model. Considering more parameters than are strictly necessary is not a problem, since SLAW can efficiently discard the less relevant parameters. (ii) We assume that the physical system is studied under a single regime. This means that the same physical factors (there is no need to know exactly which) are dominant for all of the observations used to build the input data set. This implies that f is of the order of magnitude of 1.

An important observation, which is enforced explicitly in the algorithm, is that the model given by Eq. (4) has the same units as the characteristic value Y. This additional constraint, denoted the *units constraint*, is expressed by

$$\text{units of } Y = \prod_{j=1}^{n} (\text{units of } X_j)^{\Sigma_{i=0}^{m} a_{ji}} \tag{5}$$

For this approach to be well defined, it is necessary that some combination of the parameters considered can yield the units of the characteristic value, and thus it is possible to satisfy the units constraint.

2.2 Constrained Linear Regression. For any application, the model postulated is faced with uncertainties that arise, for example, from working with experimental data or considering only n independent variables and disregarding the possibly tiny effect of other variables. By taking the logarithm of Eq. (4), and considering the existing uncertainties in the model, we can express the model as

$$\log Y = \beta_0 + \sum_{j=1}^{n} \beta_j \log X_j + \varepsilon \tag{6}$$

where the coefficients are $\beta_0 = \log a$ and $\beta_j = \Sigma_{i=0}^{m} a_{ij}$, and ε is an error term that captures the model uncertainties. Additive errors in logarithms of measurements is a common assumption in fitting scaling laws, see [15,19], and it is justified by Benford's law [20,21], which states that variations of physical quantities are evenly distributed in a logarithmic scale.

Considering p experimental observations of the physical process, we obtain estimators of the coefficients of Eq. (6) using standard linear regression machinery. We denote the p observations of the dependent variable Y by $y_1, \ldots, y_p$, and the observations for the jth independent variable X_j by $x_{1j}, \ldots, x_{pj}$. We assume independent experimental observations, which implies that observed errors $\varepsilon_1, \ldots, \varepsilon_p$ are independent identically distributed (IID) random variables, where $\varepsilon_i = \log y_i - \beta_0 - \Sigma_{j=1}^{n} \beta_j \log x_{ij}$. Using matrix notation, we have

$$\tilde{y} = \begin{pmatrix} \log y_1 \\ \vdots \\ \log y_p \end{pmatrix}, \quad \text{and } \tilde{X} = \begin{bmatrix} 1 & \log x_{11} & \cdots & \log x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \log x_{p1} & \cdots & \log x_{pn} \end{bmatrix}$$

The estimate for the coefficients in model (6) that minimizes the residual sum of squares is the solution to the system of normal equations $\tilde{X}^T \tilde{X} \beta = \tilde{X} \tilde{y}$, where the superscript T denotes the transpose of a matrix. We denote this estimate by the $n+1$ dimensional vector $\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_n)$, and the estimate of the independent variable becomes $\hat{Y} = e^{\hat{\beta}_0} \Pi_{j=1}^{n} X_j^{\hat{\beta}_j}$.

The estimate $\hat{\beta}$, however, will generally not satisfy the units constraint. Therefore, we have to select the coefficients that minimize the residual sum of squares only among those that satisfy Eq. (5), which as we explain below is equivalent to a linear constraint of the form $R\beta = b$. With this additional constraint, the estimate of the coefficients in model (6) that minimizes the residual sum of squares and satisfies the units constraint is the solution to the problem

$$\begin{aligned} &\min_{\beta} (\tilde{y} - \tilde{X}\beta)^T (\tilde{y} - \tilde{X}\beta) \\ &\text{s.t. } R\beta = b \end{aligned} \tag{7}$$

To represent the units constraint in linear form, assume that q reference units (m, kg, s,…) are the building blocks for the units of the dependent and all independent variables in the problem. The units constraint can be expressed by $R\beta = b$, where b is a q-dimensional vector such that b_i is the exponent of reference unit i in the units of the dependent variable Y, and matrix R is q by $n+1$ such that R_{ij} is the exponent of reference unit i in the units of variable X_j, for $j = 0, 1, \ldots, n$. Note that we include a variable X_0, which accounts for the constant dimensionless term, thus $R_{i0} = 0$ for all i. For example, this notation implies that $R\bar{a} = b$, for $\bar{a}^T = (a_0, a_{01}, \ldots, a_{0n})$.

2.3 Generation of Dimensionless Groups. The solution to problem (7), denoted with the vector $\beta^0 = (\beta_0^0, \beta_1^0, \ldots, \beta_n^0)$, estimates the coefficients that construct the model in Eq. (4). This solution, however, can have all coefficients different from zero, which leads to a model that, although satisfies the units constraint and has minimal residual sum of squares, is not simple as it includes all independent variables and lacks physical interpretation. We aim to separate this complicated model into a simpler power law and dimensionless groups, as in Eq. (3). For this, consider $\tilde{\beta}$ another $n+1$ dimensional vector that satisfies the units constraint and is simpler, i.e., has few nonzero coefficients, and let $\delta = \beta^0 - \tilde{\beta}$. Using this decomposition, we express the model with β^0 in the following form, in line with Eq. (3):

$$\hat{Y} = e^{\beta_0^0} \prod_{j=1}^{n} X_j^{\beta_j^0} = e^{\tilde{\beta}_0} \prod_{j=1}^{n} X_j^{\tilde{\beta}_j} \left(e^{\delta_0} \prod_{j=1}^{n} X_j^{\delta_j} \right)$$

The first factor in the right side of this equation corresponds to a simple power law as $\tilde{\beta}$ has few nonzero coefficients and, since $R\delta = 0$ by construction, the second corresponds to a dimensionless group.

The proposed algorithm simplifies the model, and in the process identifies the dimensionless groups by removing independent variables from the model one at a time. A variable X_j is removed from the model by forcing the solution to satisfy $\beta_j = 0$, which is enforced by the linear constraint $e_j^T \beta = 0$, where e_j is the jth canonical vector in $n+1$ dimensions. After k iterations of the algorithm, exactly k independent variables, let us say $X_{i_1}, \ldots, X_{i_k}$, have been removed from the model. Therefore, feasible models now must also satisfy $e_{i_1}^T \beta = 0, \ldots, e_{i_k}^T \beta = 0$, which can be written in matrix form as $M_k \beta = 0$ for the k by $n+1$ matrix $M_k = [e_{i_1}, \ldots, e_{i_k}]^T$. For each iteration k, let $\beta^k = (\beta_0^k, \ldots, \beta_n^k)$ be the solution to Eq. (7) that satisfies, in addition, $M_k \beta = 0$. To remove an additional variable X_j with $\beta_j^k \neq 0$ from the model, we simply add the constraint $\beta_j = 0$ to the problem. Therefore, the reduced model is given as the solution to the optimization problem

$$\begin{aligned} z_k^*(j) = &\min_{\beta} (\tilde{y} - \tilde{X}\beta)^T (\tilde{y} - \tilde{X}\beta) \\ &\text{s.t. } R\beta = b \\ &M_k \beta = 0 \\ &\beta_j = 0 \end{aligned} \tag{8}$$

Equation (8) minimizes a strictly convex function over linear constraints and therefore has a unique solution that can be computed solving the first-order optimality conditions, which for this problem are a linear system of equation [22].

From all variables with $\beta_j^k \neq 0$, we eliminate at iteration $k+1$ the variable i_{k+1}, which makes $z_k^*(j)$ smallest. The resulting model, with coefficients β^{k+1}, best fits the given data in a least-squares sense, satisfies the units constraint, and has $k+1$ coordinates equal to zero. This last constraint is encoded in the matrix $M_{k+1}^T = [M_k^T e_{i_{k+1}}]$. After eliminating k parameters from the model, we obtain a simplified model β^k and the dimensionless groups $\delta^i = \beta^{i-1} - \beta^i$ for $i = 1, \ldots, k$. These vectors satisfy $\beta^0 = \Sigma_{i=1}^{k} \delta^i + \beta^k$, which implies the following expression:

$$\hat{Y} = e^{\beta_0^k} \prod_{j=1}^{n} X_j^{\beta_j^k} \left[\prod_{i=1}^{n} \left(e^{\delta_0^i} \prod_{j=1}^{n} X_j^{\delta_j^i} \right) \right]$$

The process of sequentially eliminating variables is possible while the linear system of constraints has a solution. We show below that the dimensionless groups generated δ^k are linearly independent; therefore, the number of iterations can be at most $n+1-\text{rank}(R)$.

To see that $\delta^1, \ldots, \delta^K$ are linearly independent, assume without loss of generality that coordinate k is eliminated to construct β^k, then by construction we have that $\delta_j^k = 0$ for all $j < k$ and $\delta_k^k \neq 0$, thus the matrix $[\delta^1, \ldots, \delta^K]$ is lower triangular with nonzero diagonal elements, which implies the linear independence. In addition we know there are $\text{rank}(R)$ linearly independent rows of R, and thus the vector space orthogonal to the rows of R is of dimension $n+1-\text{rank}(R)$. Since all δ^k are orthogonal to the rows of R and linearly independent, there can be at most $n+1-\text{rank}(R)$ of them.

The process of generating solutions β^k, each with k variables removed, can be interpreted as backward elimination for linear regression, see [23]. Here, the backward elimination is applied to a linearly constrained linear regression problem.

3 Algorithm SLAW

The algorithm SLAW (Scaling LAWs) uses experimental data, information regarding the units of variables, and information about the accuracy needed. It outputs a physically meaningful simple model and a set of dimensionless groups that explain the dependent variable in order of importance.

This algorithm can be broken down into four steps:

1. Find the sequence of models $\{\beta^k\}$ through the backward elimination process that solves Eq. (8).
2. Determine, with user input, which model of the sequence $\{\beta^k\}$ to select, say its $\hat{\beta}$.
3. Round the coefficients in $\hat{\beta}$, obtaining a physically meaningful simple model β^*.
4. Perform backward elimination again, to identify dimensionless groups in what is not explained by β^*.

Step 4 is needed to identify the correct dimensionless groups because the rounding procedure in Step 3 creates a model β^* that is slightly different from the model derived in the original regressions. This second application of backward elimination has a significant difference from the first. While the goal of the first application of backward elimination is to find the smallest meaningful scaling law, we start the second set of iterations with a scaling law already in place, β^*. Although the change in scaling law is typically small, the dimensionless groups corresponding to the new law are not necessarily the same, and we perform another backward elimination to find them. The goal of this second backward elimination is simply to find a scaling law for the error between the input data and the rounded scaling law, characterized by $\delta = \beta^0 - \beta^*$. We now describe each step of the SLAW algorithm.

3.1 Model Reduction Algorithm. The following algorithm identifies the dimensionless groups of the linear regression model in order of significance to the dependent variable. The inputs are the experimental data and units constraint data; the output are the sequence of estimators $\{\beta^k\}$ and dimensionless groups $\{\delta^k\}$.

—Algorithm MODEL REDUCTION $(\tilde{X}, \tilde{y}, R, b)$:

Step 1: Solve (7), let β^0 be the solution. $k=0$.

Step 2: Find the coordinate j_k that minimizes $z_k^*(j)$ [problem (8)]. Let β^{k+1} be the solution to $z_k^*(j_k)$. Let $\delta^{k+1} = \beta^k - \beta^{k+1}$.

Step 3: Let $k=k+1$. Repeat Step 2 while some variable can be eliminated from β^k.

—Output: $\{\beta^k\}_{k=0}^{n-\text{rank}(R)}$ and $\{\delta^k\}_{k=1}^{n-\text{rank}(R)}$.

3.2 Selecting Model. By definition, β^0 defines the model that best explains the dependent variable out of the sequence generated by the DIMENSION REDUCTION algorithm. This solution has the smallest residual sum of squares (RSS) but is not a simple solution, as it uses all n variables and constant term. In contrast, $\beta^{n-\text{rank}(R)}$ is the simplest solution that satisfies the units constraints and at the same time has the largest RSS. User input is used to select a model that balances simplicity with accuracy.

The inputs are the experimental data, the sequence of estimators $\{\beta^k\}$, and a user-supplied tolerance TOL. The output is the estimator $\hat{\beta}$, the simplest estimator in the sequence with average relative error $<$TOL.

—Algorithm Selection $(\tilde{X}, \tilde{y}, \{\beta^k\}_k, \text{TOL})$

Step 1: Let $\hat{\beta}$ be the model β^k with less coefficients $\neq 0$ that satisfies

$$\alpha_k = \sqrt{\frac{1}{p}(\tilde{y} - \tilde{X}\beta^k)^T(\tilde{y} - \tilde{X}\beta^k)} \leqslant \text{TOL}$$

—Output: $\hat{\beta}$.

The quantity α_k used to determine the cutoff is the square root of an average RSS of the linear models considered. We opted for this criteria due to the engineering interpretation of this quantity, outlined below. There are a number of classic statistical tests, such as the F test, that are used in linear regressions. However, such tests do not have the direct interpretation of α_k, and do not apply to the linearly constrained regressions we are considering; additionally, they require extra assumptions on the distribution of the errors.

Note that if we let $\hat{y}^k$ be such that $\log \hat{y}^k = \tilde{X}\beta^k$, and define $\Delta y^k = \hat{y}^k - y$, then we have

$$\alpha_k^2 = \frac{1}{p}(\tilde{y} - \tilde{X}\beta^k)^T(\tilde{y} - \tilde{X}\beta^k) = \frac{1}{p}\sum_{i=1}^{p} (\log y_i - \log \hat{y}_i^k)^2$$

$$= \frac{1}{p}\sum_{i=1}^{p} \log^2\left(1 + \frac{\Delta y_i^k}{y_i}\right)$$

Since for small values of $\Delta y_i^k / y_i$ we have that $\log(1 + (\Delta y_i^k / y_i)) \sim (\Delta y_i^k / y_i)$, an interpretation for the cutoff criteria of algorithm selection is that the quantity α_k^2 corresponds to the average squared relative error

$$\alpha_k^2 = \frac{1}{p}\sum_{i=1}^{p} \log^2\left(1 + \frac{\Delta y_i^k}{y_i}\right) \sim \frac{1}{p}\sum_{i=1}^{p} \left(\frac{\Delta y_i^k}{y_i}\right)^2$$

We refer to α_k as the average relative error (avg. RE).

3.3 Rounding the Model. The input to function Round is an estimator $\hat{\beta}$, and its output is a related estimator β^* that satisfies the units constraints, has all coefficients rounded to a number with decimals, either 0, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, or $\frac{3}{4}$, and minimizes the increase in error. Note that since 0 is a round number, β^* has all the zeros of $\hat{\beta}$. The choice of using quarters and thirds as the finest division strikes a balance between independence from experimental error and physical meaning. A large number of known laws for engineering problems, probably the vast majority, involve exponents consistent with this choice. For example, scaling laws for a boundary layer involve whole numbers and halves in the exponents, and if the boundary layer involves heat transfer, we obtain exponents with thirds [24]. Additionally, rounding the exponents reduces their variation with experimental error.

To obtain β^*, the algorithm sequentially fixes the closest exponent to its rounded version, and solves a linearly constrained linear regression similar to problem (8). The problems solved at each iteration include the units constraint and linear constraints that fix

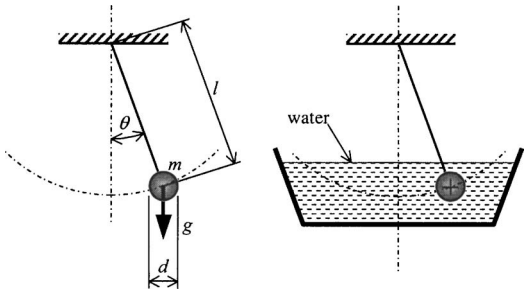


Fig. 2 Representation of a simple pendulum and its elements

coordinates to their rounded values. This is iterated until all exponents are fixed; the resulting β^* satisfies the units constraint, as this is enforced at very iteration.

3.4 Overall SLAW Algorithm. The overall algorithm performs the four steps outlined at the start of this section. For an input of statistical data $\tilde{y}$, $\tilde{X}$, and units constraint R and b, the algorithm SLAW does

Step 1: Run Dimension Reduction $(\tilde{X},\tilde{y},R,b)$, and obtain the sequence $\{\beta^k\}_k$, $\{\delta^k\}_k$.

Step 2: Run Selection $(\tilde{X},\tilde{y},\{\beta^k\},\mathrm{TOL})$, and obtain $\hat{\beta}$.

Step 3: Run Round $(\hat{\beta})$, obtaining β^*.

Step 4: Run Dimension Reduction $(\tilde{X},\tilde{y}-\tilde{X}\beta^*,R,0)$, to obtain $\{\tilde{\beta}^k\}_k$ and $\{\tilde{\delta}^k\}_k$.

—Output: β^* and $\{\tilde{\delta}^k\}_k$.

4 Examples

In this section, we present the results of applying the algorithm SLAW to three different physical experiments: a pendulum, the bonding of ceramics to metals, and a "punch test" recently developed at Exponent, Inc. As we show below, in each case SLAW "rediscovered," with a minimum of human input, the scaling laws that match previous analysis of the problem.

The input files for these three examples can be downloaded from [25]. For each problem, there are two text input files, one containing the experimental data and the other describing the units matrix.

4.1 Period of a Pendulum. Figure 2 shows a schematic of a simple pendulum and some of its elements. When the only relevant force acting on the pendulum is the force of gravity, and for small oscillations, the period of the pendulum is given by the following formula:

$$T = 2\pi\sqrt{l/g} \tag{9}$$

where T is the period (the dependent variable in this problem), l is the length of the string, and g is the acceleration of gravity. In what follows, we denote by T the observed period, and by T_{theory} the quantity in Eq. (9). We now use the SLAW algorithm to discover this relationship, assuming that the period of the pendulum is determined by the parameters in Table 1.

Table 1 Parameters involved in the pendulum example

Symbol	Units	Description
T	s	period (dependent variable)
m	kg	mass of the bob
l	m	length of the pendulum
θ	rad	initial angle
d	m	characteristic dimension of the bob
ρ	$\mathrm{kg/m^3}$	density of fluid surrounding the bob
g	$\mathrm{m/s^2}$	acceleration of gravity

These parameters account not only for inertial forces and gravity, but also for drag, rotational inertia of the bob, and initial angle. If these last three effects are neglected, dimensional analysis alone can solve Eq. (9), with the exception of the constant, which can be estimated by a single experiment. When these three effects are considered, dimensional analysis alone does not provide a unique formula. In this work, we considered these effects, which make the problem more complex, but more representative of real situations.

4.1.1 Input for SLAW. The input for SLAW are the set of experimental data and a matrix describing the units of the dependent and independent variables. The set of experimental data consists of a table listing the measured value of the period for different values of the parameters. Table 2 below displays the matrix R, which contains the units of the dependent variable and the parameters. This matrix was constructed using the units listed in Table 1; each element corresponds to the exponent of a unit (listed in the left column) for a given parameter (listed in the top row).

4.1.2 Output From SLAW. To illustrate the workings of SLAW, we here illustrate the output of the different steps of the SLAW algorithm. Table 3 displays the result of the iterations to obtain the initial simplest, meaningful, and dimensionally correct scaling law.

In this table, the first iteration corresponds to a power law without the units constraint. This model corresponds to what is typically used in engineering to fit experimental data. It is the most mathematically accurate of all possible power laws; however, it is physically incorrect, as the estimates of the model do not have the right units. The second iteration is a modification of the first power law, chosen to provide the correct units with the minimum increase in fitting error. These first two iterations use all parameters of the problem, regardless of their relevance. The third to sixth iterations remove the least significant independent variables one at a time; for example, the exponent for θ in the third iteration

Table 2 Matrix of reference units *R* for the pendulum example

Units	T	m	l	θ	d	ρ	g
m	0	0	1	0	1	−3	1
kg	0	1	0	0	0	1	0
s	1	0	0	0	0	0	−2

Table 3 Results of Step 1 of SLAW, for the pendulum example

Iter.	Param.	Constant	m	l	θ	d	ρ	g	avg. RE
1	6	0.117	0.000	0.502	0.023	−0.004	0.021	0.266	0.0197
2	6	2.050	−0.021	0.506	0.025	0.057	0.021	−0.500	0.0197
3	5	1.998	−0.016	0.500	0.000	0.049	0.016	−0.500	0.0197
4	4	1.861	−0.006	0.517	0.000	0.000	0.006	−0.500	0.0200
5	3	1.872	0.000	0.500	0.000	0.000	0.000	−0.500	0.0260
6	2	0.000	0.000	0.500	0.000	0.000	0.000	−0.500	1.8582

Table 4 Result of Step 3 of SLAW, for the on pendulum example

Iter.	Param.	Constant	m	l	θ	d	ρ	g	avg. RE
1	6	−0.001	0.000	0.002	0.023	−0.004	0.021	−0.002	0.020
2	5	0.178	−0.021	0.006	0.025	0.057	0.021	0.000	0.021
3	4	0.195	−0.021	0.000	0.022	0.064	0.021	0.000	0.022
4	3	0.125	−0.016	0.000	0.000	0.048	0.016	0.000	0.026
5	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.044

is set to zero after determining that the corresponding average relative error in this case was smaller than that of setting to zero the exponent of any other parameter. The sixth iteration is the minimum expression that provides the correct units after removing parameters in this order. Since the average relative error (avg. RE) jumps from 2.6% to 185.8% in the last iteration, we decide to use Iteration 5 as the basis of the final scaling law. This corresponds to the simplest model that provides an acceptable error. The selection of this model constitutes Step 2 of SLAW and it is automatically performed using a criteria that the average relative error (avg. RE) has a tolerance of TOL=0.2.

The behavior of exhibiting dramatic increases in the fitting error of models as variables are removed is the norm, not the exception, for physical processes that can be described with scaling laws. In this case, there is little error as the less relevant terms are discarded, and the error jumps suddenly when one of the essential terms of the power law is eliminated. This sudden jump in error can often be used by SLAW to automatically identify a cutoff point in the simplifications. The scaling law automatically obtained by SLAW for this problem is

$$T_{\mathrm{SLAW1}} = e^{1.872}\sqrt{l/g} \tag{10}$$

where T_{SLAW1} is the estimation of the period based in this scaling law. The constant factor obtained is within a 3% error from the exact coefficient 2π, and the functional dependence is the same as predicted by theory in Eq. (9).

Table 4 displays the iterations that determine the dimensionless groups ranked by relevance. Each line of this table corresponds to a dimensionless group that can be used to improve the predictions of the scaling law.

The first iteration corresponds to fitting the error of the scaling law (10) with a dimensionless group that considers all parameters. The fourth iteration corresponds to the simplest dimensionless group obtained after eliminating the less relevant parameters: $\Pi_1=(\rho d^3/m)^{0.016}$. The last iteration (the fifth in this case) always corresponds to a constant numerical factor. In this case, this factor is one (this row contains only zeros) because the scaling law obtained in the first iteration already has a constant numerical factor. Incorporating the most relevant dimensionless group yields the following scaling law:

$$T_{\mathrm{SLAW2}} = e^{1.997}\sqrt{\frac{l}{g}}\left(\frac{\rho d^3}{m}\right)^{0.016} \tag{11}$$

which has a better predictive value for these experiments and incorporates the effect of fluid drag. Fluid drag is relevant in this case because we measured pendulums surrounded by air and water, which differ by a factor of three orders of magnitude in their density. In the set of relevant parameters, we did not include the fluid viscosity. This choice is based on engineering insight that viscous drag is negligible. A metric for the relative relevance of viscosity is the Reynolds number, which in our experiments varied between 10^2 and 10^4. These values correspond to a flow in which viscous drag is unimportant.

Figure 3 illustrates the predictive capabilities of the scaling law obtained with the SLAW algorithm. Figure 3 plots the observed period on the vertical axis versus the known law T_{theory} and the scaling laws, T_{SLAW1} and T_{SLAW2}, on the horizontal axis. We also plot the identity for comparison purposes. We note that all three models predict the observed period of the pendulum very accurately when the pendulum is surrounded by air. However, both T_{theory} and T_{SLAW1} are slightly off the observed values when the bob was submerged in water. This effect is corrected by including the most important dimensionless group that accounts for the effect of drag, and can be seen in the plot of T_{SLAW2}.

Table 5 presents the correlation with the observed period and the average residual sum of squares for T_{theory}, T_{SLAW1}, and T_{SLAW2}. The correlation provides an indication of how well the different models capture the functional dependence; in engineering terms, it captures the *precision* of the models. The average relative error captures not only the trends, but also how close the predictions and the measured data are. In engineering terms, this would be the *accuracy* of the models.

4.2 Strain Energy in Ceramic-to-Metal Joining. Figure 1 shows the geometry of the problem, which consists of two long cylinders; one made of ceramic and the other of metal. These two cylinders are joined at their circular bases at high temperature. The temperature variation between the hot joining temperature and the cooler room temperature causes the ceramic and the metallic cylinder to decrease slightly in size. Typically, the metallic cylinder will shrink more than the ceramic cylinder, causing very large stresses on and around the interface of the joint. These stresses weaken the joint; therefore, the calculation of these

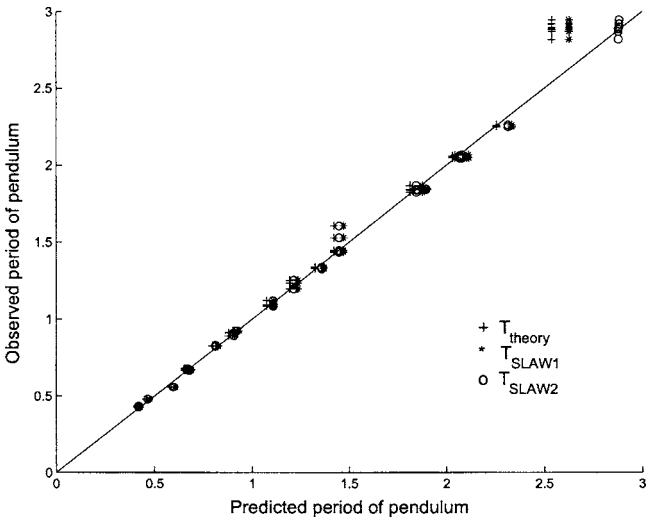


Fig. 3 Theoretical and SLAW-generated scaling laws for the pendulum example

Table 5 Correlation and error of the scaling laws for the pendulum example

Model	Correlation with T	avg. RE
T_{theory}	0.992	0.056
T_{SLAW1}	0.992	0.044
T_{SLAW2}	0.999	0.025

Table 6 Parameters involved in the ceramic-to-metal bonding example

Symbol	Units	Description
U	m^3 Pa	total elastic strain energy in the ceramic (dependent variable)
E_c	Pa	elastic modulus of the ceramic
E_m	Pa	elastic modulus of the metal
σ_Y	Pa	yield stress of metal
r	m	radius of cylinders
ε_T		differential thermal shrinkage

stresses is essential. The metric for these stresses is the "total elastic strain energy" U (units=Pa·m^3) accumulated in the ceramic. The total strain energy is the integral of the volumetric strain energy over the total volume of the ceramic cylinder.

Scaling factors exist for cases in which the metallic cylinder behaves elastically [26]. Similar scaling factors for when the metallic cylinder experiences nonlinear plasticity have been obtained only recently, by manual analysis of computational experiments [27]. In this paper, we will show how SLAW automatically obtains the same scaling factor of [27]. Similar to what is done in [27], we consider that the parameters of Table 6 represent the total strain energy in the ceramic. These parameters correspond to an elastic-plastic metal and a linear elastic ceramic.

4.2.1 Input for SLAW. Table 7 lists the elastic strain energy, which is the dependent variable that we wish to analyze, and the problem parameters for nine numerical simulations of ceramic-to-metal joints. In this problem, an engineering criterion was used to discard the length of the cylinders as a relevant parameter. The reason is that the cylinders considered are long enough, such that the far end does not influence the joined faces.

Table 8 displays the matrix of units R for this example.

4.2.2 Output From SLAW. We now present the output obtained from the SLAW algorithm for the above input data. The main outputs of SLAW are a simple scaling law, with rounded coefficients, and dimensionless groups that identify the most relevant parameters to describe the fitting error of this scaling law. The following scaling law was obtained after Step 1 of SLAW, using an avg. RE criteria with a tolerance of TOL=0.2, and the rounding procedure of Step 3.

$$U_{\text{SLAW1}} = \frac{\sigma_Y^2 r^3}{E_c} \tag{12}$$

This is the same scaling law that is obtained in [27] through ad hoc analysis and physical considerations. Table 9 displays the second set of iterations that determine the dimensionless groups and

Table 8 Matrix of reference units R for the ceramic-to-metal bonding example

Units	U	E_c	E_m	σ_Y	r	ε_T
Pa	1	1	1	1	0	0
m	3	0	0	0	1	0

numerical constant ranked by relevance.

Incorporating this last dimensionless group, the constant factor, in the scaling law, we obtain

$$U_{\text{SLAW2}} = e^{-0.333}\frac{\sigma_Y^2 r^3}{E_c} \tag{13}$$

In Table 10, we see that the average residual sum-of-square error is greatly reduced by considering this constant factor. This table also shows the correlation between the observed strain energy and what is predicted by U_{SLAW1} and U_{SLAW2}. Here, we additionally consider the scaling law $U_{\text{SLAW-lin}}$, given by the simple model selected by the avg. RE criteria, but prior to the rounding procedure.

$$U_{\text{SLAW-lin}} = \frac{\sigma_Y^{2.045} r^3}{E_c^{1.045}} \tag{14}$$

The model $U_{\text{SLAW-lin}}$ is the best simple linear regression model, but because the coefficients are not rounded, the model lacks physical interpretation. We notice that adding this physical interpretation by rounding the coefficients somewhat deteriorates the predictive value of our model; however, we still keep very high correlation with the observed strain energy and practically all the increased error can be recovered by incorporating the constant factor.

The predictive significance of these scaling laws is also observed in Fig. 4, which plots the observed strain energy versus what is predicted for the different models: $U_{\text{SLAW-lin}}$, U_{SLAW1}, and U_{SLAW2}. The figure shows that the experimental points fall closely around a straight line of slope one. We notice that the best prediction is given by model $U_{\text{SLAW-lin}}$, and that rounding the coefficients consistently overestimates the strain energy. This effect is compensated by incorporating the constant term in U_{SLAW2}.

4.3 Maximum Stress in the Punch Test. The punch test is an ASTM standard test, developed to determine mechanical properties of materials, such as ultrahigh molecular weight polyethylene used in surgical implants [28]. The test consists of using a spherical tip to push the center of a disk constrained at the edge. Figure 5 shows the geometry of the problem, where a sphere of radius r is pushing up on the center of a disk of radius L and width t.

The goal in this example is to obtain a scaling law that reproduces the maximum stress in the disk (stress at point A in Fig. 5)

Table 7 Input database containing the results of nine numerical experiments [27] for the ceramic-to-metal bonding example

Ceramic	Metal	U 10^{-2} Pa·m^3	E_c 10^{11} Pa	E_m 10^{11} Pa	σ_Y 10^8 Pa	r 10^{-3} m	ε_T 10^{-3}
Si_3N_4	Cu	0.423	3.04	1.28	7.58	6.25	6.85
Si_3N_4	Ni	1.52	3.04	2.08	1.48	6.25	5.15
Si_3N_4	Nb	2.80	3.04	1.03	2.40	6.25	2.10
Si_3N_4	Inco600	3.78	3.04	2.06	2.50	6.25	5.15
Si_3N_4	AISI 304	3.88	3.04	2.06	2.56	6.25	7.10
Si_3N_4	AISI 316	4.91	3.04	1.94	2.90	6.25	7.00
Al_2O_3	Ti	1.04	3.58	1.20	1.72	6.25	0.505
Al_2O_3	Inco600	3.00	3.58	2.06	2.50	6.25	2.95
Al_2O_3	AISI 304	3.16	3.58	2.00	2.56	6.25	4.90

Table 9 Result of Step 3 SLAW for the ceramic-to-metal bonding example

Iter.	Param.	Constant	E_c	E_m	σ_Y	r	ε_T	avg. RE
1	5	0.138	−0.151	0.192	−0.224	−0.701	0.128	0.026
2	4	−1.051	0.050	0.171	−0.221	−0.000	0.143	0.028
3	3	−1.087	−0.000	0.221	−0.221	0.000	0.131	0.030
4	2	0.642	−0.000	0.000	−0.000	−0.000	0.174	0.082
5	0	−0.333	−0.000	0.000	−0.000	0.000	−0.000	0.163

as a function of the displacement induced by the spherical tip. This problem has been addressed in [29], by an ad hoc analysis of computational experiments. Here, we show how SLAW automatically recreates the results obtaining the same scaling law. Similar to the work in [29], we consider a number of problem parameters that describe the elastic regime of the problem; these problem parameters and the dependent variable are listed in Table 11.

Table 10 Correlation and error of the scaling laws for the ceramic-to-metal bonding example

Model	Correlation with U	avg. RE
$U_{\text{SLAW-lin}}$	0.9840	0.1694
U_{SLAW1}	0.9835	0.3704
U_{SLAW2}	0.9835	0.1626

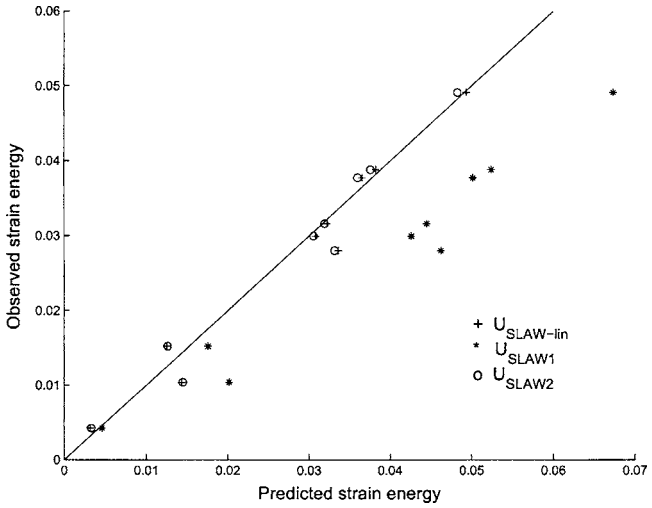


Fig. 4 SLAW-generated scaling laws for the ceramic-to-metal bonding example

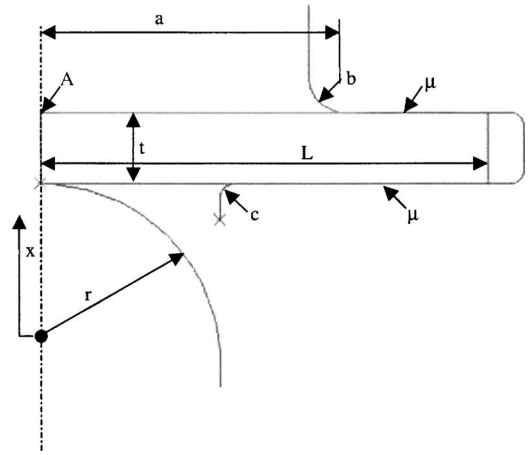


Fig. 5 Geometry of the punch test

4.3.1 Input for SLAW. This example considers the effect of Poisson's modulus υ. In solid mechanics, this parameter typically appears as a combination of υ, $1-\upsilon$, and $1+\upsilon$; therefore, we assigned three columns corresponding to each possibility. Inputs for SLAW are the table of experimental data and the units information of the variables, which is displayed in Table 12 below.

4.3.2 Output From SLAW. We now present the output obtained from the SLAW algorithm for the input data presented above. The principal outputs from SLAW are the simple, physically meaningful, scaling law and the most relevant dimensionless groups in describing the fitting error of the rounded scaling law. By running Step 1, using the avg. RE criteria with a tolerance of TOL=0.2, and the rounding procedure of Step 3, we obtain the following scaling law:

$$\sigma_{\text{SLAW1}} = \frac{E x_{\max}}{a} \tag{15}$$

We note that Step 1 of the algorithm realized 13 total iterations, and the last model with an avg. RE less than the tolerance of 0.2 is iteration 12. This scaling law is consistent with [29]. Again, the SLAW algorithm found this law automatically, while in [29] it took several hours of identifying trends manually. There is, however, a significant difference between Eq. (15) and the result in [29], which is the absence of Poisson's coefficient in the former, while the latter indicates that $1-\upsilon$ should appear in an analytical expression. SLAW indicates that the error of neglecting υ (or any of its variations) in the error of the approximation is similar to the error of neglecting other parameters that [29] also ignores. This does not mean that υ would not appear in an algebraic deduction; it means that for the data set analyzed, neglecting the effect of υ does not introduce a significant error. For comparison purposes, we will present the correlation and predictive value of the model found in [29], which we refer to as σ_{BK}.

We can reduce the error of σ_{SLAW1} by using the dimensionless groups and numerical constants obtained in Step 4 of SLAW. Table 13 displays a subset of the second group of iterations to determine the dimensionless groups and numerical constant, ranked by

Table 11 Parameters involved in the punch test example

Symbol	Units	Description
$\sigma_{\max}$	Pa	maximum stress at A (dependent variable)
L	m	radius of circular sample
t	m	thickness of circular sample
r	m	radius of spherical punch tip
a	m	radius of contact of top constraint
b	m	corner radius of top constraint
c	m	corner radius of bottom constraint
E	Pa	Elastic modulus of sample material
μ		Friction coefficient
$x_{\max}$	m	Maximum displacement of punch tip
υ		Poisson's modulus
$1-\upsilon$		1−Poisson's modulus
$1+\upsilon$		1+Poisson's modulus

Table 12 Matrix of reference units *R* for the punch test example

Units	σ_{max}	L	t	r	a	b	c	E	μ	x_{max}	v	$1-v$	$1+v$
Pa	1	0	0	0	0	0	0	1	0	0	0	0	0
m	0	1	1	1	1	1	1	0	0	1	0	0	0

Table 13 Summary of results of Step 4 of SLAW for the punch test example

Iter.	Param.	Constant	L	t	r	μ	x_{max}	v	$1+v$	avg. RE
1	12	1.19	0.40	−0.10	−0.14	−0.10	−0.17	0.72	−0.72	0.030
2	11	0.69	0.37	−0.10	−0.12	−0.08	−0.16	0.96	−2.57	0.031
3	10	0.72	0.37	−0.10	−0.11	−0.08	−0.16	0.97	−2.67	0.031
4	9	0.69	0.38	−0.10	−0.11	−0.08	−0.16	0.96	−2.62	0.032
5	8	0.72	0.38	−0.11	−0.11	−0.07	−0.15	0.96	−2.63	0.032
6	7	0.78	0.38	−0.11	−0.11	−0.07	−0.16	0.98	−2.73	0.032
7	6	−0.66	0.37	−0.12	−0.10	−0.08	−0.15	0.37	0.00	0.033
8	5	−0.49	0.35	−0.11	−0.10	0.00	−0.14	0.36	0.00	0.038
9	4	0.01	0.22	−0.12	−0.10	−0.00	−0.00	0.39	−0.00	0.044
10	3	0.10	0.15	−0.15	−0.00	−0.00	−0.00	0.43	−0.00	0.049
11	1	0.35	−0.00	0.00	−0.00	−0.00	−0.00	0.39	−0.00	0.061
12	0	−0.14	0.00	0.00	0.00	−0.00	0.00	0.00	−0.00	0.092

relevance.

The constant numerical factor compensates for the error introduced in the rounding of the exponents. Including this constant factor, we obtain the following scaling law:

$$\sigma_{SLAW2} = e^{-0.138}\frac{Ex_{max}}{a} \tag{16}$$

In Table 14, we show the correlation to maximum stress and average residual sum-of-square error of σ_{BK}, σ_{SLAW1}, and σ_{SLAW2}. We note that the average residual sum-of-square error is greatly reduced by adding the constant factor to σ_{SLAW1}. Although this table also shows that the scaling laws obtained from SLAW are slightly less correlated than σ_{BK} to the observed maximum stress, they are comparable and were obtained through an automatic methodology. In fact, σ_{SLAW2} is slightly less correlated to σ_{max} than σ_{BK}, but has smaller average residual sum of squares.

The predictive significance of these scaling laws is also observed in Fig. 6, which plots the observed maximum stress versus what is predicted for the different models: σ_{BK}, σ_{SLAW1}, and σ_{SLAW2}. The figure shows that the experimental points fall closely around a straight line of slope one; with model σ_{BK} overestimating the maximum stress, and model σ_{SLAW1} underestimating the maximum stress. Here, again, the model that incorporates the most significant dimensionless group (in this case, a constant factor) is seen as more representative of the observed σ_{max}, which is quantified with a smaller average residual sum of squares.

5 Conclusions

In this paper, we propose a new algorithm, named SLAW, to obtain scaling laws for complex systems from experimental data. The algorithm SLAW combines the ideas of dimensional analysis with statistical linear regression to obtain a representative model of the complex system. The algorithm automatically generates models of the system that not only achieve a high correlation with the observed experimental data, but are at the same time simple and physically meaningful. From linear regression, SLAW uses the concept of backward elimination to simplify the models considered, and to obtain a rank of the importance of problem parameters. Dimensional analysis is used to obtain a model that satisfies the units constraints and that considers only simple exponents. The proposed SLAW algorithm is able to:

1. find the simple scaling law that rules many real-life engineering problems.
2. provides a ranking of the significance of the parameters in the problem.

These two features of SLAW make it a useful tool for engineering design, where simple and approximate laws can be used by engineers to narrow down on a configuration during the conceptual stage of design. Even though the SLAW algorithm is based on the standard tools of dimensional analysis and linear regressions, it outperforms both of these techniques in identifying useful mod-

Table 14 Correlation and error of the scaling laws for the punch test example

Model	Correlation with σ_{max}	avg. RE
σ_{BK}	0.9592	0.1527
σ_{SLAW1}	0.9573	0.2507
σ_{SLAW2}	0.9573	0.1326

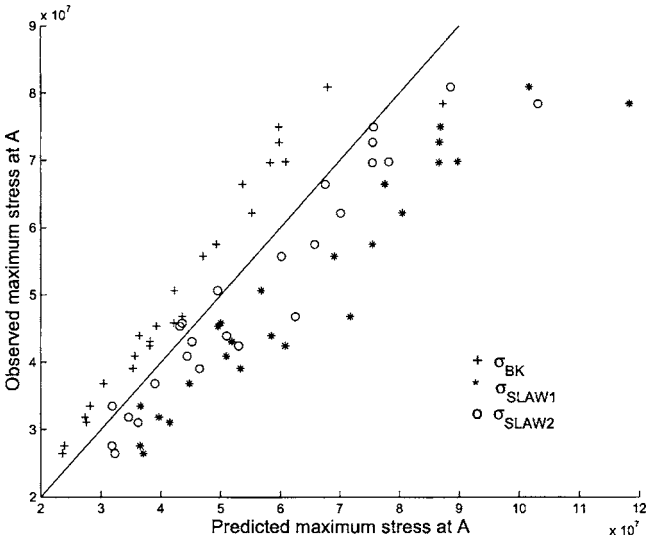


Fig. 6 SLAW-generated scaling laws for the punch test example

els for complex systems. The reason for this is that, more often than not, complex engineering problems can be explained using only a few of the physical parameters, and SLAW is able to identify these important parameters.

The SLAW algorithm makes two central assumptions on the physical system: (i) that a single physical regime is being modeled and (ii) that at least all the relevant parameters of the system have been included. The first assumption is necessary to obtain the simplest possible model. Since different regimes are characterized by different scaling laws, a data set that includes more than one regime would necessarily consider more parameters; this increases the complexity of the expression without adding to the physical understanding of the problem. The second assumption is standard for both dimensional analysis and linear regression models for physical systems. Clearly, it is not possible to explain a system if a key parameter is omitted.

We apply the SLAW algorithm to three very different examples, and note that SLAW automatically obtains the correct scaling law in each one. The resulting scaling laws are simple, correlate well with the experimental data, and are obtained using minimal user expertise. The only alternative method to obtain scaling laws in the last two examples involve time-consuming expertise-intensive analysis of the data.

Future work in this algorithm will involve, for example, investigating efficient alternative methods of selecting simple models. Currently, the algorithm eliminates one parameter at a time through backward elimination, and therefore does not consider all possible combinations of parameters. Therefore, the algorithm can, in theory, miss the most adequate model; to evaluate the impact of this on engineering practice, it is important to investigate other methods of generating simple models such as forward selection or even an exhaustive analysis of all combinations with correct units for small problems. An implementation of forward selection (albeit for predetermined dimensionless groups) was developed by Li and Lee [14].

Another potential improvement is a refined analysis of the dimensionless groups obtained. Currently, the less relevant dimensionless groups consider a larger number of parameters, which are also considered by the most relevant dimensionless groups. The determination of simpler dimensionless groups, minimizing the overlap of the parameters considered, would enhance their intuitive meaning.

Finally, the capability of SLAW of grouping a large amount of data under a single, simple law could be potentially used for clustering algorithms, separating experimental data into fewer and well-defined groups characterized by their different simple scaling laws. Further work will be necessary to explore the application of SLAW to data sets that cover more than one regime, perhaps drawing inspiration from previous multiple-regime work by Li and Lee [14] and the artificial intelligence algorithm ABACUS [17].

Acknowledgment

The authors thank Dr. Jorgen Bergström from Exponent, Inc., for the concepts and calculations used in the punch test example, and to Dr. Jin-Woo Park of Central R&D Institute in Samsung Electro-Mechanics for providing the data related to the ceramic-to-metal bonding.

References

[1] Kokshenev, V. B., 2003, "Observation of Mammalian Similarity Through Allometric Scaling Laws," Physica A, **322**(1-4), pp. 491–505.
[2] Azad, R. K., Bernaola-Galvan, P., Ramaswamy, R., and Rao, J. S., 2002, "Segmentation of Genomic DNA Through Entropic Divergence: Power Laws and Scaling," Phys. Rev. E, **65**(5), p. 051909.
[3] Washio, T., and Motoda, H., 1999, "Extension of Dimensional Analysis for Scale-Types and Its Application to Discovery of Admissible Models of Complex Processes," in *2nd Int. Workshop on Similarity Method*, University of Stuttgart, Stuttgart, pp. 129–147.
[4] Housen, K. R., Schmidt, R. M., and Holsapple, K. A., 1983, "Crater Ejecta Scaling Laws—Fundamental Forms Based on Dimensional Analysis," J. Geophys. Res., **88**(B3), pp. 2485–2499.
[5] Cho, J. Y. N., Newell, R. E., and Sachse, G. W., 1999, "Anomalous Scaling of Mesoscale Tropospheric Humidity Fluctuations," Geophys. Res. Lett., **27**(3), pp. 377–380.
[6] Carlson, J. M., and Doyle, J., 2000, "Power Laws, Highly Optimized Tolerance and Generalized Source Coding," Phys. Rev. Lett., **84**(24), pp. 5656–5659.
[7] de Jong, F. J., and Quade, W., 1967, *Dimensional analysis for economists*, Contributions to Economic Analysis 50, North Holland, Amsterdam.
[8] Szirtes, T., and Rózsa, P., 1997, *Applied Dimensional Analysis and Modeling*, McGraw Hill, New York.
[9] Segel, L. A., 1972, "Simplification and Scaling," SIAM Rev., **14**(4), pp. 547–571.
[10] Mendez, P. F., and Ordóñez, F., 2002, "Determination of Scaling Laws From Statistical Data," in *Fifth Int. Workshop on Similarity Methods*, University of Stuttgart, Stuttgart, pp. 21–31.
[11] Kasprzak, W., Lysik, B., and Rybachuk, A. M., 1999, "Dimensional Analysis in the Identification of Mathematical Models," http://www.immt.pwr.wroc.pl/kniga
[12] Vignaux, G. A., and Scott, J. L., 1999, "Simplifying Regression Models Using Dimensional Analysis," Aust. N. Z. J. Stat., **41**(1), pp. 31–41.
[13] Vignaux, G. A., 2001, "Some Examples of Dimensional Analysis in Operations Research and Statistics," in *4th Int. Workshop on Similarity Methods*, University of Stuttgart, Stuttgart, pp. 247–265.
[14] Li, C. C., and Lee, Y. C., 1990, "A Statistical Procedure for Model-Building in Dimensional Analysis," Int. J. Heat Mass Transfer, **33**(7), pp. 1566–1567.
[15] Dovi, V. G., Reverberi, A. P., Maga, L., and De Marchi, G., 1991, "Improving the Statistical Accuracy of Dimensional Analysis Correlations for Precise Coefficient Estimation and Optimal Design of Experiments," Int. Commun. Heat Mass Transfer, **18**(4), pp. 581–590.
[16] Bradshaw, G., Langley, P., and Simon, H. A., 1980, "Bacon 4: The Discovery of Intrinsic Properties," in *Proc. of the Third Nat. Conf. of the Canadian Society for Computational Studies of Intelligence*, Victoria, BC, pp. 19–25.
[17] Kokar, M. M., 1986, "Determining Arguments of Invariant Functional Descriptions," Mach. Learn., **1**(4), pp. 403–422.
[18] Washio, T., and Motoda, H., 1996, "Discovery of Possible Law Formulae Based on Measurement Scale," in *Proc. of Fourth Int. Workshop on Rough Sets, Fuzzy Sets and Machine Discovery*, Tokyo, pp. 209–216.
[19] Li, C. C., and Lee, Y. C., 1989, "Computational Aspects of Dimensional Analysis," Int. Commun. Heat Mass Transfer, **16**(2), pp. 315–321.
[20] Newcomb, S., 1881, "Note on the Frequency of Use of the Different Digits in Natural Numbers," Am. J. Math., **4**, pp. 39–40.
[21] Benford, F., 1938, "The law of anomalous numbers," Proc. Am. Philos. Soc., **78**(4), pp. 551–572.
[22] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M., 1993, *Nonlinear Programming, Theory and Algorithms*, 2nd Edition, Wiley, New York.
[23] Freedman, D., Pisani, R., and Purves, R., 1998, *Statistics*, 3rd Edition, W.W. Norton, New York.
[24] Schlichting, H., 1987, *Boundary-Layer Theory*, 7th Edition, McGraw-Hill Classic Textbook Reissue Series, McGraw-Hill, New York.
[25] Mendez, P. F., and Ordóñez, F., 2004, "SLAW, a Package for Scaling LAWs From Statistical Data," http://illposed.usc.edu/~pat/SLAW
[26] Blackwell, B. E., 1996, *A Framework for Determining the Mechanical Properties of Dissimilar Material Joints*, Doctor of Philosophy, Massachusetts Institute of Technology, Cambridge, MA.
[27] Park, J.-W., Mendez, P. F., and Eagar, T. W., 2002, "Strain Energy Distribution in Ceramic to Metal Joints," Acta Mater., **50**, pp. 883–899.
[28] American Society for Testing and Materials, 2002, *ASTM F 2183-02. Standard Test Method for Small Punch Testing of Ultra-High Molecular Weight Polyethylene Used in Surgical Implants*, W. Conshohocken, PA.
[29] Bergström, J., and Kurtz, S., 2003, "Extraction of Strength Properties of Bone Cement From Uniaxial Tension, Uniaxial Compression, and Small Punch Data," Technical Report MT00046.0MD/C0F2/0103/0001, Exponent, January 31.

Energy Release Rates for an Edge Delamination of a Laminated Beam Subjected to Thermal Gradient

M. Toya
Professor
Department of Mechanical Engineering,
Kagoshima University, 1-21-40 Korimoto,
Kagoshima 890-0065, Japan
e-mail: toyamasa@mech.kagoshima-u.ac.jp

M. Oda
Research Assistant

A. Kado
Graduate Student

T. Saitoh
Chief Research Engineer

ULSI Device Development Division,
NEC Corporation,
Kanagawa 229-1198, Japan

Energy release rates for an edge delamination of a laminated beam subjected to through-thickness temperature gradient are analyzed on the basis of the classical beam theory. The decomposition of the energy release rate into mode I and mode II components is made by combining the analyses of the energy release rates by Toya (1992) and the two-dimensional elasticity solutions for a split-beam element by Suo and Hutchinson (1990). The energy release rate is a quadratic function of the temperatures of the top and bottom surfaces of the beam. The transition of the type of crack growth between pure mode II and mixed mode type occurs at the temperature difference corresponding to the minimum energy release rate. Numerical analyses based on finite-element method are also carried out, which show that the theory agrees well with numerical results when temperature jump across the delaminated surfaces is relatively small as compared with the temperature difference between the top and bottom surfaces of the layered beam. [DOI: 10.1115/1.1978917]

1 Introduction

Layered materials consisting of load-carrying metallic substrates and ceramic layers deposited on them are expected to be of highly temperature tolerant and wear resistant. Actually thermal barrier coatings deposited on combustor liner and turbine blades are being investigated for use in commercial aircraft engines. These plate or shell components are put under thermal gradient and consequently stresses are generated as a result of thermal mismatch between adjacent layers. Especially, when initial delaminations, or interface cracks, exist on the interface of the layers, stress fields become singular at the crack tip and these high magnitude of thermal stresses promote debonding of the interface, which lead to serious degradation of materials. Therefore to elucidate the condition of the growth of initial delaminations is required for application of layered materials to commercial structures.

Brown and Erdogan [1] and Kuo [2] have investigated thermal stresses around partially insulated two-dimensional interface cracks lying on the interface of two semi-infinite different elastic materials subjected to normal temperature gradient. Martin-Moran et al. [3] and Barber and Comninou [4] studied the corresponding disc-shaped crack problems. Hutchinson and Lu [5] analyzed the energy release rate of a partially insulated interior delamination embedded in an orthotropic beam which is subjected to temperature gradients.

Even for perfect bond, thermal stresses along the bond-line become singular when approached to the free edge (Bogy [6]; Kuo [7]). Thus delamination is most likely to occur from the free edge, and hence the study of edge delamination is practically very important. Toya et al. [8] calculated the energy release rate of a delamination at the edge of a laminated beam subjected to a uniform temperature change. They confirmed that theoretical results based on simple strength of materials theory compared well with those of finite-element method.

In this paper we analyze an interface edge crack in a laminated beam subjected to temperature gradient. The energy release rate is obtained on the basis of the strength of material theory. Then the decomposition of the energy release rate into mode I and mode II components is made by combining the two-dimensional linear beam solutions by Suo and Hutchinson [9] and the analysis of the components of the energy release rates of an interface crack by Toya [10]. Numerical analyses based on finite-element method are also carried out and it is shown that the theory is accurate enough when temperature jump between the upper and lower surfaces of the crack is small.

2 Analyses

2.1 Reduction of the Problem to a Split-Beam Subject to Externally Applied Forces and Moments. Figure 1 shows the model of an edge delamination of a bimaterial beam subject to temperature gradient. The beam is formed by bonding two different isotropic and linearly elastic rectangular beams having the same length ℓ and width b, and it contains an edge delamination with the length c. The thicknesses, Young's moduli, coefficients of thermal expansion, and coefficients of thermal conductivity of the two layers are denoted by h_1, E_1, α_1, λ_1 for the upper beam and h_2, E_2, α_2, λ_2 for the lower beam, respectively. The top surface ($y=h_1$) is subject to a uniform temperature T_1 and the bottom surface ($y=-h_2$) T_2. As in the model considered by Hutchinson and Lu [5], we also assume that there is a temperature gap on the delamination surfaces ($y=\pm 0, 0\leqslant x\leqslant c$), with the temperatures being T_A and T_B for the upper and lower faces of the delamination, respectively. At uniform temperature of $T_1=T_2=0\,^\circ\mathrm{C}$, no thermal stress is induced in the beam.

For simplicity, we assume that no heat flows in the axial (x) direction across the cross section passing through the delamination tip B, and hence the temperature gradients in both the bonded and delaminated regions are uniform along the x axis (Hutchinson and Suo [5]). With this assumption, temperature distributions are

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, June 28, 2001; final revision, November 4, 2004. Associate Editor: V. K. Kinra. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California–Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

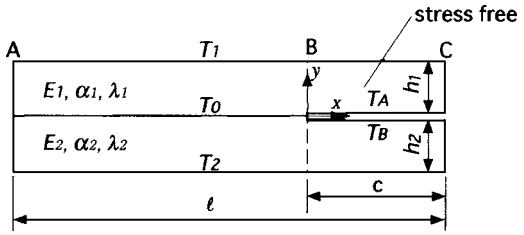


Fig. 1 Model of a laminated beam with an edge delamination subject to temperature gradient

readily determined. In the bonded part AB, denoting by T_0 the temperature at the interface, temperature distributions in each layer are given as

$$T_{01}(y) = (T_1 - T_0)y/h_1 + T_0 \quad (y > 0), \tag{1a}$$

$$T_{02}(y) = (T_0 - T_2)y/h_2 + T_0 \quad (y < 0) \tag{1b}$$

The heat fluxes in the y direction in the upper $(y>0)$ and lower $(y<0)$ layers are given as

$$q_1 = -\lambda_1(T_1 - T_0)/h_1, \quad q_2 = -\lambda_2(T_0 - T_2)/h_2. \tag{2}$$

From the condition $q_1 = q_2 (\mathrm{W/m^2})$, we obtain

$$T_0 = (T_1 + \eta T_2)/(1 + \eta), \tag{3}$$

where

$$\eta = (\lambda_2 h_1)/(\lambda_1 h_2). \tag{4}$$

Temperature distributions in the parts of the beam upper and lower to the delamination are given by Eq. (1*a*) with T_0 replaced by T_A and by Eq. (1*b*) with T_0 replaced by T_B, respectively. As is well-known, for homogeneous beams, no thermal stress is induced by the temperature distribution that is linear in y. Consequently, within the approximation that no heat conduction in the axial (x) direction occurs across the cross section at the tip of the delamination, strain energies in the parts of the beam over and below the delamination are zero. Hence, though we have started from rather general temperature condition, the energy release rate is determined solely by the strain energy in the bonded part AB, the temperatures T_A and T_B having no effect on the energy release rate. In other words, for the purpose of energy consideration, we can replace T_A and T_B by the temperature at the bond line T_0. We will later examine the accuracy of this approximation on the basis of finite-element analyses and elucidate under what condition our approximation can be validated.

The analyses of deformation of the beam may be conducted by assuming imaginative "cut and paste" procedures which consist of the two steps as illustrated in Figs. 2(*a*) and 2(*b*). At temperature 0°C, we imaginatively isolate the beam in many thin slices with

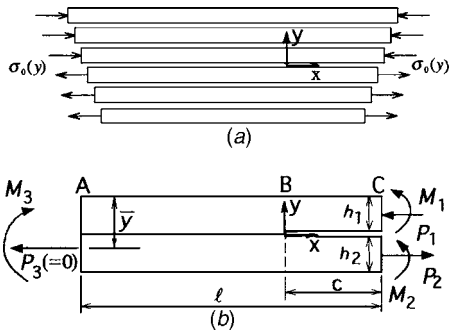


Fig. 2 Cut and paste procedure for analyzing thermal stress of a cracked laminate. ($\bar{y}$ is the distance of the neutral axis of the bonded part from the top surface.)

the thickness dy. With the temperature distribution as given by Eq. (1) with $T_A = T_B = T_0$, each slice will stretch to different lengths. In the first step, in order to make these slices equal in length, normal axial stress $\sigma_x = \sigma_0(y)$ is applied at both ends and in this stressed state all slices are bonded one another except over the surfaces of the interface crack $(y = \pm 0, 0 \leqslant x \leqslant c)$. At this stage axial resultant forces P_1, P_2, and P_3 are applied at the ends of the beam as shown in Fig. 2(*b*). In the second step, in order to realize the condition that actually no external stresses act at both ends of layers, the normal stress $\sigma_0(y)$ is removed. This removal of the stress is equivalent to the application of the moment M_3 about the neutral axis of the bonded beam that is distant $\bar{y}$ from the top surface at the left end of the beam. Likewise the moments M_1 and M_2 are produced at the right-hand ends of the upper and the lower beam of the crack, respectively [cf. Fig. 2(*b*); signs of these quantities are positive when they act in the directions as shown in Fig. 2(*b*)]. The longitudinal stress due to thermal gradient is given by the sum of $\sigma_0(y)$ and the stress induced by these moments. This technique was first introduced by Timoshenko [11] and was applied to the analysis of the deflection of bimetal plates.

Now, let us determine these moments and the axial forces. The extension of each slice Δ is given by

$$\begin{aligned} \Delta &= \{\alpha_1 T_{01}(y) + \sigma_0(y)/E_1\}\ell \quad (y > 0), \\ \Delta &= \{\alpha_2 T_{02}(y) + \sigma_0(y)/E_2\}\ell \quad (y < 0). \end{aligned} \tag{5}$$

If we assume that the extension Δ of each slice given by Eq. (5) is made equal to one another under the condition

$$\int_{-h_2}^{h_1} \sigma_o(y)dy = 0 \tag{6}$$

then the extension Δ is given by

$$\Delta = \frac{\ell}{E_1 h_1 + E_2 h_2}\left[\frac{T_1 + T_0}{2}E_1\alpha_1 h_1 + \frac{T_2 + T_0}{2}E_2\alpha_2 h_2\right] \tag{7}$$

Stress $\sigma_0(y)$ can be determined by substituting Eq. (7) into Eq. (5). For the upper layer $(y>0)$ we have

$$\sigma_0(y) = -\alpha_1 E_1(T_1 - T_0)y/h_1 + E_U \tag{8}$$

where

$$\begin{aligned} E_U = \frac{E_1}{2(E_1 h_1 + E_2 h_2)}&[(T_1 - T_0)E_1\alpha_1 h_1 - (T_0\alpha_1 - T_2\alpha_2)E_2 h_2 \\ &- (\alpha_1 - \alpha_2)T_0 E_2 h_2] \end{aligned} \tag{9}$$

and for the lower layer $(y<0)$,

$$\sigma_0(y) = -\alpha_2 E_2(T_0 - T_2)y/h_2 + E_L \tag{10}$$

where

$$\begin{aligned} E_L = \frac{E_2}{2(E_1 h_1 + E_2 h_2)}&[-(T_0 - T_2)E_2\alpha_2 h_2 + (T_1\alpha_1 - T_0\alpha_2)E_1 h_1 \\ &+ (\alpha_1 - \alpha_2)T_0 E_1 h_1] \end{aligned} \tag{11}$$

Next, we calculate P_1 and the moments M_1 and M_3 that are produced by the removal of $\sigma_0(y)$. They are given as follows:

$$P_1 = P_2 = b\int_0^{h_1} \sigma_0(y)dy = -h_1 b[\alpha_1 E_1(T_1 - T_0) - 2E_U]/2 \tag{12}$$

$$M_1 = b\int_0^{h_1} \sigma_0(y)(y - h_1/2)dy = -\alpha_1 b E_1 h_1^2(T_1 - T_0)/12 \tag{13}$$

$$M_3 = b\int_{-h_2}^{h_1} \sigma_0(y)\{y + (\bar{y} - h_1)\}dy$$
$$= -b[R(T_1 - T_2) + 3E_1E_2h_1h_2\{(Ah_1 + Ch_2)T_1 - (Bh_1 + \eta Dh_2)T_2\}]/[12(1+\eta)S] \quad (14)$$

where

$$R = \eta\alpha_1E_1^2h_1^3 + 4\eta\alpha_1E_1E_2h_1^2h_2 + 4\alpha_2E_1E_2h_1h_2^2 + \alpha_2E_2^2h_2^3 \quad (15)$$

$$S = E_1h_1 + E_2h_2 \quad (16)$$

and

$$A = 2\alpha_1 - \alpha_2, \quad B = -2\eta\alpha_1 + (1+2\eta)\alpha_2,$$
$$C = (2+\eta)\alpha_1 - 2\alpha_2, \quad D = 2\alpha_2 - \alpha_1 \quad (17)$$

Because of the condition Eq. (6), the axial force P_3 is zero. The problem of determining thermal stresses is thus reduced to the problem of a laminated beam subjected to the external axial forces and moments as shown in Fig. 2(*b*).

2.2 Energy Release Rate and Its Mode I and Mode II Components. Since the development of delamination is nothing more than the growth of the interface crack, the energy release rate becomes the most important parameter. As mentioned previously, for obtaining the strain energy of the beam, we need only to calculate the energy of the bonded part *AB*. First, we calculate the strain energy due to the application of normal stresses on both sides of thin strips [cf. Fig. 2(*a*)]. For the upper layer, the strain energy density is $u_1 = \sigma_0^2(y)/(2E_1)$, so that by integrating bu_1 from $y=0$ to h_1, we have the strain energy U_U per unit length of the beam as follows:

$$U_U = bh_1E_1[(3/S^2)\{-\bar{\alpha}_1^2E_1h_1(E_1h_1 + 2E_2h_2)(T_1 - T_2)^2 + 2\bar{\alpha}_1E_2^2h_2^2(AT_1 - BT_2)(T_1 - T_2) + E_2^2h_2^2(AT_1 - BT_2)^2\} + 4\bar{\alpha}_1^2(T_1 - T_2)^2]/[24(1+\eta)^2] \quad (18)$$

where $\bar{\alpha}_1 = \eta\alpha_1$. By the same manner, the strain energy U_L per unit length of the lower layer is given by

$$U_L = bh_2E_2[(3/S^2)\{-\alpha_2^2E_2h_2(2E_1h_1 + E_2h_2)(T_1 - T_2)^2 + 2\alpha_2E_1^2h_1^2(CT_1 - \eta DT_2)(T_1 - T_2) + E_1^2h_1^2(CT_1 - \eta DT_2)^2\} + 4\alpha_2^2(T_1 - T_2)^2]/[24(1+\eta)^2] \quad (19)$$

Next, the strain energy per unit length of the bonded part that is released by the relaxation of the stress $\sigma_0(y)$ is given by

$$U_{M3} = M_3^2/(2D') \quad (20)$$

where D' is the flexural rigidity of the laminated beam,

$$D' = D_1 + D_2 + h^2D_0/4 \quad (21)$$

with $D_1 = E_1I_1$ and $D_2 = E_2I_2$ [$I_j = bh_j^3/12$ ($j=1,2$)] being the flexural rigidities of the upper and lower layer, respectively, and

$$D_0 = b/[(E_1h_1)^{-1} + (E_2h_2)^{-1}], \quad h = h_1 + h_2 \quad (22)$$

The strain energy per unit length of the bonded part of the beam is given by $U_U + U_L - U_{M3}$ and by dividing this by the width of the beam b, the energy release rate is obtained. After somewhat tedious manipulation, arranging in terms of the temperatures T_1 and T_2, we obtain the energy release rate as follows:

$$G = \frac{1}{24(1+\eta)^2SH}(T_1 - T_2)^2Q + \frac{E_1E_2h_1h_2}{8SH}(T_1 - T_2)(f_1T_1 - f_2T_2) + G_0 \quad (23)$$

where G_0 corresponds to the energy release rate for the uniform temperature change ($T_1 = T_2$)

$$G_0 = E_1E_2h_1h_2T_1T_2(\alpha_1 - \alpha_2)^2(E_1h_1^3 + E_2h_2^3)/(2H) \quad (24)$$

The quantities Q, H, f_1, f_2 are constants that depend only on the geometry of the beam and mechanical and thermal properties of constituent materials; they are given as follows:

$$Q = SH(\eta^2\alpha_1^2E_1h_1 + \alpha_2^2E_2h_2) - R^2 \quad (25)$$

$$H = 12D'S/b \quad (26)$$

$$f_1 \equiv f_1(\alpha_1,\alpha_2,E_1,E_2,h_1,h_2,\eta) = [H\{(2+\eta)\alpha_1 - \alpha_2\}^2 - 3E_1E_2h_1h_2(Ah_1 + Ch_2)^2 - 2R(Ah_1 + Ch_2)]/(1+\eta)^2 \quad (27)$$

$$f_2 = f_1(\alpha_2,\alpha_1,E_2,E_1,h_2,h_1,1/\eta) = [H\{-\eta\alpha_1 + (1+2\eta)\alpha_2\}^2 - 3E_1E_2h_1h_2(Bh_1 + \eta Dh_2)^2 - 2R(Bh_1 + \eta Dh_2)]/(1+\eta)^2 \quad (28)$$

The energy release rate G is a quadratic function of T_1 and T_2, and the effect of the thermal conductivities on G appears only through the parameter η. As expected, for a homogeneous beam ($\alpha_1 = \alpha_2, E_1 = E_2, \eta = h_1/h_2$), G is zero. Further, Eq. (24) with both T_1 and T_2 replaced with ΔT (uniform heating or cooling) agrees with the previous result obtained by Toya et al. [8]. We also note that if $\alpha_1 = \alpha_2$, then $f_1 = f_2$, so that in this case G is proportional to $(T_1 - T_2)^2$ irrespective of the difference in the thermal conductivities, and the minimum value of G is zero. Since G does not depend on crack length, the delamination growth occurs in a steady state condition.

Equation (23) applies to plane stress condition. For plane strain condition, $E_i (i=1,2)$ appearing in Eqs. (23) and (24) and in all relevant equations [i.e., Eqs. (21)–(28)] should be replaced with $E_i/(1-\nu_i^2)$ and α_i with $\alpha_i(1+\nu_i)$, where ν_i is Poisson's ratio (Hutchinson and Lu, [5]).

In developing the theory and experiments of the condition of delamination growth, not only the total energy release rate but its mode I and mode II components are also important parameters, and hence it is desirable to separate the energy release rate into mode I and mode II components. To do this considerations based on the analyses of the near-crack-tip singular fields are needed.

With the definition of the complex factor in the notation by Suo and Hutchinson [9], which is similar to the stress intensity factor introduced by Malyshev and Salganik [12], singular stresses on the bond line are given as

$$\sigma_y + i\tau_{xy} = K(r/h_1)^{i\varepsilon}/(2\pi r)^{1/2} \quad (29)$$

where $K = K_I + iK_{II}$ (henceforth referred to as complex stress intensity factor) and r is the distance from the crack tip. The height h_1 of the upper beam is chosen as the normalizing factor for r. Further,

$$\varepsilon = (2\pi)^{-1}\ln[(1-\beta^*)/(1+\beta^*)] \quad (30)$$

where β^* is one of the two Dundurs' parameters,

$$\beta^* = [\Gamma(\kappa_2 - 1) - (\kappa_1 - 1)]/[\Gamma(\kappa_2 + 1) + (\kappa_1 + 1)] \quad (31)$$

Subscripts 1 and 2 refer to the upper and lower beams as before, and $\kappa_i = 3 - 4\nu_i$ for plane strain and $\kappa_i = (3-\nu_i)/(1+\nu_i)$ for plane stress, $\Gamma = \mu_1/\mu_2$, μ_i being shear modulus ($i=1,2$).

The normal component of the relative displacement of the two points of the upper and lower crack surfaces is given by

$$v = m\left[(K_I + 2\varepsilon K_{II})\cos\left(\varepsilon\ln\frac{r}{h_1}\right) - (K_{II} - 2\varepsilon K_I)\sin\left(\varepsilon\ln\frac{r}{h_1}\right)\right]\sqrt{r} \quad (32)$$

where

$$m=[(\kappa_1+1)/\mu_1+(\kappa_2+1)/\mu_2]/[2\sqrt{2\pi}(1+4\varepsilon^2)\cosh(\pi\varepsilon)] \tag{33}$$

Suo and Hutchinson [9] gave formulas of K_{I} and K_{II} for the split-beam with a unit width subjected to general stretching and bending as shown in Fig. 2(*b*). Their expressions modified for a beam with width b are

$$K_{\mathrm{I}}=\frac{p}{\sqrt{2b}}\left[\frac{P_0}{\sqrt{A_0h_1}}\cos\omega+\frac{M_0}{\sqrt{Ih_1^3}}\sin(\omega+\gamma)\right] \tag{34}$$

$$K_{\mathrm{II}}=\frac{p}{\sqrt{2b}}\left[\frac{P_0}{\sqrt{A_0h_1}}\sin\omega-\frac{M_0}{\sqrt{Ih_1^3}}\cos(\omega+\gamma)\right] \tag{35}$$

where

$$P_0=P_1-C_1P_3-C_2M_3/h_1 \tag{36}$$

$$M_0=M_1-C_3M_3 \tag{37}$$

The constants p, A_0, I, γ, and C_i $(i=1,2,3)$ are expressed as follows:

$$p=[(1-\alpha^*)/(1-\beta^{*2})]^{1/2} \tag{38}$$

where

$$\alpha^*=[\Gamma(\kappa_2+1)-(\kappa_1+1)]/[\Gamma(\kappa_2+1)+(\kappa_1+1)] \tag{39}$$

is a Dundurs' parameter. Further, under plane stress condition, with the notations in the present paper,

$$A_0=\frac{D_0D_2}{bE_1h_1(D'-D_1)},\quad \sin\gamma=\sqrt{\frac{(D'-\bar{D})D_1}{(D'-D_1)\bar{D}}}$$

$$I=D_2/(12\bar{D}),\quad C_1=D_0/(bE_2h_2)$$

$$C_2=h_1hD_0/(2D'),\quad C_3=D_1/D',\quad \bar{D}=D_1+D_2 \tag{40}$$

[For plane strain condition we replace E_i $(i=1,2)$ appearing in Eq. (40) and in the definitions of D_0, D_1, D_2, and D' with $E_i/(1-\nu_i^2)$.] Values of angular quantity ω, which are dependent solely on the ratio h_1/h_2 and Dundurs' parameters, α^* and β^*, are provided in Tables 1–4 in Suo and Hutchinson [9]. By substituting M_1, M_3, and P_1 given by Eqs. (12)–(14) into Eqs. (36) and (37), stress intensity factors K_{I} and K_{II} are immediately obtained from Eqs. (34) and (35).

The opening (mode I) and sliding (mode II) components of the energy, G_{I}^{Δ} and G_{II}^{Δ}, that are released during an incremental extension of an interface crack, Δa, were first calculated by Sun and Jih [13]. One of the present writers gave formulas of G_{I}^{Δ} and G_{II}^{Δ} expressed in terms of the complex stress intensity factors as follows (Toya, [10]):

$$G_{\mathrm{I}}^{\Delta}=(G/2)[1+F(\varepsilon)\cos(2\varepsilon a'+\varphi(\varepsilon)+\theta_0)]\Delta a$$

$$G_{\mathrm{II}}^{\Delta}=(G/2)[1-F(\varepsilon)\cos(2\varepsilon a'+\varphi(\varepsilon)+\theta_0)]\Delta a \tag{41}$$

where

$$F(\varepsilon)=\left[\frac{\sinh(2\pi\varepsilon)}{2\pi\varepsilon(1+4\varepsilon^2)}\right]^{1/2} \tag{42}$$

$$a'=\ln[\Delta a/(2h_1)] \tag{43}$$

$$\varphi(\varepsilon)=\sum_{n=0}^{\infty}\left(\mathrm{Tan}^{-1}\frac{\varepsilon}{1+n}-\mathrm{Tan}^{-1}\frac{\varepsilon}{\frac{1}{2}+n}\right) \tag{44}$$

$$\theta_0=\begin{cases}\mathrm{Tan}^{-1}[(1-k^2+4\varepsilon k)/k_0] & \text{for } k_0>0\\ \mathrm{Tan}^{-1}[(1-k^2+4\varepsilon k)/k_0]+\pi & \text{for } k_0\leqslant 0\end{cases} \tag{45}$$

with

$$k=K_{\mathrm{II}}/K_{\mathrm{I}},\quad k_0=2[k+\varepsilon(k^2-1)] \tag{46}$$

Values of $\varphi(\varepsilon)$ are tabulated in Table I in Toya [10]. Especially when ε is small, say $|\varepsilon|<0.05$, the following approximation may be used

$$\varphi(\varepsilon)=-2\varepsilon\ln 2=-1.3863\varepsilon \tag{47}$$

Further, G is the total energy release rate (Malyshev and Salganik, [12])

$$G=\left(\frac{\kappa_1+1}{\mu_1}+\frac{\kappa_2+1}{\mu_2}\right)(K_{\mathrm{I}}^2+K_{\mathrm{II}}^2)/[16\cosh^2(\pi\varepsilon)] \tag{48}$$

Referring to Eq. (43), we see from Eq. (41) that the conventional definition for the components of the energy release rate, i.e., $\lim_{\Delta a\to 0}G_i^{\Delta}/\Delta a$ $(i=\mathrm{I},\mathrm{II})$, cannot be applied for interfacial cracks. Instead each component should be defined as

$$G_{\mathrm{I}}\equiv G_{\mathrm{I}}^{\Delta}/\Delta a,\qquad G_{\mathrm{II}}\equiv G_{\mathrm{II}}^{\Delta}/\Delta a \tag{49}$$

for some finite crack growth step size Δa (Toya, [10]; Toya et al., [14]). There is no definite way of choosing a proper size of Δa yet, but a reasonable definition of Δa should be such that Δa is large enough as compared with the size of the zone of oscillatory singularity for which the relative displacement v becomes negative, but still small enough so that Eqs. (29) and (32) provide a good approximation for the stresses and displacements. As such, a characteristic length of delamination process, e.g., the thickness of an adhesive layer could be suggested.

It seems to be common to express mode mix by a phase angle $\mathrm{Tan}^{-1}(K_{\mathrm{I}}/K_{\mathrm{II}})$ (e.g., Hutchinson and Suo [15]), but since G_{I} and G_{II} are solely dependent on $K_{\mathrm{I}}/K_{\mathrm{II}}$, both definitions of the mode mix are essentially equivalent. The present writers believe, however, that the introduction of $G_{\mathrm{I}}/G_{\mathrm{II}}$ as a characterizing parameter for mode mix is advantageous in that in finite element analyses it is easier to obtain as compared to the determination of K_{I} and K_{II}. Further, for relatively small ε (say, when $|\varepsilon|\leqslant 0.05$), G_{I} and G_{II} are not strongly dependent on Δa (Toya, [10]).

Now returning to our problem, we have from Eqs. (36) and (37),

$$P_0/b=A^*(T_1-T_2)/(1+\eta)+D_0(-AT_1+BT_2)/[2b(1+\eta)]+hD_0^2[(Ah_1+Ch_2)T_1-(Bh_1+\eta Dh_2)T_2]/[8bD'(1+\eta)] \tag{50}$$

$$M_0/b=B^*(T_1-T_2)/(1+\eta)+D_0D_1/[4bD'(1+\eta)]\times[(Ah_1+Ch_2)T_1-(Bh_1+\eta Dh_2)T_2] \tag{51}$$

where

$$A^*=\frac{D_0}{2b}\left(\frac{bhR}{12SD'}-\bar{\alpha}_1\right),\quad B^*=\frac{D_1}{bh_1}\left(\frac{bh_1R}{12SD'}-\bar{\alpha}_1\right) \tag{52}$$

Therefore, if either T_1 or T_2 is zero, then the ratio P_0/M_0 is constant, meaning that the ratio of $G_{\mathrm{I}}/G_{\mathrm{II}}$ or $K_{\mathrm{I}}/K_{\mathrm{II}}$ is independent of temperature gradient. Otherwise, generally the ratio $G_{\mathrm{I}}/G_{\mathrm{II}}$ are functions of both T_1 and T_2. It is readily checked that for homogeneous beams, K_{I} and K_{II} are zero, and for uniform temperature change $T_1=T_2$, the energy release rate calculated from Eq. (48) coincides with Eq. (24). Moreover, we can generally prove that Eq. (48) coincides with Eq. (23). We could thus obtain the total energy release rate from Eq. (48), but the resulting expression is very lengthy and hence the straightforward calculation of the strain energy in the previous section is advantageous only for obtaining explicit expression of the total energy release rate in a compact form.

2.3 Biot Number. In our original model (Fig. 1) the temperature jump across the crack surfaces was assumed. In the approxi-

mation that there is no heat flow in the axial direction across the cross section at B, the energy release rate does not depend on the temperature gap. In the next section, we will examine the appropriateness of our assumption by using finite-element method. To this end we here introduce Biot number which may be conveniently utilized as a parameter that characterizes the degree of temperature gap ([2],[5]). When the thermal-conductivity between two crack faces is characterized by h_c, the heat flux across the jump q_c (W/m^2) may be expressed as ([2],[5])

$$q_c = h_c(T_B - T_A) \tag{53}$$

Equations similar to Eq. (2) hold for the heat flux in the upper and lower beams of the delamination, viz.,

$$q_1 = -\lambda_1(T_1 - T_A)/h_1, \quad q_2 = -\lambda_2(T_B - T_2)/h_2 \tag{54}$$

By the condition of equilibrium of heat flux, $q_1 = q_2 = q_c$, T_A and T_B are expressed as follows:

$$T_A = \frac{T_1(1+\eta) + B_c(T_1 + \eta T_2)}{(1+B_c)(1+\eta)} \tag{55}$$

$$T_B = \frac{T_2(1+\eta) + B_c(T_1 + \eta T_2)}{(1+B_c)(1+\eta)} \tag{56}$$

where

$$B_c = \frac{\lambda_1 h_2 + \lambda_2 h_1}{\lambda_1 \lambda_2} h_c \tag{57}$$

is a dimensionless Biot number that characterizes the heat flow across the interface crack. $B_c = 0$ means that the crack is perfectly insulated, and hence $T_A = T_1$, $T_B = T_2$, while $B_c = \infty$ means that the heat flux is not disturbed by the crack and therefore $T_A = T_B = T_0$. The temperature jump across the crack is given by

$$T_A - T_B = (T_1 - T_2)/(1 + B_c) \tag{58}$$

In the next section we calculate the energy release rate for several values of Biot number using finite-element method and examine to what extent of Biot number our theoretical predictions of the energy release rate may be validated.

3 Analyses by Finite-Element Method (FEM)

In this section we compare the theoretical results with those from FEM. The FEM program developed by Okusa [16] is utilized, by which mutual contact of crack faces may be readily analyzed. We improved his program so that the thermal stresses in laminated beams might be treated.

To obtain the components G_{I} and G_{II}, the finite element crack closure simulation (Rybicki and Kanninen, [17]) is performed. In this method, the nodal forces required to close the virtual crack extension, Δa, which is taken as the width of the element adjacent to the crack-tip, are computed using the local compliance method (cf. Armanios et al. [18]). Both components of the energy release rate are then given by

$$G_{\mathrm{I}} = \frac{1}{2}\frac{Y\Delta v}{\Delta a}, \quad G_{\mathrm{II}} = \frac{1}{2}\frac{X\Delta u}{\Delta a} \tag{59}$$

where Y and X are normal and tangential nodal forces required to hold two nodes together at the crack-tip, and Δv and Δu are the normal and tangential relative displacements of the extended crack, respectively. The total energy release rate is given by the sum of both components.

The beam with length $\ell = 100$ mm and containing an edge crack with length $c = 30$ mm is chosen as a basic model beam (except the case where the effect of the crack length on the energy release rate is examined: cf. Fig. 8). The thicknesses of the beams h_1, h_2 are varied between 2 and 4 mm. The triangular element with the base length of 0.5 mm is chosen as a basic one. Element sizes are made smaller towards the crack tip, and the mesh of triangular elements with the base length $\Delta a = 0.125$ mm is used in the small region surrounding the crack-tip. A schematic diagram of the finite element discretization is shown in Fig. 3. The temperature T_2 was taken to be $T_2 = 0°C$, $250°C$, $500°C$, while T_1 is varied so that $|T_1 - T_2|$ ranges from $0°C$ to a maximum $1000°C$. We assume several Biot numbers ranging from 1 to infinity, and temperatures T_A and T_B determined from Eqs. (55) and (56) are imposed on the upper and lower crack faces. The support condition of the beam is taken as indicated in Fig. 3 and both side surfaces are assumed to be thermally insulated. Computations are performed under plane stress condition.

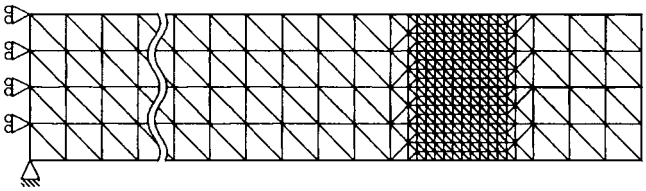


Fig. 3 Schematic diagram of the mesh division

Materials chosen are alumina (Al_2O_3), titanium alloy, Pyrex™ glass, and 0.2% C carbon steel. Young's moduli (E), Poisson's ratios (ν), coefficients of thermal expansion (α), and thermal conductivities (λ) of these materials are listed in Table 1. Model beams, for example, such as consisting of alumina and Pyrex glass are intended for verifying the accuracy of the theory and we bear no realistic application in mind. Further, we assume that the values of materials constants are independent of temperature.

In the first step of computation, we assume no contact of the crack faces, and check the sign of the relative displacements in the y direction of the nodes on crack faces nearest to the crack tip. If $v < 0$, this implies the contact of the crack faces. In this case the condition of contact, viz. the relative displacement in the y direction should equal to zero, is imposed for all nodes on the crack faces and calculation is freshly conducted. Next if tensile nodal forces are detected for nodes on the crack faces, the condition of contact is relaxed at these nodes at the next calculation. This procedure is repeated until no tensile nodal forces are detected at all nodes on the crack faces. In some cases, however, no convergence was found, in which case calculations were truncated after several repetitions. By these procedures, the relative displacements u and v between the two nodes nearest to the delamination tip are obtained. The mode I and mode II energy release rates are then calculated from Eqs. (59). The total energy release rate is obtained by the sum of these components.

Variations of the energy release rate and its mode I and mode II components with the temperature difference in the case of infinitely large Biot number are shown in Figs. 4 and 5 to compare the theoretical results with FEM solutions. The cases of $T_2 = 0°C$ (Fig. 4) and $T_2 = 500°C$ (Fig. 5) are chosen. Model materials combinations are alumina/titanium alloy (thicknesses with $h_1 = h_2 = 4$ mm) and alumina/Pyrex glass (thicknesses with $h_1 = h_2 = 2$ mm). In deriving theoretical values of each component of the energy release rates, we referred to the values of ω summarized in tables in Suo and Hutchinson [9]. According to their tables, $\omega \approx 51.7°$ (degree) for titanium alloy (upper layer)/alumina (lower layer) composite, and $\omega \approx 44.1°$ for alumina (upper layer)/Pyrex glass (lower layer). For the computations of alumina (upper

Table 1 Mechanical and thermal properties of model materials

	E (GPa)	ν	α ($\times 10^{-6}/°C$)	λ (W/m°C)
Alumina (Al_2O_3)	200.0	0.25	8.00	20
Titanium alloy	111.0	0.32	8.90	7.1
Pyrex glass	65.46	0.20	3.25	0.9
0.2%C Carbon steel	206.0	0.30	14.30	46

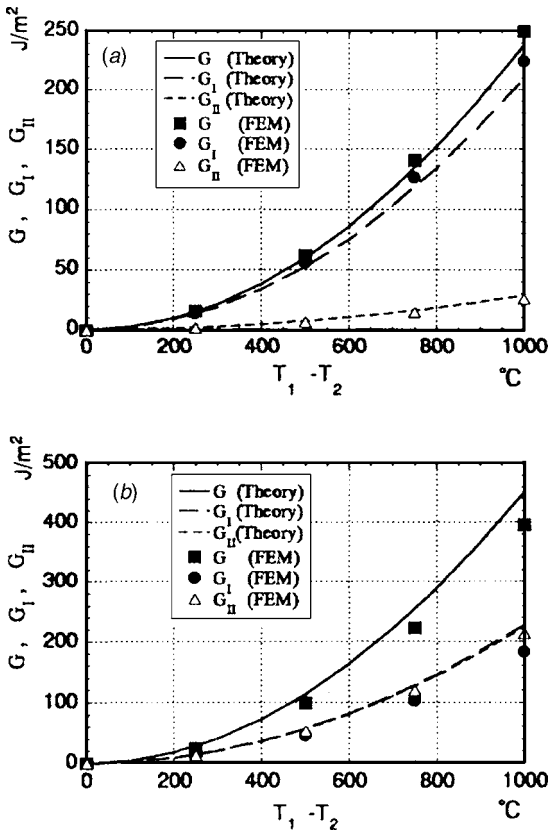


Fig. 4 Variation of energy release rate and its components with temperature difference for the case of $T_2=0°C$ and $B_c=\infty$. (a) Alumina (upper layer)/titanium alloy (lower layer), $h_1=h_2=4$ mm; (b) alumina/Pyrex glass, $h_1=h_2=2$ mm.

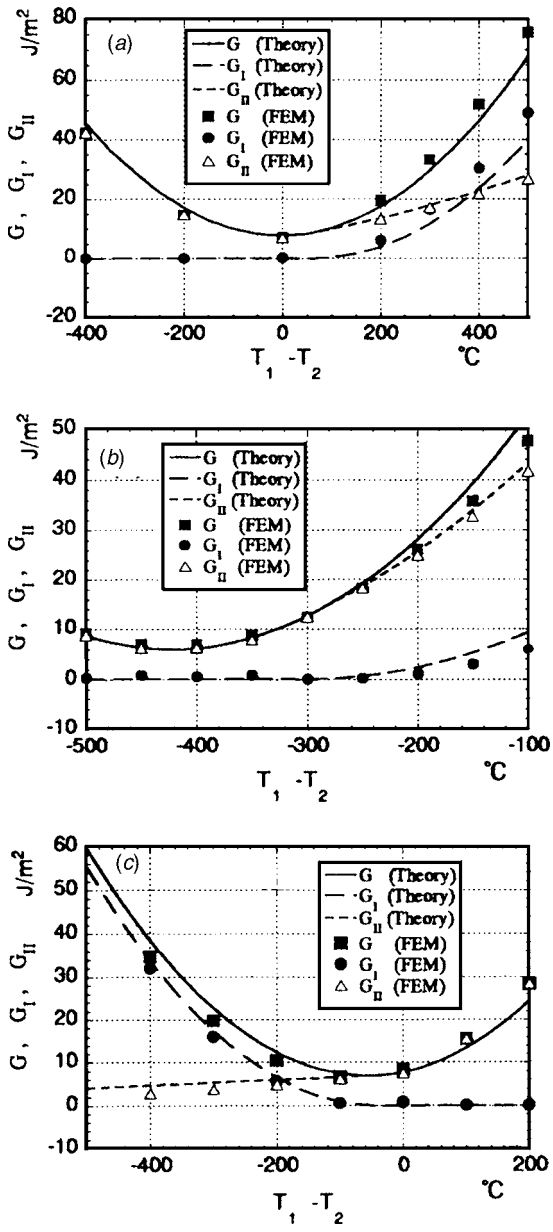


Fig. 5 Variation of energy release rate and its components with temperature difference for the case of $T_2=500°C$ and $B_c=\infty$. (a) Alumina (upper layer)/titanium alloy (lower layer), $h_1=h_2=4$ mm; (b) alumina/Pyrex glass, $h_1=h_2=2$ mm; (c) titanium alloy/alumina, $h_1=h_2=4$ mm.

layer)/titanium alloy (lower layer) composite [Figs. 4(*a*) and 5(*a*)], we reverse the beam and T_2 is varied while T_1 is held constant. Using these values and substituting Eqs. (50) and (51) into Eqs. (34) and (35), we obtain K_{I} and K_{II}. We then check the sign of relative displacement v from Eq. (32) at the distance $\Delta a=0.125$ mm. If v is positive, we calculate G_{I} and G_{II} according to Eqs. (49). While if negative, this means that $G_{\mathrm{I}}=0$ and hence G_{II} is equal to the total energy release rate G.

In Figs. 4 and 5 we observe that the theory agrees well with the computational results from FEM, the difference being less than 10%, which confirms the validity of the theory for infinitely large Biot numbers ($T_A \cong T_B$). When $T_2=0°C$, the energy release rate increases in proportion to the square of T_1-T_2 and becomes zero when $T_1-T_2=0°C$, agreeing with the theoretical prediction. When $T_2=500°C$, the condition $T_1-T_2=0°C$ implies the uniform heating with 500°C, and in this case the energy release rate is no longer zero. Further, the values of T_1-T_2 corresponding to the minimum energy release rate are not zero in general [cf. Eq. (23)]. For alumina/titanium alloy combination, for which values of thermal-expansion coefficients are close one another, the energy release rate becomes minimum at small values of T_1-T_2 as predicted from the theory [Figs. 5(*a*) and 5(*c*)].

In Figs. 4(*a*) and 4(*b*), we see that when $T_2=0°C$, the ratio of $G_{\mathrm{I}}/G_{\mathrm{II}}$ is constant independent of the temperature difference, whereas when $T_2=500°C$ (Fig. 5) the ratio of $G_{\mathrm{I}}/G_{\mathrm{II}}$ varies according as T_1-T_2 is varied. Theoretical values of $G_{\mathrm{I}}/G_{\mathrm{II}}$ and those from FEM are tabulated in Table 2, in which reasonable agreement between the theory and numerical results may be observed.

Further, we observe from Figs. 5(*a*)–5(*c*) that the transition of the type of delamination growth from pure mode II to mixed mode type, or in the reversed direction, occurs at the temperature difference corresponding to the minimum energy release rate. This mode transition by the temperature difference is also consistent to the change of the sign of v given by Eq. (32).

The effects of Biot number on the energy release rate are shown in Fig. 6 for alumina/Pyrex glass combination for the case of $T_2=0°C$ [Fig. 6(*a*)] and 500°C [Fig. 6(*b*)]. In Fig. 6(*a*), agreeing with our qualitative expectation, the energy release rate becomes larger as Biot number decreases from infinity to zero. In Fig. 6(*b*) the minimum value of the energy release rate becomes larger as B_c decreases. On the other hand, in Fig. 6(*a*) theoretical results compare well with the numerical solutions when Biot number is 1, while in Fig. 6(*b*) the theory compares well when Biot number is larger than 7, for which the differences between the theory and numerical results are less than 15%. Similar results, which are not shown here for economy of space, were obtained for alumina/titanium alloy composite. Overall, for the materials combinations

Table 2 Ratio of the components of the energy release rate G_I/G_{II} (percent) for alumina/Pyrex glass ($h_1=h_2=2$ mm)

	$T_2=0°C$		$T_2=250°C$		$T_2=500°C$	
T_1-T_2	Theory	FEM	Theory	FEM	Theory	FEM
250°C	49.60/50.40	46.17/53.83	37.22/62.78	32.47/67.53	32.85/67.15	27.76/72.24
500°C			41.50/58.50	37.15/62.85	37.23/62.77	32.47/67.53
750°C			43.59/56.41	39.46/60.54	39.81/60.19	35.28/64.72
1000°C			44.82/55.18	40.83/59.17	41.50/58.50	37.15/62.85

considered in the present paper, we may conclude that the theoretical results would give reasonable value of the energy release rates for Biot number that is larger than 7.

It is to be noted that actually the parameter h_c introduced in Eq. (53) is a function of crack opening, δ, and Biot number is inversely proportional to δ ([5],[19]), which means B_c becomes infinite when approached to the crack tip (no temperature gap at the crack tip). Thus the condition of uniform temperature on crack faces should be regarded as an effective or averaged condition. Since the energy release rate will be strongly dependent on the temperature distribution near the crack tip while in the present problem it is little affected by the temperatures on crack faces well away from the crack tip, the condition of perfect insulation will be too extreme. Thus it is probable that the average Biot number may be taken to be rather large and the present theory assuming no temperature jump on crack faces would give reasonable predictions of the exact energy release rate.

Graphs of the energy release rate as a function of the thickness ratio, h_1/h_2, are shown in Fig. 7 for alumina (upper layer)/steel (lower layer) laminate, where T_1 is taken to be either 250°C or 500°C while T_2 is kept 0°C. At the FEM computation the height h_2 is fixed as 2 mm and h_1 is varied between 0.25 mm and 10 mm. It is confirmed that theoretical curves agree well with computational results. We note that for this combination of materials the energy release rate becomes the largest at relatively thin upper layer ($h_1/h_2 \approx 0.2$).

Both the present analyses and the theory of Suo and Hutchinson [9] are based on the assumption that the lengths of the cracked and bonded parts of the beam are much longer than the height of the beams, so that the application of our theory is limited to these cases. Figure 8 shows the variation of the energy release rate with the length of delamination, where the alumina/Pyrex glass laminate with $h_1=h_2=4$ mm is assumed and temperatures are chosen to be $T_1=250°C$ and $T_2=0°C$. We find that when $c>0.75\ h_1(h_2)$,

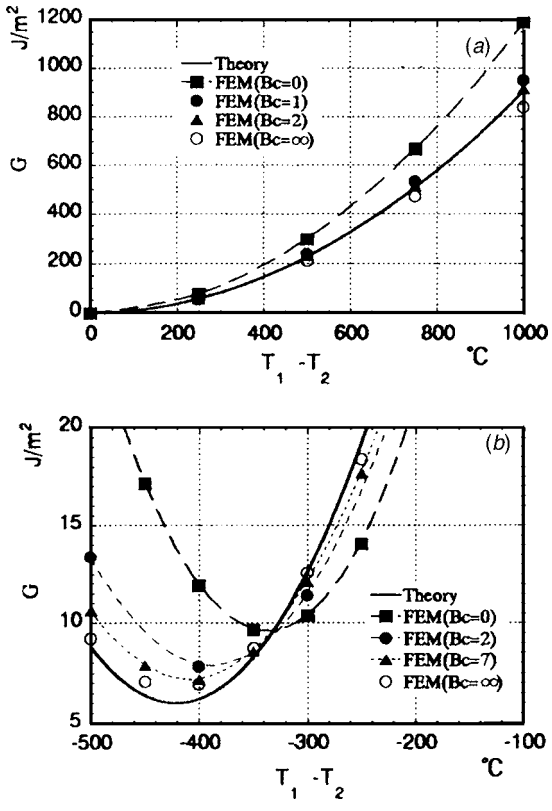


Fig. 6 Effect of Bio number on the relation between energy release rate and temperature difference for alumina (upper layer)/Pyrex glass (lower layer) laminate of $h_1=h_2=2$ mm. (a) $T_2=0°C$; (b) $T_2=500°C$.

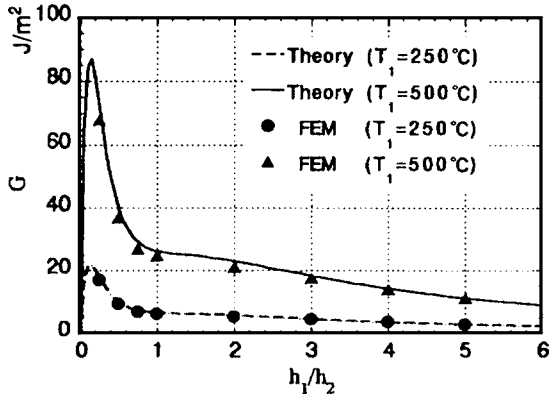


Fig. 7 Variation of energy release rate with thickness ratio for alumina (upper layer)/carbon steel (lower layer) laminate ($B_c=\infty$, $T_2=0°C$)

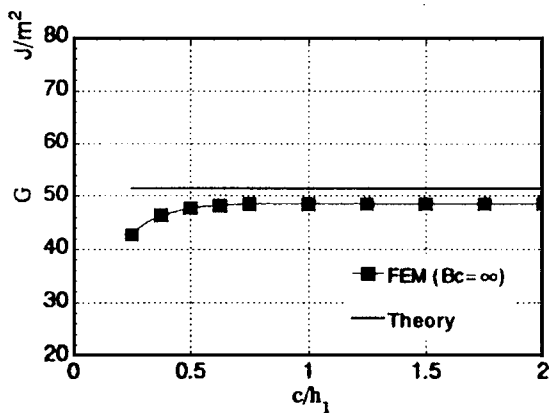


Fig. 8 Variation of energy release rate with the length of delamination for alumina (upper layer)/Pyrex glass (lower layer) laminate ($B_c=\infty$, $T_1=250°C$, $T_2=0°C$, $h_1=h_2=4$ mm)

the energy release rate reaches a constant that is close to the theoretical value and hence the present analysis may well be expected to apply to the crack length longer than the beam heights. Moreover, the present analyses give the largest value of the energy release rate. These features agree with the expectation of Hutchinson and Lu [5].

4 Concluding Remarks

The edge delaminations in layered beams subjected to thermal gradients have been analyzed. We theoretically derived the energy release rate on the basis of classical beam theory. Combining Toya's energy release rate analysis for an interface crack [10] and Suo and Hutchinson's formulas [9] of the complex stress intensity factor, the decomposition of the energy release rate into mode I and mode II components was made. The energy release rate was found to be a quadratic function of the temperatures of the top and bottom surfaces of the beam and independent of the beam and crack lengths. By comparing with the results of numerical analyses based on FEM, present theory has been found to be accurate enough when the length of delamination is longer than the thickness of the upper layer and there is no temperature gap between crack surfaces.

Further, if either T_1 or T_2 equals to zero, then the ratio of G_I/G_{II} is constant independent of temperature difference T_1-T_2. While if both T_1 and T_2 are not zero, then the ratio changes according as T_1-T_2 is varied. Theoretical analyses also predicted that the change of the mode of delamination growth from pure mode II type to mixed mode type, or in the reversed direction, occurs at temperature difference corresponding to the minimum energy release rate. This prediction was also confirmed in FEM computations.

FEM analyses for the cases where there is a temperature gap between upper and lower surfaces of the crack were also carried out. It was found that in the case where temperature gap is small (when Biot number is larger than 7), the difference between the theory and the results from FEM was less than 15%. Hence present theory will be particularly effective for crack growth in pure mode II, because in this case crack faces are contacting one another so that the temperature gap is expected to be very small.

It was argued that, since in fact there is no temperature gap at the crack tip, the average Biot number near the crack tip may be taken to be rather large even when crack faces are open. It is probable that the present theory assuming no disturbance of heat flow by the crack would give reasonable predictions of the energy release rates even for mixed mode delamination. However, the verification of this qualitative expectation is left for future study.

Acknowledgment

This work was supported by the Ministry of Education, Science and Culture, Japan (Grant in Aid for Scientific Research C(2), No. 09650110).

References

[1] Brown, E. J., and Erdogan, F., 1968, "Thermal Stresses in Bonded Materials Containing Cuts on the Interface," Int. J. Eng. Sci., **6**, pp. 517–529.

[2] Kuo, A. Y., 1990, "Interface Crack Between Two Dissimilar Half Spaces Subjected to a Uniform Heat Flow at Infinity-Open Crack," ASME J. Appl. Mech., **57**, pp. 359–364.

[3] Martin-Moran, C. J., Barber, J. R., and Comninou, M., 1983, "The Penny-Shaped Interface Crack with Heat Flow. I. Perfect Contact," ASME J. Appl. Mech., **50**, pp. 29–36.

[4] Barber, J. R., and Comninou, M., 1983, "The Penny-Shaped Interface Crack with Heat Flow. II. Imperfect Contact," ASME J. Appl. Mech., **50**, pp. 770–776.

[5] Hutchinson, J. W., and Lu, T. J., 1995, "Laminate Delamination Due to Thermal Gradients," ASME J. Eng. Mater. Technol., **117**, pp. 386–389.

[6] Bogy, D. B., 1968, "Edge-Bonded Dissimilar Orthogonal Elastic Wedges Under Normal and Shear Loading," ASME J. Appl. Mech., **35**, pp. 460–466.

[7] Kuo, A. Y., 1989, "Thermal Stresses at the Edge of a Bimetallic Thermostat," ASME J. Appl. Mech., **56**, pp. 585–589.

[8] Toya, M., Miyawaki, T., and Kirioka, K., 1992, "Analyses of Delamination of Laminated Beams by Bending," *Proc. 18th International Symposium on Space Technology and Science*, T. Hashiguchi, ed., pp. 527–534.

[9] Suo, Z., and Hutchinson, J. W., 1990, "Interface Crack Between Two Elastic Layers," Int. J. Fract., **43**, pp. 1–18.

[10] Toya, M., 1992, "On Mode I and Mode II Energy Release Rates of an Interface Crack," Int. J. Fract., **56**, pp. 345–352.

[11] Timoshenko, S. P., 1925, "Analysis of Bi-Metal Thermostats," J. Opt. Soc. Am., **11**, pp. 233–255.

[12] Malyshev, B. M., and Salganik, R. L., 1965, "The Strength of Adhesive Joints Using the Theory of Cracks," Int. J. Fract. Mech., **1**, pp. 114–128.

[13] Sun, C. T., and Jih, C. J., 1987, "On Strain Energy Release Rates for Interfacial Cracks in Bi-Material Media," Eng. Fract. Mech., **28**, pp. 13–20.

[14] Toya, M., Chosa, A., and Aritomi, M., 1997, "Energy Release Rates for an Interface Crack Embedded in a Laminated Beam Subjected to Three-Point Bending," ASME J. Appl. Mech., **64**, pp. 375–382.

[15] Hutchinson, J. W., and Suo, Z., 1992, "Mixed Mode Cracking in Layered Materials," Adv. Appl. Mech., **29**, pp. 63–191.

[16] Okusa, K., 1983, "Studies on the Shearing of Wood," Bulletin of the Faculty of Agriculture, Kagoshima University, **33**, pp. 193–202.

[17] Rybicki, E. F., and Kanninen, M. F., 1977, "A Finite Element Calculation of Stress Intensity Factors by a Modified Crack Closure Integral," Eng. Fract. Mech., **9**, pp. 930–938.

[18] Armanios, E. A., Regfield, L. M., and Reddy, A. D., 1986, "Design Analysis and Testing for Mixed-Mode and Mode II Interlaminar Fracture of Composites," *Composite Materials: Testing and Design*, J. M. Whitney, ed., ASTM STP 893, pp. 232–255.

[19] Lu, T. J., and Hutchinson, J. W., 1995, "Effect of Matrix Cracking on the Overall Thermal Conductivity of Fiber-Reinforced Composites," Philos. Trans. R. Soc. London, Ser. A, **351**, pp. 595–610.

Using Series-Series Iwan-Type Models for Understanding Joint Dynamics

D. Dane Quinn
Department of Mechanical Engineering,
The University of Akron,
Akron, OH 44325-3903
e-mail: quinn@uakron.edu

Daniel J. Segalman[1]
Sandia National Laboratories,
P. O. Box 5800, MS 0847,
Albuquerque, NM 87185-0847
e-mail: djsegal@sandia.gov

In mechanical assemblies, the energy loss induced by joints and interfaces can account for a significant portion of the overall structural dissipation. This work considers the dynamical behavior of an elastic rod on a frictional foundation as a model for the dissipation introduced by micro-slip in mechanical joints. In a quasi-static loading limit, the deformation of the rod and hence the frictional dissipation can be solved in closed form. The resulting model is a continuum model of series arrangements of parallel Jenkins elements. For a general class of normal load distributions, the resulting energy loss per forcing cycle follows a power-law and is qualitatively similar to observed experimental findings. Finally, these results are compared with those obtained from a discrete formulation of the rod including inertial effects. For loading conditions that are consistent with mechanical joints, the numerical results from the discrete model are consistent with the closed form predictions obtained in the quasistatic limit. [DOI: 10.1115/1.1978918]

1 Introduction

In many structures of great engineering importance, such as air frames or jet engines, the primary source of vibration damping is often just the frictional damping of interfaces associated with mechanical joints. This damping is associated with slip in outer regions of the contact patches and is known to be strongly amplitude dependent (and hence nonlinear) [1,2]. These mechanisms have traditionally been accommodated in structural dynamics only indirectly. For instance, one may use a finite element code to deduce modes and frequencies but then wait for data taken from a prototype tested at amplitudes of interest to obtain nominal values for modal damping coefficients.

As the need and expectation of *predictive* structural dynamics simulation grows, the requirement of systematically accounting for the role of joints in the structural response becomes more urgent [3]. Unfortunately, the most direct method of accommodating joint mechanics into finite element analysis—meshing the joint regions finely enough to capture any relevant micromechanics [4,5]—proves to be impractical for large-scale structural systems because of the prohibitively small time steps required and/or matrix ill-conditioning that results from the attempt to resolve the interfaces. A more practical approach is to devise constitutive models for the overall behavior of individual joints and to incorporate that constitutive response locally into the structural model. In the following we discuss a class of models that captures important qualitative properties of mechanical joints in a manner that can be integrated into conventional finite element codes.

The qualitative behavior generally found for joints is illustrated by two types of experiments. The first is a unidirectional lateral pull test. At small loads, the force-displacement curve appears linear, though some amount of micro-slip in the interface does take place. At larger loads, as slip increases, the curve begins to level out and finally, at macro-slip the curve becomes flat. Of course, if there is a bolt at the core of the joint, that bolt will eventually come into shear and a new apparently linear portion of the curve begins.

The second experiment that illustrates the core features of joint response is a lap joint subject to small amplitude lateral oscillatory loads. When energy dissipation per cycle is plotted against the force amplitude on a log-log scale, the result is generally well-approximated by a straight line, whose slope lies between 2.0 and 3.0 [6,7]. (It is interesting to note that Goodman pointed out forty years ago that the Mindlin solution for spheres pressed together and subject to similar *small* oscillatory loads predicts similar power-law dissipation with a slope of 3 [8].) Though the departure of power-law slope of experimental values from the Goodman value of 3.0 has often been ascribed to nonlinear material response [2], recent work [5] indicates that the slope may also depend on geometric nonlinearities reflective of the geometry of the joint.

This paper considers a simplified model for the prediction of energy dissipation in mechanical joints and interfaces. The model represents an elastic rod on a frictional foundation with time-varying shear loads and spatially varying normal pressure. In the quasi-static limit, these continuum equations are solved in closed form to determine the force-displacement relationship and thereby deriving an expression for the energy dissipation per cycle of harmonic forcing. Finally, the continuum model is discretized and the resulting n-dof model is studied for non-zero forcing frequency and as the model order n varies.

Iwan Models. Iwan considered two permutations of a spring and frictional damper—arranged in series and connected in parallel, sometimes referred to as Jenkins elements [9,10] (see Fig. 1). The parallel element allows for changes in force with zero change in displacement while the series element allows for changes in displacement with zero force change. Moreover, researchers including Iwan have considered both series and parallel collections of each of these elements. This gives rise to four combinations: series arrangements of parallel elements (series-parallel), series-series, parallel-parallel, and parallel-series. Of these, the most well-known systems are parallel arrangements of series elements. In this configuration, all spring stiffnesses are set to be identical, the sliders are all connected to ground, and the left node is left

[1]Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract No. DE-AC04-94AL85000.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, January 3, 2003; final revision, August 10, 2004. Associate Editor: A. A. Ferri. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

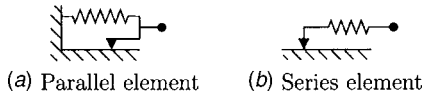


Fig. 1 Iwan elements

free. Iwan's parallel-series network has had some popularity, primarily because of equations that Iwan presented to deduce model parameters in terms of gross force-displacement behavior [11]. A similar analysis has not been available for the series-parallel system, and it has been generally assumed that the parallel-series network is unique in having the useful relationships between model parameters and gross behavior.

Iwan's equations for the deduction of parameters for a parallel network employs the force-displacement curve from monotonic loading for the boundary of the system. We show in the following that similarly useful relations can be obtained for another of the four permutations of Iwan network types: the series-series network. In the following, we do our best to derive equivalent expressions for the above-defined series-series system. In particular, the series-series arrangement provides a straightforward point of departure for considering the dynamical behavior of mechanical joints. In this work we attempt to relate the characteristics of these simplified models to experimentally observed behavior from mechanical lap joints as well as simulations based on a discrete model for a rod on a frictional foundation.

Physical model. The physical model under consideration can be described as a uniform elastic rod of length L held in place by a frictional surface, as illustrated in Fig. 2. The continuum model for this system can be written as

$$\rho\frac{\partial^2\tilde{u}}{\partial\tilde{t}^2}(\tilde{x},\tilde{t}) - EA\frac{\partial^2\tilde{u}}{\partial\tilde{x}^2}(\tilde{x},\tilde{t}) = \tilde{G}(\tilde{x},\tilde{t}), \tag{1}$$

where $\tilde{u}(\tilde{x},\tilde{t})$ is the lateral displacement of the rod at location $\tilde{x}$ and time $\tilde{t}$ while EA is the rod stiffness. The function $\tilde{G}(\tilde{x},\tilde{t})$ describes the force acting on the rod arising from friction. A Coulomb model is used to describe friction and it is assumed that the coefficient of friction does not vary over the interface. Also, no distinction is made between static and kinetic coefficients of friction. More realistic interface models would necessarily consider generalized descriptions of friction such as mesoscopic asperity based models [12,13] and friction laws dependent on interfacial variables, including slip rate [14]. The Amontons law of friction can be uniquely specified in terms of two quantities: $\mu N(\tilde{x})$, the frictional intensity, and $\mathcal{G}^{\mathrm{eq}}(\tilde{x},\tilde{t})$, the value of the friction force that would be required to maintain static equilibrium in a state of sticking. With these, the friction force is described by

$$\tilde{G}(\tilde{x},\tilde{t}) = \begin{cases} -\mu N(\tilde{x})\mathrm{sgn}(\dot{\tilde{u}}(\tilde{x},\tilde{t})), & \dot{\tilde{u}}(\tilde{x},\tilde{t}) \neq 0, \\ -\min\left(|\tilde{\mathcal{G}}^{\mathrm{eq}}(\tilde{x},\tilde{t})|, \mu N(\tilde{x})\right)\cdot \mathrm{sgn}\left(\tilde{\mathcal{G}}^{\mathrm{eq}}(\tilde{x},\tilde{t})\right), & \dot{\tilde{u}}(\tilde{x},\tilde{t}) = 0. \end{cases} \tag{2}$$

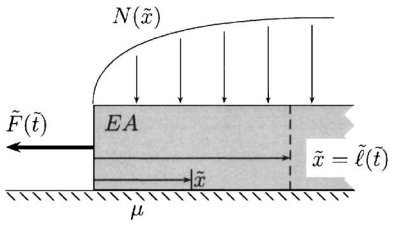


Fig. 2 Physical system. The rod slips over the interval $0 \leq \tilde{x} < \tilde{\ell}(\tilde{t})$.

The frictional intensity $\mu N(\tilde{x})$ is the product of the magnitude of the normal force and the coefficient of friction. Each of these quantities can depend independently on the spatial position along the rod. However, as they appear only in the combination $\mu \cdot N$, we do not differentiate between the $N(\tilde{x})$ and $\mu(\tilde{x})$. The function sgn is the sign of nonzero arguments and zero if its argument is zero. Finally, $\tilde{\mathcal{G}}^{\mathrm{eq}}(\tilde{x},\tilde{t})$, the force required to maintain sticking, is

$$\tilde{\mathcal{G}}^{\mathrm{eq}}(\tilde{x},\tilde{t}) = -EA\frac{\partial^2\tilde{u}}{\partial\tilde{x}^2}(\tilde{x},\tilde{t}) \tag{3}$$

This model is subject to a suitable initial state and a natural boundary condition at $\tilde{x}=0$:

$$\frac{\partial\tilde{u}}{\partial\tilde{x}}(0,t) = \frac{\tilde{F}(\tilde{t})}{EA}, \tag{4}$$

where positive values of $\tilde{F}$ imply the end of the bar is in tension.

We nondimensionalize this model through the transformations:

$$\tilde{x} = L\,x, \quad \tilde{t} = \sqrt{\frac{\rho L^2}{EA}}t, \quad \tilde{u} = \left(\frac{L}{EA}\int_0^L \mu N(\xi)d\xi\right)u,$$

$$\tilde{F} = \left(\int_0^L \mu N(\xi)d\xi\right)F, \tag{5}$$

yielding the following equations:

$$\frac{\partial^2 u}{\partial t^2}(x,t) - \frac{\partial^2 u}{\partial x^2}(x,t) = G(x,t), \quad x \in (0,1) \tag{6}$$

with:

$$G(x,t) = \begin{cases} -\psi(x)\mathrm{sgn}(\dot{u}(x,t)), & \dot{u}(x,t) \neq 0, \\ -\min\left(|\mathcal{G}^{\mathrm{eq}}(x,t)|, \psi(x)\right)\mathrm{sgn}(\mathcal{G}^{\mathrm{eq}}(x,t)), & \dot{u}(x,t) = 0, \end{cases} \tag{7}$$

where $\psi(x)$ represents the frictional intensity and in terms of dimensional quantities:

$$\psi(x) = \frac{\mu N(L\,x)L}{\displaystyle\int_0^L \mu N(\xi)d\xi} \tag{8}$$

Finally, the boundary condition at $x=0$ can be written as

$$\frac{\partial u}{\partial x}(0,t) = F(t) \tag{9}$$

A few words concerning this nondimensionalization are in order. The frictional intensity satisfies the constraint:

$$\int_0^1 \psi(\xi)d\xi = 1, \tag{10}$$

which implies that the interface can support a maximum total (nondimensional) friction force of one. Also, the nondimensional length of the rod is now unity and time has been scaled by the period required for a longitudinal wave to traverse the interface. For typical applications, the joint is expected to extend over a very small interval and therefore the corresponding frequency will be much larger than typical frequencies associated with the forcing $\tilde{F}(\tilde{t})$. In terms of the nondimensional time t, the forcing frequencies are expected to be much less than one. Consequently, in many problems of interest the results do not depend on inertia, but the presence of inertial terms stabilizes the numerical results. Likewise, the dimensional forcing amplitude has been scaled by the force required to induce gross slip in the interface. Realistic amplitudes of the nondimensional forcing F are usually expected to be much less than one for most structural problems.

2 Continuum Results

A closed form solution to the continuum problem is unavailable in terms of elementary functions due to the nonlinear form of the interfacial friction, except for the case $\rho=0$, which we consider in the following. Under nonzero external force with $|F|<1$, there exists an interval $0\leqslant x<\ell(t)$ over which the interface exhibits slip. We further assume that for the non-slip region, that is, $x>\ell(t)$, the displacement profile of the rod is specified. In particular the quantities:

$$u(\ell(t),t), \quad \frac{\partial u}{\partial x}(\ell(t),t) \tag{11}$$

are known. Neglecting the inertial terms in the above-noted differential equation, over the interval of slip the deformation of the rod is described as

$$\frac{\partial^2 u}{\partial x^2}(x,t) = \psi(x)\ \mathrm{sgn}(\dot{u}(x,t)), \tag{12}$$

while the boundary condition at $x=0$ remains unchanged. Finally, because of the quasi-static approximation (and sign conventions) the velocity of slip has sign opposite to the term $\dot{F}$, that is:

$$\mathrm{sgn}\big(\dot{u}(0,t)\big) = \mathrm{sgn}\big(-\dot{F}(t)\big) \equiv \sigma(t) \tag{13}$$

It is worth noting that this model is closely related to the Menq model [15,16], which generalizes the frictional interface considered here to an elastoplastic shear layer at the interface. However, the present model incorporates spatial variations in the frictional intensity, whereas the Menq model considers only uniform pressure.

2.1 General Response. We consider general loading conditions applied at $x=0$. Integration of the governing equation and application of the boundary conditions yields

$$\frac{\partial u}{\partial x}(x,t) = F(t) + \sigma(t)\int_0^x \psi(s)ds, \tag{14a}$$

$$u(x,t) = u(\ell(t),t) - F(t)\big(\ell(t)-x\big) - \sigma(t)\int_x^{\ell(t)}\int_0^s \psi(\xi)d\xi\, ds. \tag{14b}$$

The integral of $\psi(x)$ will appear throughout the following analysis. Therefore to simplify the resulting expressions we identify

$$\int_0^x \psi(\xi)d\xi \equiv \Psi(x), \tag{15}$$

and we note that $\Psi(0)=0$. In addition, because of the nondimensionalization $\Psi(1)=1$ (see Eq. (10)). Physically, $\Psi(x)$ represents the total frictional loading over the interval $[0,x)$, provided the interval is uniformly slipping. Therefore the above-noted solutions are written as

$$\frac{\partial u}{\partial x}(x,t) = F(t) + \sigma(t)\Psi(x), \tag{16a}$$

$$u(x,t) = u(\ell(t),t) - F(t)\big(\ell(t)-x\big) - \sigma(t)\int_x^{\ell(t)}\Psi(s)ds. \tag{16b}$$

Evaluating Eq. (16*a*) at $x=\ell(t)$, we find

$$\frac{\partial u}{\partial x}(\ell(t),t) = F(t) + \sigma(t)\Psi(\ell(t)), \tag{17}$$

and evaluating Eq. (16*b*) at $x=0$, the displacement at the end of the rod is seen to be

$$u(0,t) = u(\ell(t),t) - F(t)\ell(t) - \sigma(t)\int_0^{\ell(t)}\Psi(s)ds \tag{18}$$

Moreover, the gradient of this terminal displacement with respect to the external load is simply:

$$\frac{d}{dF(t)}\big(u(0,t)\big) = -\ell(t) \tag{19}$$

2.2 Unidirectional Loading.

2.2.1 Slip Zone. On unidirectional loading from an initially undeformed state, we anticipate a region of slip of length $\ell_1(t)$, in which the deformation gradient at $x=0$ reflects the external force $F(t)=f_1(t)$. (In the following description the subscript "1" indicates the initial loading into the undeformed material.) At the other end of the slip region, the displacement is zero. Also, because the right-hand side abuts an undeformed domain, the force there must be zero also. Therefore:

$$\frac{\partial u_1}{\partial x}(0,t) = f_1(t), \quad u_1(\ell_1,t) = 0, \quad \frac{\partial u_1}{\partial x}(\ell_1,t) = 0, \tag{20}$$

where $u_1(x,t)$ is the displacement field resulting from the application of $F(t)=f_1(t)$.

Applying these boundary conditions we find

$$f_1(t) = -\sigma_1(t)\Psi(\ell_1(t)), \tag{21}$$

and the displacement at $x=0$ may be reduced to

$$u_1(0,t) = \sigma_1(t)\left[\ell_1(t)\Psi(\ell_1(t)) - \int_0^{\ell_1(t)}\Psi(s)ds\right] \tag{22}$$

with $\sigma_1(t)=\mathrm{sgn}(-\dot{f}_1(t))$.

As in Iwan's original work, we may relate the force-displacement curve to the model parameters, in this case the distribution of the normal force over the interface. Making repeated use of the chain rule for differentiation, we obtain

$$\frac{d}{df_1(t)}\big(u_1(0,t)\big) = -\ell_1(t), \tag{23}$$

as above, and differentiating again, using the definition of $\Psi(x)$:

$$\frac{d^2}{df_1^2}\big(u_1(0,t)\big) = \frac{1}{\psi(\ell_1(t))} \tag{24}$$

Knowledge of the force-displacement curve generated by unidirectional loading into undeformed material can be used to identify the tangential tractions acting at the interface. In practice, Eq. (24) typically yields reliable estimates for ψ only for large arguments since $\partial^2 u_1/\partial f_1^2$ is hard to evaluate meaningfully for small argument.

2.2.2 Power-Law Scalings of Dissipation. Say in the vicinity of $x=0$, the frictional intensity can be described as $\psi(x)=(\alpha+1)x^\alpha$, or equivalently:

$$\Psi(x) = x^\beta, \tag{25}$$

with $\alpha=\beta-1$. The solution to Eq. (16), subject to unidirectional loading is

$$\frac{\partial u_1}{\partial x}(x,t) = f_1(t)\left[1-\left(\frac{x}{\ell_1(t)}\right)^\beta\right], \tag{26a}$$

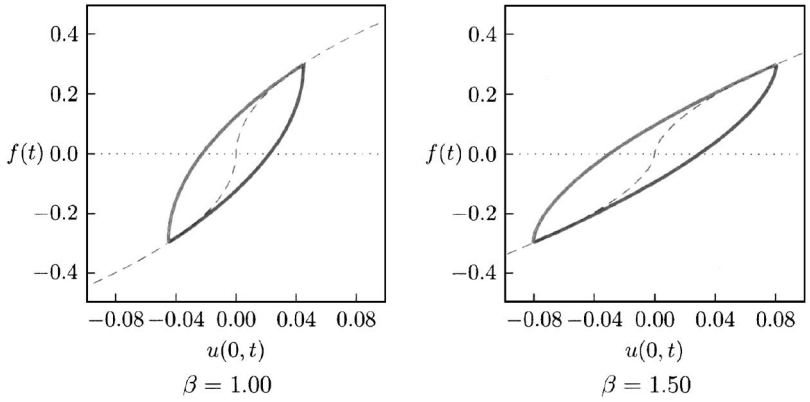


Fig. 3 Force-displacement curve. The dashed curve represents the force-displacement curve generated from loading into undeformed material [Eq. (28)]. In each panel the loading amplitude is 0.30.

$$u_1(x,t) = f_1(t)\ell_1(t)\left[\frac{1}{\beta+1}\left(1-\left(\frac{x}{\ell_1(t)}\right)^{\beta+1}\right)-\left(1-\frac{x}{\ell_1(t)}\right)\right], \tag{26b}$$

with

$$\ell_1(t) = (|f_1(t)|)^{1/\beta} \tag{27}$$

Consequently, the deformation of the rod can be written as

$$u_1(0,t) = \frac{\beta}{\beta+1} f_1(t)|f_1(t)|^{1/\beta} \tag{28}$$

(The above presented equations apply equally for tensile loading on the free surface.)

We now observe that the energy dissipation D due to small oscillatory loads is four times the dissipation due to a single monotonic loading of the same amplitude:

$$D = 4\int_0^{\ell_1(t)} u_1(x,t)\psi(x)dx = \frac{4\beta}{(\beta+1)(2\beta+1)} f_1(t)^{2+1/\beta}. \tag{29}$$

This last result is very interesting. If we associate the frictional intensity $\psi(x)$ with the normal traction in contact and recall that in Hertzian contact the normal traction goes as $\sqrt{x}$, then we set $\beta = 3/2$. In that case, the dissipation goes as $f_1^{8/3}$, which is reasonably close to the experimental values.

2.3 Cyclic Loading. Say that after we have pushed the rod by a force $F(t)=f_1^*$ to obtain slip out to a length ℓ_1^*, we then reduce the applied load, indicated as $F(t)=f_2(t)$. Reduction of the loading induces a new slip zone initiating at the free end of the rod. Notice that from Eq. (19), at the instant of this reversal the stiffness of the rod is infinite. Labeling the length of that new slip zone as $\ell_2(t)$, within that slip zone, the governing equation remains unchanged. However, $\sigma_2=-\sigma_1^*$ and the relevant boundary conditions for the deformation gradients become

$$\frac{\partial u_2}{\partial x}(0,t) = f_2, \tag{30a}$$

$$\frac{\partial u_2}{\partial x}(\ell_2(t),t) = \frac{\partial u_1^*}{\partial x}(\ell_2(t)) = f_1^* + \sigma_1^*\Psi(\ell_2(t)), \tag{30b}$$

where $u_1^*(x)$ represents the deformation profile of the rod at the point of reversal. The length $\ell_2(t)$ of the new slip zone can be found from Eq. (16*a*) subject to the above-noted boundary conditions, from which one finds

$$f_2(t) + \sigma_2\Psi(\ell_2(t)) = f_1^* + \sigma_1^*\Psi(\ell_2(t)) \tag{31}$$

Therefore, solving for the length of the new slip zone:

$$\ell_2(t) = \Psi^{-1}\left(\frac{f_1^* - f_2(t)}{2\sigma_2}\right), \tag{32}$$

provided of course that $\ell_2(t) < \ell_1^*$. Recalling Eq. (19), the force-displacement curve can be determined to be

$$\int_{f_1^*}^{f_2(t)} \left\{\frac{du(0,t)}{df(t)} = -\Psi^{-1}\left(\frac{f_1^* - f_2(t)}{2\sigma_2}\right)\right\} df_2(t), \tag{33}$$

$$u(0,t)\big|_{f_2(t)} = u(0,t)\big|_{f_1^*} - \int_{f_1^*}^{f_2(t)} \Psi^{-1}\left(\frac{f_1^* - f_2(t)}{2\sigma_2}\right) df_2(t). \tag{34}$$

For our power-law representation of the frictional intensity, we find

$$\ell_2(t) = \left(\frac{f_1^* - f_2(t)}{2\sigma_2}\right)^{1/\beta}, \tag{35}$$

and therefore the force-displacement curve takes the form

$$u(0,t)\big|_{f_2(t)} = u(0,t)\big|_{f_1^*} - \frac{2\beta}{\beta+1}\sigma_1^*\left(\frac{f_1^* - f_2(t)}{2\sigma_2}\right)^{1+1/\beta}. \tag{36}$$

Over a harmonic cycle of loading, we obtain hysteresis curves as illustrated in Fig. 3 for $\beta=1.00$ and $\beta=1.50$. In addition, the above-presented description of the force-displacement curve can be applied to more general loading conditions.

2.4 Relationship With Parallel-Series Models. We briefly review the force-displacement response of parallel-series system. Iwan showed ([10,11]) that for a parallel-series system (continuum of series elements in parallel) all having the same spring stiffness k, but a distribution $\tilde{\rho}(\tilde{\phi})$ of sliders of strength $\tilde{\phi}$, the instantaneous force-displacement relationship is

$$\tilde{F}(\tilde{u}(t)) = k\int_0^\infty \tilde{\rho}(\tilde{\phi})\left[\tilde{u}(t) - \tilde{x}_{\tilde{\phi}}(t)\right]d\tilde{\phi}, \tag{37}$$

where $\tilde{u}(t)$ is the imposed extension and $\tilde{x}_{\tilde{\phi}}(t)$ is the displacement of sliders of strength $\tilde{\phi}$ at time t.

The break-free force of the system is that which can cause gross slip to occur:

$$\bar{\phi} = \int_0^\infty \tilde{\phi}\tilde{\rho}(\tilde{\phi})d\tilde{\phi} \tag{38}$$

Equation (37) is nondimensionalized by dividing both sides by break-free force $\bar{\phi}$ and scaling the displacements by $\bar{\phi}/k$:

$$F(u(t)) = \int_0^\infty \rho(\phi)\left[u(t) - x_\phi(t)\right]d\phi, \tag{39}$$

where $\phi = \tilde{\phi}/\bar{\phi}$ and $\rho(\phi) = \bar{\phi}\tilde{\rho}(\bar{\phi}\phi)$. Note that the above also normalized ρ:

$$\int_0^\infty \phi\ \rho(\phi)d\phi = 1, \tag{40}$$

implying that gross slip occurs at an applied load of $F=1$.

As in the previous analysis, as this Iwan system is loaded from the underformed state, the slider displacements can be deduced to be as follows:

$$x_\phi(t) = u(t) - \phi \text{ for all } \phi < |u(t)|, \tag{41}$$

and

$$x_\phi(t) = 0 \text{ for all } \phi \geqslant |u(t)| \tag{42}$$

At this initial state we now break the above-presented force integral into two parts:

$$F(t) = \int_0^{u(t)} \phi\ \rho(\phi)d\phi + K_u u(t), \tag{43}$$

where

$$K_u = \int_{u(t)}^\infty \rho(\phi)d\phi \tag{44}$$

The quantity K_u represents the elastic stiffness arising from those elements which do not undergo slip during this loading.

Having achieved a displacement u^* with a force F^*, we now consider the force response as the system is reversed. For each spring-slider unit, characterized by the value of ϕ, the response is initially elastic as $u(t)$ withdraws from $u^*$to $u^*-2\phi$ Sliding takes place as $u(t)$ further reduces from $u^*-2\phi$. We can use the above observation to identify those Jenkins elements that are sliding as the system reaches a value of $u(t)$. This causes us to further divide Eq. (39):

$$F(u(t)) = \int_0^{[u^*-u(t)]/2} \phi\ \rho(\phi)d\phi + \int_{[u^*-u(t)]/2}^{u^*} \left[u(t) - (u^* - \phi)\right]\rho(\phi)d\phi + K_{u^*}u(t). \tag{45}$$

With the above-presented definition of F^* (Eq. (43)), this can be written as

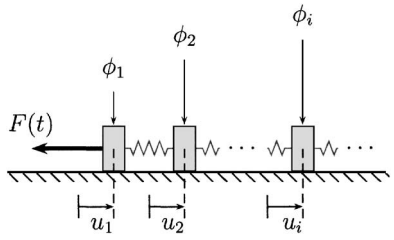


Fig. 4 Discrete model

$$F(u(t)) - F^* = -2\left[\int_0^{[u^*-u(t)]/2} \phi\ \rho(\phi)d\phi + (u(t) - u^*)\right]\left[K_{u^*} + \int_{[u^*-u(t)]/2}^{u^*} \rho(\phi)d\phi\right] \tag{46a}$$

$$= -2\left[\int_0^{[u^*-u(t)]/2} \phi\ \rho(\phi)d\phi + (u(t) - u^*)\right]\left[K_0 - \int_0^{[u^*-u(t)]/2} \rho(\phi)d\phi\right] \tag{46b}$$

Selecting $\rho(\phi) = (2+\chi)\phi^\chi$ for small values of ϕ, to provide power-law dissipation at low forces, Eq. (46*b*) becomes

$$F(u(t)) - F^* = \frac{2}{1+\chi}\left(\frac{u^* - u(t)}{2}\right)^{2+\chi} - K_0(u^* - u(t)) \tag{47}$$

Subtracting off the elastic term, this expression is similar to the force-displacement curve derived form the series-series continuum model, given in Eq. (36). In addition, the frictional dissipation per forcing cycle may be expressed as

$$D = \frac{4}{(1+\chi)(3+\chi)}(u^*)^{3+\chi} \tag{48}$$

As discussed in Segalman [11], in the micro-slip regime the gross displacement of the Iwan system is dominated by the elastic response, so that $u^* \sim F^*$. Therefore, comparing this equation for the parallel-series model with the equivalent expression for the continuum rod, given in Eq. (29), we find that the exponents of the distributions can be related as

$$\chi = \frac{1}{\beta} - 1, \quad \beta = \frac{1}{1+\chi}$$

Therefore, either the series-series or the parallel-series formulation may be used to generate power-law dissipation scalings arising from micro-slip. Segalman constructed a parallel-series Iwan model to have this power-law behavior at low and medium amplitude loads and other desirable properties at high loads [17].

3 Discrete Formulation

Although the above-described quasi-static continuum model can in principle be solved in closed form for arbitrary load histories, the requirement of keeping track of the slip reversal location can be cumbersome. This issue is obviated by considering the corresponding discrete series of Jenkins elements. The direct solution of the resulting nonlinear algebraic equations is notoriously awkward. We have chosen to regularize the problem by returning the intertial terms and solving a differential problem in time. Thus we are led to consider an n-degree-of-freedom discrete approximation to the continuum model given in Eq. (6) as shown in Fig. 4. The quasi-static result is recovered in the limit of forcing frequencies much less than one (the characteristic frequency of the system). This finite degree-of-freedom system corresponds to an

n-element series-series Iwan model. The discretization is obtained by a collocation method with quadratic comparison functions and in the following, the displacements are represented as

$$u_i(t) = u(x_i,t) \quad \text{with} \quad x_i = \frac{i}{n}\left(i - \frac{1}{2}\right), \quad i = 1,\ldots,n \tag{49}$$

With this, the discrete equations of motion become

$$\begin{aligned}
\frac{1}{n}\ddot{u}_1 + n(u_1 - u_2) &= \frac{G_1}{n} - F(t),\\
\frac{1}{n}\ddot{u}_2 + n(-u_1 + 2u_2 - u_3) &= \frac{G_2}{n},\\
&\vdots\\
\frac{1}{n}\ddot{u}_j + n(-u_{j-1} + 2u_j - u_{j+1}) &= \frac{G_j}{n},\\
&\vdots\\
\frac{1}{n}\ddot{u}_n + n\left(-\frac{4}{3}u_{n-1} + 4u_n\right) &= \frac{G_n}{n},
\end{aligned} \tag{50}$$

The description of the friction force G_i follows from the continuum model, i.e.,

$$G_i(t) = \begin{cases} -\phi_i \operatorname{sgn}(\dot{u}(t)), & \dot{u}_i(t) \neq 0,\\ -\min\left(|\mathcal{G}_i^{\text{eq}}(t)|, \phi_i\right)\operatorname{sgn}(\mathcal{G}_i^{\text{eq}}(t)), & \dot{u}_i(t) = 0, \end{cases} \tag{51}$$

where $\phi_i = \phi(x_i)$ and $\mathcal{G}_i^{\text{eq}}$ represents the force required to maintain static equilibrium on the i-th element. In particular, $\mathcal{G}_1^{\text{eq}}$ reduces to

$$\mathcal{G}_1^{\text{eq}} = n^2(u_1 - u_2) + nF(t), \tag{52}$$

so that in this discrete model, the boundary ($i=1$) element cannot support an equilibrium state if

$$\left|(u_1 - u_2) + \frac{F(t)}{n}\right| > \frac{\phi_1}{n^2}, \tag{53}$$

and in the undeformed configuration ($u_i=0$) the boundary element begins to slip if $F(t) \geqslant \phi_1/n$. As n increases, the minimum force necessary to induce initial slip decreases to zero. A similar model was considered in [18].

Our focus in this work is to characterize the effect of distributed friction, as represented by the number of discrete masses in the model, on the dynamic response of the interface as well as the energy dissipated by the frictional forces. To accomplish this, we investigate the response of the above-presented systems to external forces of the form:

$$F(t) = \alpha \sin(\tau), \quad \tau = \omega t + \tau(0) \tag{54}$$

and over one complete forcing cycle, τ varies over 2π regardless of the forcing frequency. The following numerical investigations are restricted to $0 < \omega \ll 1$ and $0 < \alpha \ll 1$, which correspond to physically common values of these nondimensional parameters. The results shown in the following are often represented as a function of τ rather than time. Finally, the instantaneous (nondimensional) power dissipated by the frictional forces can be expressed as

$$P_{\text{loss}}(t) = \frac{1}{n}\sum_{i=1}^{n} G_i(t)\dot{u}_i(t), \tag{55}$$

while the frictional work is calculated as the integral of this power over time.

The numerical results were obtained using a fourth-order Runge-Kutta method, with an integration step size of $\Delta t = 0.001/n$. Coulomb friction is incorporated through a regularized model proposed by Quinn [19]. To verify the accuracy of the results, the step size was reduced by a factor of 100 and the resulting simulated behavior showed no qualitative change (Fig. 4).

In Fig. 5(*a*) the dynamical behavior of the joint is shown for $\omega = 0.25$, $\alpha = 0.25$, and $n = 64$. In the figure the displacement of each element is traced versus τ, where the leftmost curve represents the terminal end of the chain (to which the forcing is applied). The asymptotic response is shown for one forcing cycle; the transient behavior (not shown) was removed by integrating over five forcing cycles before showing the results. Although this external load is one-quarter of that expected to initiate slip in the 1-dof model, a slip zone is seen at the end of the chain. However, only the 25 elements nearest the boundary of the rod experience slip—the majority of the joint remains stationary throughout the forcing cycle. In addition, the evolution of the slip interface is consistent with that predicted from the quasi-static continuum model described in the previous section.

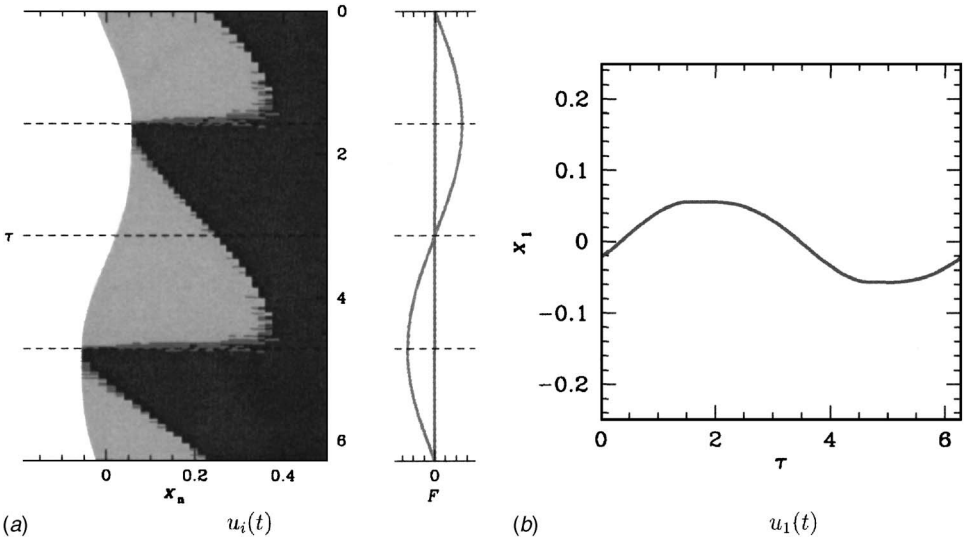


Fig. 5 Interfacial behavior with $n=64$ ($\omega=0.25$, $\alpha=0.25$). The displacements have been marked according to the slip velocity—for the lightest points $|\dot{u}_i| > v_0$ and for the darker points $v_0 > |\dot{u}_i| > v_0^2/n$. The velocity of the darkest points is $v_0^2/n > |\dot{u}_i|$, with $v_0 = 2\omega(\Delta t) = 7.8125 \times 10^{-6}$. The slip velocity is in the same direction as that of the end of the interface.

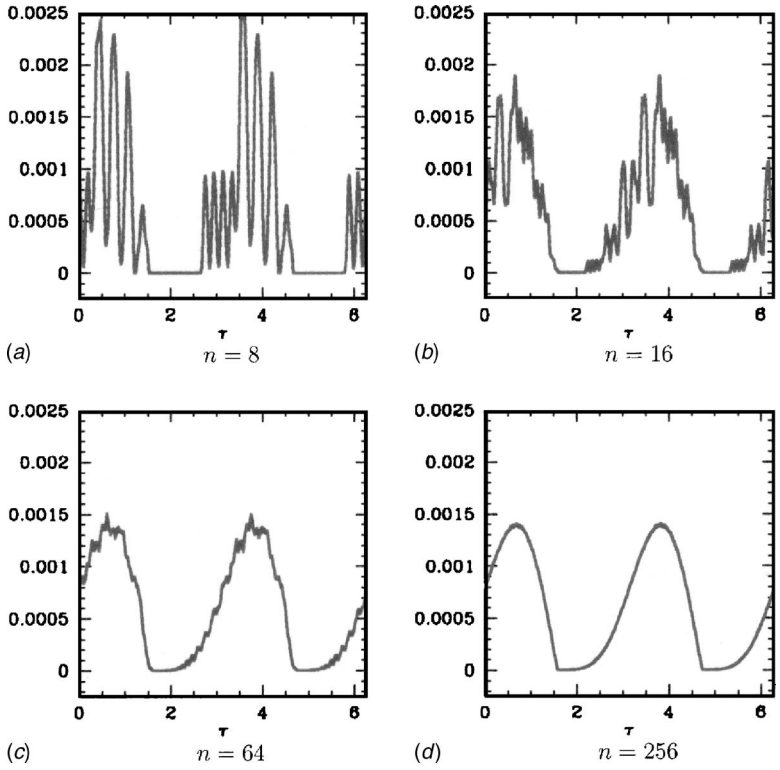


Fig. 6 Power dissipated over one steady-state cycle, $t=\omega\tau(\omega=0.25, \alpha=0.25)$

The power dissipated by the frictional forces is sensitive to the number of elements in the model. In Fig. 6 the power dissipated is shown over one cycle of motion for $n=8$, 16, 64, and 256. In the simulation the rod was initially at rest in an undeformed state, and then integrated for five cycles of the external forcing to remove the transient behavior before generating the observed figures. As the number of elements is decreased, the trace of the power dissipated becomes less smooth. For $n=256$ the trace of the instantaneous power dissipation is relatively smooth, while as the degree of freedom is decreased to $n=8$, power is dissipated in relatively short, large amplitude bursts, indicating significant stick-slip motions.

Surprisingly, the total work done per unit cycle is rather insensitive to the number of elements, provided a sufficient number are chosen to admit a time-dependent state, i.e., $n > 1/\alpha$. In Fig. 7 the frictional work per unit cycle is shown as the number of elements varies from $n=2$ to $n=32$. As n increases beyond 32, the work remains fairly constant. For example, at $n=32$, the work is found to be $W=-0.015233$ while as n is increased to 256, the frictional work becomes $W=-0.015070$, which represents a 1.1% change for an eightfold increase in model size. Moreover, the dissipation predicted by the quasi-static continuum model is $W=-0.014882$ (see Eq. (29)).

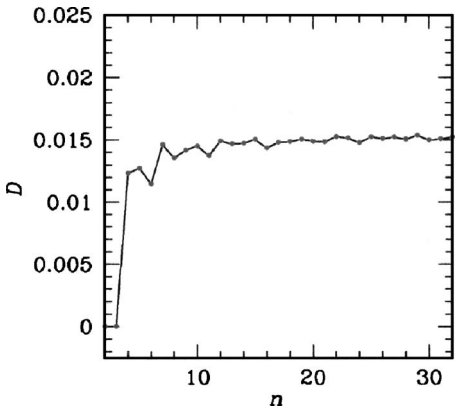


Fig. 7 Frictional dissipation per unit cycle as n varies ($\omega=0.25, \alpha=0.25$)

In contrast to variations in n, the work done by the frictional forces is strongly dependent on α. In Fig. 8(a), the frictional work is shown as α is varied, holding $\omega=0.25$ and $n=256$ fixed. As illustrated in the figure, the dissipation is well-represented by a power law and for this simulation the slope of this curve is approximately $m=2.672$. The numerical value closely approximates the predicted value of $m=8/3$ obtained from the quasi-static continuum model.

The work done by frictional forces can also be evaluated as the forcing frequency varies. In Fig. 8(b) both n and α are held fixed while $0.01 < \omega < 0.25$. Although some dependence on ω is seen ($m=0.002$), it is slight compared with the variation seen in W as α is varied. This implies that for forcing frequencies much less than the lowest characteristic frequency of the joint, the dissipation predicted by the quasi-static continuum model closely approximates the response predicted by the model with inertia. Moreover, the dissipation predicted by the discrete model rapidly approaches that of the quasi-static continuum model as the number of degrees-of-freedom (n) increases, provided that n is sufficient to resolve the slipping at the free end of the chain.

4 Discussion and Conclusions

We have considered an elastic rod sliding on a frictional surface subject to an external force across the structure. When the rod is assumed massless, the continuum partial differential equations can be solved exactly, based only on the amplitude of external force across the rod. This solution can then be used to predict the frictional dissipation and generate force-displacement curves for

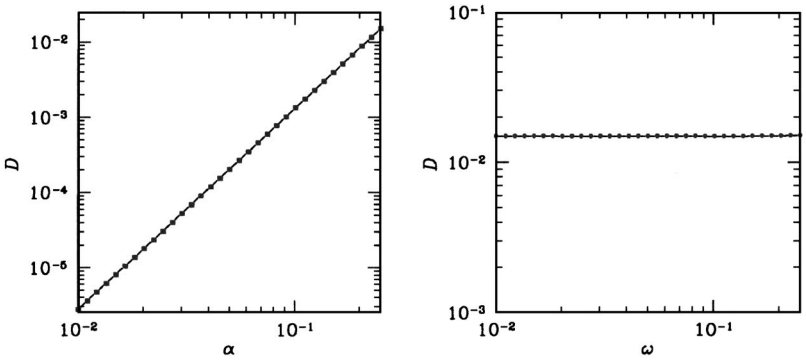


Fig. 8 Frictional work per unit cycle (n=256). The quantity m, represents the slope of this curve, as determined from linear regression

comparison with experimental data. When mass is included, the partial differential equations are discretized to develop a finite degree-of-freedom model.

The above-noted system is appropriate for the modeling of dissipation induced by mechanical joints, provided the forcing amplitude and frequency are limited. If the degrees-of-freedom of the discrete model is sufficiently large (roughly twice the value necessary to allow for micro-slip, c f., Eq. (53)), the predictions of the massless continuum model agree with those of the discrete formulation for the frictional dissipation per unit cycle. Moreover, in both formulations the dissipation per cycle is seen to depend sensitively on the distribution of the normal load over the rod, while for the parameter ranges applicable to joint dynamics it is insensitive to the frequency of the external loading.

Finally, this model yields power-law behavior in the dissipation per forcing cycle. Specifically, if the normal traction varies as x^{α} near the free edge of the rod, then the frictional dissipation per cycle scales with the forcing amplitude to a power of $(3+2\alpha)/(1+\alpha)$. As a special case, if the normal traction follows a Hertzian distribution, so that $\alpha=1/2$, the dissipation per cycle scales with the forcing amplitude to the 8/3 power which is similar to experimentally observed results [6,7]. These results indicate that the model under consideration is a reasonable first step toward the development of physically based reduced-order models for the incorporation of interfaced-induced dissipation in larger structural models.

References

[1] Mindlin, R. D., and Deresiewicz, H., 1953, "Elastic Spheres in Contact Under Varying Oblique Forces," ASME J. Appl. Mech., **20**, p. 327.

[2] Johnson, K. L., 1985, *Contact Mechanics*, Cambridge University Press, Cambridge.

[3] Dohner, J. L., 2001, "On the Development of Methodologies for Constructing Predictive Models of Structures with Joints and Interfaces," Tech. Rep. SAND2001-0003P, Sandia National Laboratories.

[4] Berger, E. J., Begley, M. R., and Mahajani, M., 2000, "Structural Dynamic Effects on Interface Response: Formulation and Simulation Under Partial Slipping Conditions," ASME J. Appl. Mech., **67**, pp. 785–792.

[5] Heinstein, M., and Segalman, D. J., 2001, "Bending Effects in the Energy Dissipation of Bolted Interfaces," presented at the 2001 ASME Design Engineering Technical Conferences, Pittsburgh, PA. September 9–12, DETC2001/VIB-21517.

[6] Gregory, D. L., Smallwood, D. O., Coleman, R. G., and Nusser, M. A., 1999, "Experimental Studies to Investigate Damping in Frictional Shear Joints," *Proceedings of the 70th Shock and Vibration Symposium.*

[7] Smallwood, D. O., Gregory, D. L., and Coleman, R. G., 2000, "Damping Investigations of a Simplified Frictional Shear Joint," Tech. Rep. SAND2000-1929C, Sandia National Laboratories.

[8] Goodman, L. E., and Brown, C. B., 1962, "Energy Dissipation in Contact Friction: Constant Normal and Cyclic Tangential Loading," ASME J. Appl. Mech., **29**, p. 17.

[9] Iwan, W. D., 1966, "A Distributed-element Model for Hysteresis and Its Steady-state Dynamic Response," ASME J. Appl. Mech., **33**, pp. 893–900.

[10] Iwan, W. D., 1967, "On a Class of Models for the Yielding Behavior of Continuous and Composite Systems," ASME J. Appl. Mech., **89**, pp. 612–617.

[11] Segalman, D. J., 2001, "An Initial Overview of Iwan Modeling for Mechanical Joints," Tech. Rep. SAND2001-0811, Sandia National Laboratories.

[12] Tworzydlo, W. W., Cecot, W., Oden, J. T., and Yew, C. H., 1998, "Computational Micro- and Macroscopic Models of Contact and Friction: Formulation, Approach and Applications," Wear, **220**, pp. 113–140.

[13] Dankowicz, H., 1999, "On the Modeling of Dynamic Friction Phenomena," Z. Angew. Math. Mech., **79** (6), pp. 399–409.

[14] Ruina, A. L., 1983, "Slip Instability and State Variable Friction Laws," J. Geophys. Res., **88** (B12), pp. 10359–10370.

[15] Menq, C.-H., Bielak, J., and Griffin, J. H., 1986, "The Influence of Microslip on Vibratory Response, Part I: A New Microslip Model," J. Sound Vib., **107**, pp. 279–293.

[16] Menq, C.-H., Griffin, J. H., and Bielak, J., 1986, "The Influence of Microslip on Vibratory Response, Part II: A Comparison with Experimental Results," Sound Vib., **107**, pp. 295–307.

[17] Segalman, D. J., 2002, "A Four-parameter Iwan Model for Lap-type Joints," Tech. Rep. SAND2002-3828, Sandia National Laboratories.

[18] Quinn, D. D., 2001, "Distributed Friction and Microslip in Mechanical Joints with Varying Degrees-of-freedom," presented at the 2001 ASME Design Engineering Technical Conferences, Pittsburgh, PA. September 9-12, DETC2001/VIB-21514.

[19] Quinn, D. D., 2004, "A New Regularization of Coulomb Friction," ASME J. Vibr. Acoust., **126**, pp. 391–397.

Mechanics of Bimaterial Interface: Shear Deformable Split Bilayer Beam Theory and Fracture

Jialai Wang
Assistant Professor
Department of Civil Engineering,
North Dakota State University,
Fargo, ND 58105-5285
e-mail: jialai.wang@ndsu.edu

Pizhong Qiao[1]
Associate Professor
Advanced Materials and Structures Research Group,
Department of Civil Engineering,
The University of Akron,
Akron, OH 44325-3905
e-mail: qiao@uakron.edu

A novel split beam model is introduced to account for the local effects at the crack tip of bi-material interface by modeling a bi-layer composite beam as two separate shear deformable beams bonded perfectly along their interface. In comparisons with analytical two-dimensional continuum solutions and finite element analysis, better agreements are achieved for the present model, which is capable of capturing the local deformation at the crack tip in contrast to the conventional composite beam theory. New solutions of two important issues of cracked beams, i.e., local buckling and interface fracture, are then presented based on the proposed split bi-layer shear deformable beam model. Local buckling load of a delaminated beam considering the root rotation at the delamination tip is first obtained. By considering the root rotation at the crack tip, the buckling load is lower than the existing solution neglecting the local deformation at the delamination tip. New expressions of energy release rate and stress intensity factor considering the transverse shear effect are obtained by the solution of local deformation based on the novel split beam model, of which several new terms associated with the transverse shear force are present, and they represent an improved solution compared to the one from the classical beam model. Two specimens are analyzed with the present model, and the corresponding refined fracture parameters are provided, which are in better agreement with finite element analysis compared to the available classical solutions. [DOI: 10.1115/1.1978920]

1 Introduction

A bi-material or bi-layer system is a common configuration in structural applications, and it is usually manufactured by monolithically forming the two parts together. Interlaminar delamination is one of the most popular failure modes in this type of layered structures. A split bi-material beam is resulted from the delamination of a bi-layer structure. Requirement of effective analysis of the split beam is encountered frequently, such as the delamination buckling of laminated composites [1], data reduction technique of fracture tests [2], crack identification [3], and vibration analysis of delaminated structures [4]. Conventional analysis of split beam in the literature simulates the cracked segment of the beam as two separate beams and the uncracked segment as one composite beam. At the connection of the cracked and uncracked segments where a joint is formed to connect three beams, the cross sections of the three beams are assumed to remain in one plane and perpendicular to the mid-plane of the virgin beam. This conventional model neglects the elastic deformation of the joint, such as the root rotation at the crack tip [5] and thus forms a rigid connector. Extra errors are introduced, and unfavorable results are obtained by this conventional split beam model, such as the unconservative loading of delamination buckling of composites [6], under-evaluated energy release rate of fracture [5], and rough dynamic analysis at the crack tip [3]. The reason for this unfavorable feature of the available split beam model is explained by the nature of the assumptions used in the beam model, which are unable to describe the severe local deformation at the crack tip of the split beam. In the cases where the local deformation is of no interest or of little importance, the conventional split beam model is applicable; however, in the cases where the local deformation is significant, a new and improved model is required to account for the deformation at the crack tip.

In this study, for the convenience of application, a closed-form solution of a split bi-material beam is presented first, of which the local deformation at the crack tip is captured by modeling the split beam as a bi-layer shear deformable beam system. Compared with the conventional composite beam theory, the proposed novel split bi-layer beam model predicts more accurate stress and deformation distributions near the crack tip. Then, the present solution is applied to solve two important issues of cracked bi-layer beams: local buckling analysis and interface fracture analysis.

2 Split Bi-Layer Beam Theory

Consider an interface fracture problem of Fig. 1, where a crack lies along the straight interface of the top beam "1" and bottom beam "2" with thickness of h_1 and h_2, respectively. Two beams are made of homogeneous, orthotropic, and elastic materials, with the orthotropy axes along the coordinate system. The length of the uncracked region L is relatively large compared to the thickness of the whole beam h_1+h_2. This configuration essentially represents a crack tip element [7], a small element of a split beam, on which the generic loads are applied, as already determined by a global beam or beam analysis. It is assumed that the lengths of cracked and uncracked parts of the beam are relatively large compared to the bi-layer beam thickness. Therefore, a beam theory can be used to model the behavior of the top and bottom layers. Plane strain formulation is used in this study.

[1]Corresponding author. Tel.: +1-330-972-5226; fax: +1-330-972-6020.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, August 18, 2003; final revision; February 2, 2005. Associate Editor. K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

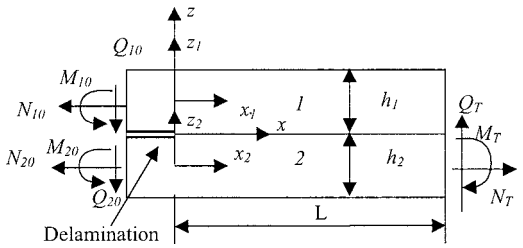


Fig. 1 A bi-layer beam system under generic loadings

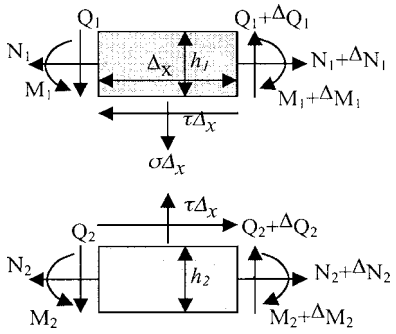


Fig. 2 Overall equilibrium of bi-layer beam system

In the conventional split beam analysis [8], this problem is modeled as three classical beams: the top beam in the cracked region, the bottom beam in the cracked region, and a single composite beam of the whole uncracked region. As shown in Ref. [9], it is not appropriate to model the uncracked portion of the laminate using a single beam element in order to capture the actual shear deformation. To this end, Wang and Qiao [10] recently modeled the uncracked region as two separate beams: the top beam "1" and bottom beam "2" instead of only a single beam. These two beams are perfectly bonded along their interface to keep the continuity of displacement; while the two beams in cracked region deform separately. The first-order shear deformation theory or Reissner-Mindlin plate theory is used to account for the transverse shear deformation.

2.1 Analysis of a Bi-layer Beam System. According to Reissner-Mindlin plate theory, the deformations of two beams have the form:

$$U_i(x_i,z_i) = u_i(x_i) + z_i\phi_i(x_i), \tag{1}$$

$$W_i(x_i,z_i) = w_i(x_i) \tag{2}$$

where subscript $i=1, 2$, representing the beams 1 and 2 in Fig. 1, respectively. x_i and z_i are the local coordinates in beam i. The strains in these two beams are given as

$$\epsilon_i^0 = \frac{du_i}{dx}, \quad \kappa_i = \frac{d\phi_i}{dx}, \quad \gamma_{xy}^j = \phi_i + \frac{dw_i}{dx} \tag{3}$$

The constitutive equations are written in the conventional way as

$$\begin{pmatrix} N_i \\ M_i \end{pmatrix} = \begin{pmatrix} C_i & 0 \\ 0 & D_i \end{pmatrix} \begin{pmatrix} \dfrac{du_i}{dx} \\ \dfrac{d\phi_i}{dx} \end{pmatrix}, \quad Q_i = B_i\left(\phi_i + \frac{dw_i}{dx}\right), \tag{4}$$

where N_i, Q_i, and M_i are, respectively, the resulting axial force, transverse shear force, and bending moment per unit width of beam i; C_i, B_i, and D_i are the axial, shear, and bending stiffness coefficients of the beam i under the plane strain condition and given as

$$C_i = \frac{E_{xxi}h_i}{1-\nu_{xzi}\nu_{zxi}}, \quad B_i = \kappa G_{xzi}h_i, \quad D_i = \frac{E_{xi}h_i^3}{12(1-\nu_{xzi}\nu_{zxi})} \quad (i=1,2),$$

where $E_{xi}, G_{xzi}, \nu_{xzi}$, and ν_{zxi} are the longitudinal modulus, transverse shear modulus, and Poisson's ratios of beam i. κ is the shear correction coefficient chosen as 5/6 for a rectangular section in this study. For a plane stress problem, the corresponding stiffness/compliance parameters must be used.

Considering a typical infinitesimal isolated body of bi-layered beam system (Fig. 2), the following equilibrium equations are established:

$$\frac{dN_1(x)}{dx} = \tau(x), \frac{dN_2(x)}{dx} = -\tau(x),$$

$$\frac{dQ_1(x)}{dx} = \sigma(x), \quad \frac{dQ_2(x)}{dx} = -\sigma(x), \tag{5}$$

$$\frac{dM_1(x)}{dx} = Q_1(x) - \frac{h_1}{2}\tau(x), \quad \frac{dM_2(x)}{dx} = Q_2(x) - \frac{h_2}{2}\tau(x),$$

where $N_1(x)$ and $N_2(x), Q_1(x)$ and $Q_2(x), M_1(x)$, and $M_2(x)$ are the axial forces, transverse shear forces, and bending moments in layers 1 and 2, respectively; h_1 and h_2 are the thickness of layers *1* and *2*, respectively; $\sigma(x)$ and $\tau(x)$ are the interface normal (peel) and shear stresses, respectively.

At the interface of the bi-layer beam system, the displacement continuity requires

$$u_1 - \frac{h_1}{2}\phi_1 = u_2 + \frac{h_2}{2}\phi_2, \tag{6}$$

$$w_1 = w_2 \tag{7}$$

Differentiating Eq. (6) with respect to x once and then substituting it into the first equation of Eq. (4) yield

$$\frac{N_1(x)}{C_1} - \frac{h_1}{2}\frac{M_1(x)}{D_1} = \frac{N_2(x)}{C_2} + \frac{h_2}{2}\frac{M_2(x)}{D_2} \tag{8}$$

Considering the global equilibrium conditions in Fig. 1, we have

$$N_1 + N_2 = N_{10} + N_{20} = N_T \tag{9}$$

$$Q_1(x) + Q_2(x) = Q_{10} + Q_{20} = Q_T, \tag{10}$$

$$M_1 + M_2 + N_1\frac{h_1+h_2}{2} = M_{10} + M_{20} + N_{10}\frac{h_1+h_2}{2} + Q_T x = M_T(x), \tag{11}$$

where N_{10}, N_{20}, Q_{10}, Q_{20}, and M_{10}, M_{20} are the applied axial forces, transverse shear forces, and bending moments, respectively, at the crack tip (see Fig. 1); N_T, Q_T, and M_T are the total resulting applied axial force, transverse shear force, and bending moment of the bi-layer beam system about the mid-plane of the layer 2 (Fig. 1), respectively.

For simplicity, there are no distributed loads applied to the composite beam, and therefore, N_T and Q_T are constant and M_T is a linear function of x. Substituting Eqs. (9) and (11) into Eq. (8) gives

$$\eta N_1 - \xi M_1 = \frac{N_T}{C_2} + \frac{h_2}{2D_2}M_T, \tag{12}$$

where

$$\xi = \frac{h_1}{2D_1} - \frac{h_2}{2D_2}, \tag{13}$$

$$\eta = \frac{1}{C_1} + \frac{1}{C_2} + \frac{(h_1 + h_2)h_2}{4D_2} \tag{14}$$

2.2 Conventional Composite Beam Theory. Conventional composite beam theory is used most widely in the literature to analyze the bi-layer beam, in which the cross sections of two sublayers are assumed to remain in the same plane after deformation, i.e.,

$$\phi_1(x) = \phi_2(x) \tag{15}$$

Differentiating Eq. (15) with respect to x gives

$$\frac{M_1}{D_1} = \frac{M_2}{D_2} \tag{16}$$

By substituting Eq. (16) into Eq. (12) and considering Eq. (11), the governing equation of composite beam model is obtained as

$$\left(\left(\frac{1}{D_1} + \frac{1}{D_2}\right)\eta + \frac{(h_1 + h_2)}{2D_2}\xi\right)N_1(x) = F(x), \tag{17}$$

where

$$F(x) = \left(\left(\frac{1}{D_1} + \frac{1}{D_2}\right)\frac{h_2}{2D_2} + \frac{\xi}{D_2}\right)M_T + \left(\frac{1}{D_1} + \frac{1}{D_2}\right)\frac{N_T}{C_2}$$

The resultant forces of the beam are thereby obtained as

$$N_{1C} = \frac{(D_1 + D_2)h_2 + \xi D_1 D_2}{2D_2(D_1 + D_2)\eta + \xi D_1 D_2(h_1 + h_2)}M_T + \frac{2(D_1 + D_2)}{C_2(2(D_1 + D_2)\eta + (h_1 + h_2)D_1\xi)}N_T,$$

$$Q_{1C} = \left(\left(\frac{\eta}{\xi} + \frac{h_1}{2}\right)\frac{(D_1 + D_2)h_2 + D_1 D_2\xi}{2D_2(D_1 + D_2)\eta + \xi D_1 D_2(h_1 + h_2)} - \frac{1}{\xi}\frac{h_2}{2D_2}\right)Q_T,$$

$$M_{1C} = \frac{\eta}{\xi}N_{1C} - \frac{1}{\xi}\left(\frac{N_T}{C_2} + \frac{h_2}{2D_2}M_T\right) \tag{18}$$

$$N_{2C} = N_T - N_{1C}, \quad Q_{2C} = Q_T - Q_{1C}, \quad M_{2C} = M_T - M_{1C} - \frac{h_1 + h_2}{2}N_{1C}$$

The subscript C is used to refer to the conventional composite beam solution.

2.3 Shear Deformable Bi-layer Beam Theory. Although the conventional composite beam model is very simple and widely used, it is fairly approximate in nature due to neglecting the local deformation at the crack tip. To account for this deformation, a shear deformable bi-layer beam theory [10] has recently been developed, in which the restraint on the rotation in Eq. (15) is released, i.e., each sub-layer in the virgin beam portion rotates separately. By using Eqs. (5)–(7), the governing equation for this model is obtained by Wang and Qiao [10] as

$$\left(\frac{1}{B_1} + \frac{1}{B_2}\right)\left(\eta + \frac{\xi h_1}{2}\right)\frac{d^2N_1(x)}{dx^2} + \left(\left(\frac{1}{D_1} + \frac{1}{D_2}\right)\eta + \frac{(h_1 + h_2)}{2D_2}\xi\right)N_1(x) = F(x) \tag{19}$$

Compared to the governing equation of conventional composite beam model [see Eq. (17)], Eq. (19) has an extra second-order differential term, and the solution of Eq. (17) is a particular solution of Eq. (19). Therefore, the resultant forces of the sub-beams are given as [10]

$$N_1(x) = ce^{-kx} + N_{1C}, \quad Q_1(x) = -\left(\frac{\eta}{\xi} + \frac{h_1}{2}\right)cke^{-kx} + Q_{1C},$$

$$M_1(x) = \frac{\eta}{\xi}ce^{-kx} + M_{1C},$$

$$N_2(x) = -ce^{-kx} + N_{2C}, \quad Q_2(x) = \left(\frac{\eta}{\xi} + \frac{h_1}{2}\right)cke^{-kx} + Q_{2C},$$

$$M_2 = -\left(\frac{\eta}{\xi} + \frac{h_1 + h_2}{2}\right)ce^{-kx} + M_{2C}, \tag{20}$$

where

$$k = \sqrt{\frac{B_1 B_2(2(D_1 + D_2)\eta + D_1(h_1 + h_2)\xi)}{D_1 D_2(B_1 + B_2)(2\eta + h_1\xi)}}, \quad c = \frac{(2M + h_1 N)\xi}{h_1\xi + 2\eta} \tag{21}$$

where k is the decay rate determined by the geometries of the specimen and properties of materials, and

$$M = M_{10} - M_{1C}|_{x=0}, \quad N = N_{10} - N_{1C}|_{x=0}, \quad Q = Q_{10} - Q_{1C}|_{x=0} \tag{22}$$

N_{1C}, M_{1C}, and Q_{1C} are given by Eq. (18). Equation (20) shows that the shear deformable bi-layer beam solution consists of two parts: the exponential terms which describe the local effect and the conventional composite beam solution terms which are dominant away from the vicinity of the crack tip. Equations (20) and (21) also show that the exponential terms in the present bi-layer beam solutions are only determined by M, N, and Q.

It should be pointed out that when the bi-layer beam system is homogeneous and symmetric about the interface, $\xi = 0$ and moment M_1 and axial force N_1 decouple (Eq. (12)). This special case was solved by Bruno and Greco [9]. In the present model, we only need to rewrite the expression of moments and shear forces.

2.4 Rotational Flexible Joint Deformation Model. Considering the constitutive equation of the beam given by Eq. (4), we have

$$\phi_1(L) - \phi_1(x) = \int_x^L \frac{M_1}{D_1}dx = \int_x^L \frac{\left(\frac{\eta}{\xi}ce^{-kx} + M_{1C}\right)}{D_1}dx = \frac{\eta}{\xi}\frac{ce^{-kx}}{kD_1} + \int_x^L \frac{M_{1C}}{D_1}dx \tag{23}$$

Note that

$$\int_x^L \frac{M_{1C}}{D_1}dx = \phi_C(L) - \phi_C(x), \tag{24}$$

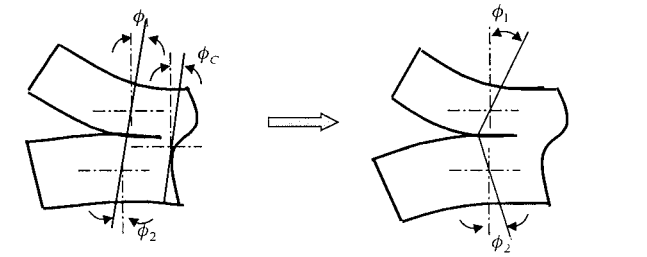
where ϕ_C is the rotation angle of the uncracked portion based on the conventional composite beam model, i.e., both the top and bottom beams have the same rotation. When L is very large (see Fig. 1), we have

$$\phi_1(L) = \phi_2(L) = \phi_C(L) \tag{25}$$

Therefore, combining Eqs. (23) and (24) yields

$$\phi_1(x) = \phi_{1C}(x) - \frac{\eta}{\xi}\frac{ce^{-kx}}{kD_1} = \phi_{1C} - \frac{e^{-kx}}{kD_1}\frac{(2M + h_1 N)}{h_1\xi + 2\eta}\eta \tag{26}$$

In the same way, the deformation at the crack tip ($x = 0$ in the given coordinates in Fig. 1) is then obtained by



(*a*) Conventional rigid joint (*b*) Present rotational flexible joint

Fig. 3 Joint (crack tip) deformation model

$$\begin{pmatrix} u_1(0) \\ u_2(0) \\ \phi_1(0) \\ \phi_2(0) \\ w_1(0) \\ w_2(0) \end{pmatrix} = \begin{pmatrix} u_{1C}(0) \\ u_{2C}(0) \\ \phi_{1C}(0) \\ \phi_{2C}(0) \\ w_{1C}(0) \\ w_{2C}(0) \end{pmatrix} - \frac{\xi}{h_1\xi + 2\eta} \begin{pmatrix} \frac{1}{C_1 k} \\ \frac{-1}{C_2 k} \\ \frac{\eta}{D_1 k \xi} \\ \frac{-1}{D_2 k}\left(\frac{\eta}{\xi} + \frac{h_1 + h_2}{2}\right) \\ \frac{\eta}{D_1 k^2 \xi} - \frac{1}{B_1}\left(\frac{\eta}{\xi} + \frac{h_1}{2}\right) \\ \frac{\eta}{D_1 k^2 \xi} - \frac{1}{B_1}\left(\frac{\eta}{\xi} + \frac{h_1}{2}\right) \end{pmatrix} \times (h_1\ 2) \begin{pmatrix} N \\ M \end{pmatrix} \tag{27}$$

Equation (27) gives a new solution of deformation of the joint (crack tip), which is different from the conventional rigid joint as shown in Fig. 3. In the conventional bi-layer beam analysis [Fig. 3(*a*)], the compatibility conditions require the rotations of three-beam (i.e., two cracked beams and one uncracked bi-layer beam) segments at the joint to be the same. Thus, the joint is rotationally rigid, leading to the underestimated crack tip deformation. Equation (27) allows the relative rotation between the top and bottom beams [as shown in Fig. 3(*b*)] and therefore describes a rotationally flexible joint (crack tip).

2.5 Verification and Comparison. To evaluate the present split bi-layer beam model, especially the crack tip deformation model of Eq. (27), a double cantilever beam (DCB) specimen (Fig. 4) used widely in the interface fracture is examined numerically by finite element analysis (FE). The specimen is modeled by a commercial finite element package ANSYS as a two-dimensional (2D) problem with 8-node isoparametric plane element (PLANE82). The specimen has a symmetric geometry with $a/h_1=16, a/L=1$ and $h_1=h_2$. Two scenarios of materials are considered in the finite element analysis: (a) the top and bottom beams are made of the same materials ($E_1=E_2=1, \nu_1=\nu_2=0.3, G_1=G_2=E_1/7$) simulating a symmetric DCB configuration, and (b) the top and bottom beams are made of different materials ($E_1=1, \nu_1=\nu_2=0.3, G_1=E_1/7, E_2=5E_1, G_2=5G_1$) representing an asymmetric DCB specimen. Figure 4 compares the deformations

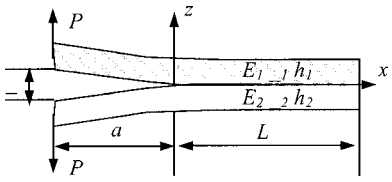
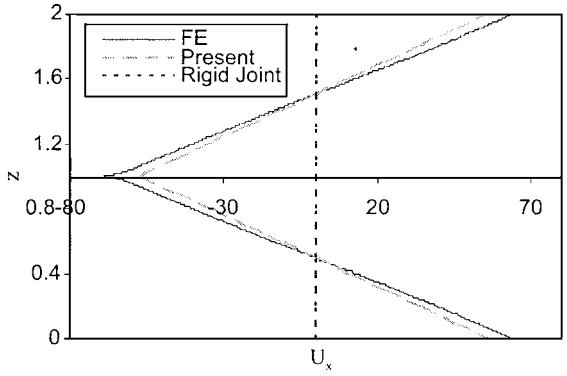
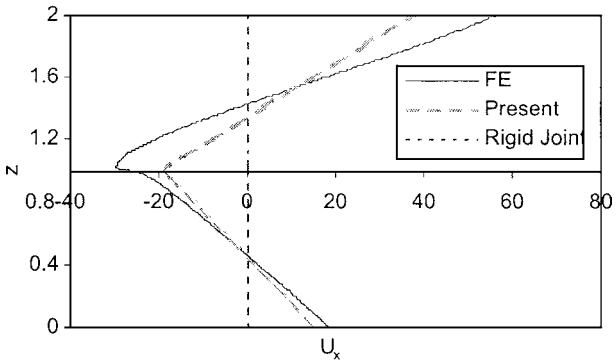


Fig. 4 Double cantilever beam (DCB) specimen



(*a*) Symmetric DCB Specimen



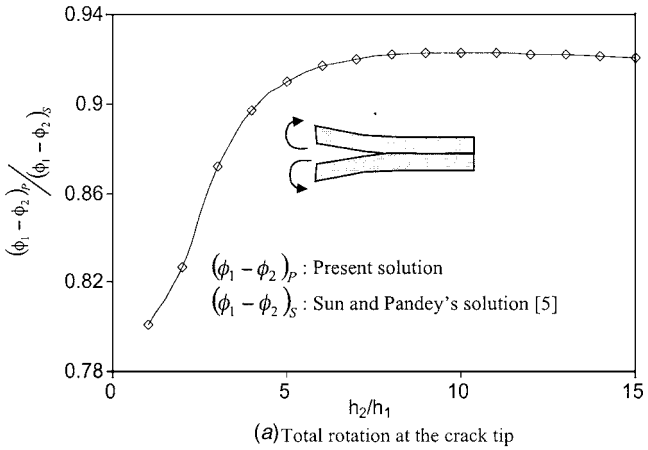
(*b*) Asymmetric DCB Specimen

Fig. 5 Comparisons of joint deformation at the crack tip

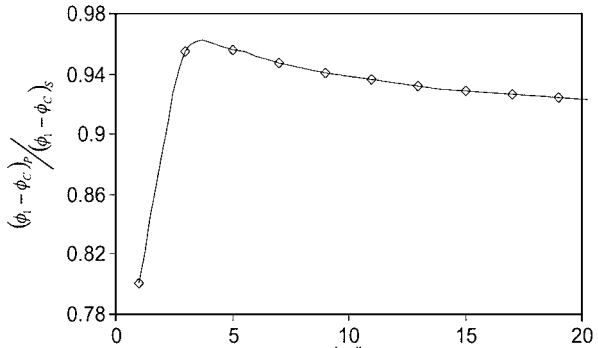
calculated by the FE analysis, present solution [Eq. (27)], and conventional composite beam model at the crack tip of the given two DCB specimens. A considerable rotation at the crack tip is found by the finite element analysis in contrast with the zero rotation predicted by the conventional composite beam model for both the symmetric and asymmetric DCB specimens. Figure 5 shows that a significant amount of root rotation at the crack tip is captured by the present analysis for both the symmetric DCB specimens (88% of FE result) and asymmetric DCB specimen (70% of FE result); while the conventional composite beam analysis just simply excluded this rotation and thus produces a rigid joint model (Fig. 5).

An approximate 2D elasticity solution of crack tip rotation was obtained by Sun and Pandey [5] for a DCB specimen subjected to the opposite bending moments. As a comparison, the crack tip rotations obtained by the present study [Eq. (27)] and by Sun and Pandey's solution [5] are shown in Fig. 6. It is observed that the present solution is about 80% to 92% of the approximate elastic solution for a large range of thickness ratio. Note that the solution in Ref. [5] was only valid for isotropic, homogeneous materials; while the present solution of Eq. (27) is applicable to a general orthotropic bi-layer beam system.

Compared to the exact finite element solution or 2D approximation, the present solution seems to underestimate the rotation at the crack tip. This may be due to the assumption of the first-order displacement field in Eqs. (1) and (2) adopted in this study. The actual displacement field is more complex at the crack tip as demonstrated by the finite element solution (Fig. 5), and therefore, a higher order beam theory may be required to better describe the deformation of the crack tip in order to improve the present solution. However, the simplicity of the analytical solution will be lost.



(*a*) Total rotation at the crack tip



(*b*) Root rotation of beam 1

Fig. 6 Comparison of crack tip rotation

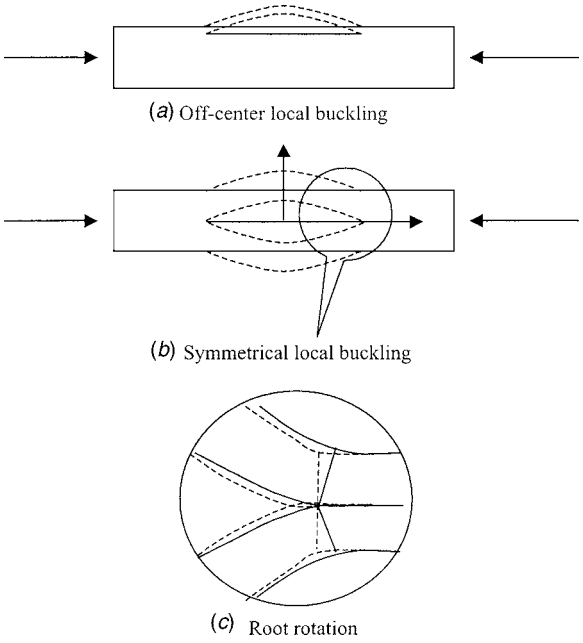


Fig. 7 Local buckling of laminated composites

3 Application Examples

In this section, two application problems (i.e., local delamination buckling and interface fracture) using the proposed split bilayer model are solved to show the significant effect of local deformation at the crack tip. The closed-form solutions based on the present model are given for the convenience of reference since they are not available in the literature.

3.1 Local Delamination Buckling. Local delamination buckling is a common failure mode of laminated composites. Typical analytical solution of local delamination assumes the clamped boundary conditions at the delamination tip [1]. The deformation at the delamination tip is ignored in the assumption as shown in Fig. 7(*c*). Vizzini and Lagace [11] used a beam on elastic foundation model to study the effect of the deformation conditions at the delamination tip on the delamination buckling via Rayleigh-Ritz energy method. Here, a new closed-form solution of local delamination buckling considering the delamination tip rotation is presented. According to Eq. (4), the governing equation of local buckling can be written as

$$\frac{d^3w_1}{dx^3}+\frac{P_1}{D_1\left(1-\dfrac{P_1}{B_1}\right)}\frac{dw_1}{dx}=0 \tag{28}$$

Due to symmetry of the delamination to its center line, the solution of Eq. (28) can be written as

$$w_1=c_1(\cos(\lambda_1 x)-1),\quad \phi_1=-c_1\lambda_1(\sin(\lambda_1 x)), \tag{29}$$

where

$$\lambda_1^2=\frac{P_1}{D_1\left(1-\dfrac{P_1}{B_1}\right)} \tag{30}$$

Using the forces at the delamination tip [Fig. 7(*c*)], the root rotations at the delamination tip $x=a$ can be obtained by Eq. (27) as

$$\phi_1=-\frac{2\eta}{kD_1(h_1\xi+2\eta)}M=-\frac{2\eta}{k(h_1\xi+2\eta)}\frac{d\phi_1}{dx} \tag{31}$$

Substituting Eq. (31) into Eq. (29), we have

$$\tan(\lambda_1 a)=-\frac{2\eta}{k(h_1\xi+2\eta)}\lambda_1 \tag{32}$$

By solving Eq. (32), λ_1 is obtained and then the local delamination buckling can be written as

$$P_1=\frac{\lambda_1^2D_1}{1+\dfrac{\lambda_1^2D_1}{B_1}} \tag{33}$$

Normalized by Euler value, Eq. (33) becomes

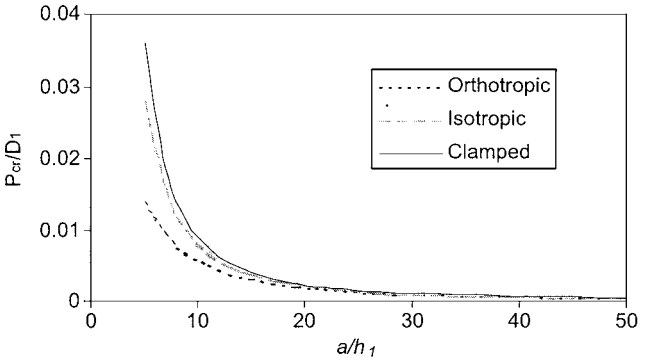
$$\bar{P}_1=\frac{\mu^2}{1+s\mu^2}, \tag{34}$$

where

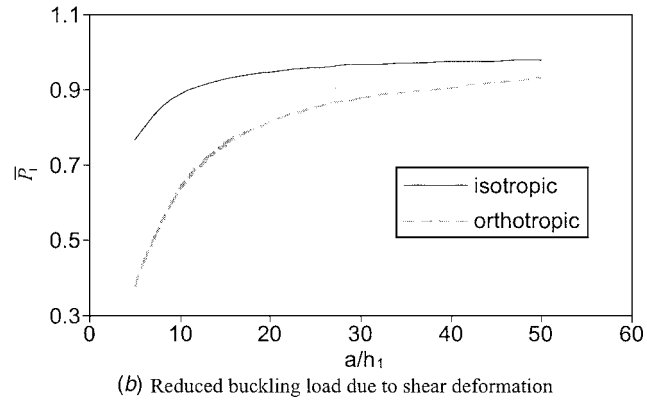
$$\bar{P}_1=\frac{P_1}{P_E},\quad P_E=\frac{\pi^2D_1}{a^2},$$

$$s=\frac{\pi^2D_1}{B_1a^2}=\frac{\pi^2}{12(1-\nu_{13}\nu_{31})}\frac{E_1}{\kappa G_1}\left(\frac{h_1}{a}\right)^2,\quad \mu=\frac{\pi\lambda_1}{a}$$

The effect of delamination tip rotation can be shown by a numerical example presented by Fig. 8 (where $h_1=1$ for the convenience of calculation). Both isotropic and orthotropic materials are considered. The orthotropic material in the calculation is obtained by reducing the shear modulus of the isotropic materials by 10 times. Figure 8 shows that the conventional clamped boundary condition overestimates the local buckling load, especially when the shear modulus of the material is relatively low, which is the



(*a*) Effect on buckling loading



(*b*) Reduced buckling load due to shear deformation

Fig. 8 End rotation effect on local buckling loading

case for the laminated composites. In other words, the local buckling loading evaluated by clamped boundary condition is not conservative.

3.2 Interface Fracture. Interface cracking is one of common failure modes in multi-layered structures. Typical examples include delamination of composites laminates, debonding of adhesive joints, and decohesion of thin film from substrates. The shear deformation in the cracked and uncracked regions is not considered in the existing model since the classical beam or plate theory was basically used [12,13]. As a result, the ERR is always underestimated by this method as evidenced by Davidson and Sundararaman [14]. As a matter of fact, the shear deformation effect on the ERR for anisotropic materials with relative low transverse shear modulus (e.g., polymer composite laminates) is even more significant as shown in Bruno and Greco [9], where the portion contributed by the shear deformation was found to be more than half of the total ERR for an orthotropic double cantilever beam specimen. Therefore, it is necessary to account for the shear deformation in computation and prediction of the ERR, especially when the materials with relatively low transverse shear modulus and moderate thickness are concerned. Notable effort to incorporate the shear deformation into the ERR was made by Bruno and Greco [9]. However, the closed-form solutions of the ERR are only obtained for certain simple configurations. In this study, the new solution of interface fracture problem shown in Fig. 1 considering transverse shear are obtained based on the new split beam bi-layer model developed in Sec. 2.

3.2.1 Interface Fracture Solution With Transverse Shear Deformation. J-integral [15] can be used to calculate the energy release rate of interface fracture problem shown in Fig. 1:

$$J=\frac{1}{2}\left(\frac{N_1^2}{C_1}+\frac{N_2^2}{C_2}+\frac{Q_1^2}{B_1}+\frac{Q_2^2}{B_2}+\frac{M_1^2}{D_1}+\frac{M_2^2}{D_2}-2Q_1\phi_1-2Q_2\phi_2\right)\Bigg|_{x=L}^{x=0}, \tag{35}$$

which can be further simplified as

$$J=\frac{1}{2}(C_N N^2+C_M M^2+C_{MN}MN+C_Q Q^2-2Q(\phi_1(0)-\phi_2(0))) \tag{36}$$

where

$$C_N=\frac{1}{C_1}+\frac{1}{C_2},\quad C_M=\frac{1}{D_1}+\frac{1}{D_2},\quad C_Q-\frac{1}{B_1}+\frac{1}{B_2},\quad C_{MN}=\frac{h_1+h_2}{D_2} \tag{37}$$

It can be seen that the ERR depends on not only the three loading parameters but also the relative rotation at the joint (crack tip). Equation (36) clarifies the major argument made by Li et al. [16] on the effects of shear on interface fracture in the layered materials. In their study, Li et al. [16] pointed out that the crack tip deformation only affects the shear components of the ERR. Two terms of transverse shear Q are present in Eq. (36) which represent the transverse shear components of the total ERR of the interface fracture: (a) the stable-state part $(1/B_1+1/B_2)Q^2/2$ which is the contribution of the shear deformation in the cracked region, and (b) $-Q(\phi_1(0)-\phi_2(0))$ which is the contribution of the shear deformation in the uncracked region of a bi-layer beam. Only the latter part (part (b)) of the transverse shear component is dependent of the local deformation at the crack tip, and more exactly, only dependent on the relative rotation of two sublayers at the crack tip. It should be noted that the local deformation is not a contributor to the ERR physically. As a matter of fact, the relative rotation in Eq. (36) is a reflection of the complex local displacement field, and it disappears once the conventional composite beam model is used.

Substituting the solution of rotation at the crack tip (Eq. (27)), we have

$$J=\frac{1}{2}(C_N N^2+C_Q Q^2+C_M M^2+C_{MN}MN+C_{NQ}NQ+C_{MQ}MQ) \tag{38}$$

where

$$C_{NQ}=k\left(\frac{1}{B_1}+\frac{1}{B_2}\right)h_1,\ C_{MQ}=2k\left(\frac{1}{B_1}+\frac{1}{B_2}\right)$$

Equation (38) is similar to the expression of the ERR obtained by Suo and Hutchinson [12], except that three new terms associated with Q from the contribution of transverse shear to the total ERR are introduced. In Eq. (38), $C_Q Q^2/2$, which is due to the transverse shear deformation in the cracked region, is present; while $C_{NQ}NQ/2$ and $C_{MQ}MQ/2$ arise from the shear deformation in the uncracked region. The closed-form solutions obtained by Bruno and Greco [9] for geometrically symmetric plates can be easily derived using the present formula of Eq. (38) by substituting the specific loading and laminate properties.

Based on the split beam model developed in Sec. 2, the ERR can be decomposed as modes I and II [17]:

$$G_{\mathrm{I}}=\frac{1}{2}\left(\frac{1}{B_1}+\frac{1}{B_2}\right)\left(Q+k\left(M+\frac{h_1 N}{2}\right)\right)^2=\frac{1}{2}\delta_Q Q_C^2, \tag{39}$$

$$G_{\mathrm{II}}=\frac{1}{h_1\xi+2\mu}(\xi M-\eta N)^2=\frac{1}{2}\delta_N N_C^2, \tag{40}$$

where

$$\delta_N = \frac{h_1}{2}\xi + \eta = \frac{1}{C_1} + \frac{1}{C_2} + \frac{h_1^2}{4D_1} + \frac{h_2^2}{4D_2}, \quad \delta_Q = \frac{1}{B_1} + \frac{1}{B_2},$$

$$N_C = \frac{2(M\xi - N\eta)}{h_1\xi + 2\eta}, \quad Q_C = -Q - k\left(M + \frac{h_1 N}{2}\right)$$

While following the procedure described by Suo and Hutchinson [12], the stress intensity factor K is obtained as [10]

$$K = K_1 + iK_2 = \left(\sqrt{C_N}N - ie^{i\gamma_1}\sqrt{C_M}M - ie^{i\gamma_2}\sqrt{C_Q}Q\right)\frac{p}{\sqrt{2}}h_1^{-i\varepsilon}e^{i\omega} \tag{41}$$

where P is defined in the same way as in Suo and Hutchinson [12] and

$$\sin(\gamma_1) = \frac{C_{MN}}{2\sqrt{C_M C_N}}, \quad \sin(\gamma_2) = \frac{C_{NQ}}{2\sqrt{C_N C_Q}} \tag{42}$$

It is convenient to use the combination $Kh_1^{i\varepsilon}$ as suggested by Rice [18] and define

$$Kh_1^{i\varepsilon} = K_{\mathrm{I}} + iK_{\mathrm{II}} = |K|e^{i\psi} \tag{43}$$

Then the individual stress intensity factors are given by

$$K_{\mathrm{I}} = \frac{p}{\sqrt{2}}\left(\sqrt{C_N}N\cos(\omega) + \sqrt{C_M}M\sin(\omega+\gamma_1) + \sqrt{C_Q}Q\sin(\omega+\gamma_2)\right), \tag{44}$$

$$K_{\mathrm{II}} = \frac{p}{\sqrt{2}}\left(\sqrt{C_N}N\sin(\omega) - \sqrt{C_M}M\cos(\omega+\gamma_1) - \sqrt{C_Q}Q\cos(\omega+\gamma_2)\right) \tag{45}$$

The phase angle ψ defined in Eq. (43) is given by

$$\psi = \tan^{-1}\left(\frac{\sqrt{C_N}N\sin(\omega) - \sqrt{C_M}M\cos(\omega+\gamma_1) - \sqrt{C_Q}Q\cos(\omega+\gamma_2)}{\sqrt{C_N}N\cos(\omega) + \sqrt{C_M}M\sin(\omega+\gamma_1) + \sqrt{C_Q}Q\sin(\omega+\gamma_2)}\right), \tag{46}$$

where the bi-material constant ε was defined in Ref. [10].

Note that we can rewrite the global decomposition expressed by Eqs. (39) and (40) in term of stress intensity factor:

$$K_G = \frac{p}{\sqrt{2}}\left(\sqrt{\delta_Q}Q_c + i\sqrt{\delta_N}N_c\right) = |K_G|e^{i\psi_G}, \tag{47}$$

where

$$\psi_G = \arctan\left(\frac{\sqrt{\delta_N}N_C}{\sqrt{\delta_Q}Q_C}\right) = \arctan\left(\sqrt{\frac{G_{\mathrm{II}}}{G_{\mathrm{I}}}}\right) \tag{48}$$

Equation (41) can be rewritten as

$$K = \left(\sqrt{\delta_Q}Q_C + i\sqrt{\delta_N}N_C\right)\frac{p}{\sqrt{2}}h_1^{-i\varepsilon}e^{i\omega_1} = K_G e^{i\omega_1}h_1^{-i\varepsilon} \tag{49}$$

By comparing Eqs. (49) and (41), we obtain

$$\omega_1 = \omega + \gamma_2 - \frac{\pi}{2} \tag{50}$$

By following the same definition of the stress intensity factor as in Eq. (43), a relation between the local and global stress intensity factor is obtained as

$$Kh_1^{i\varepsilon} = |K|e^{i\psi} = K_I + iK_{II} = K_G e^{i\omega_1} \tag{51}$$

It appears that the local SIF can be obtained by shifting the global SIF phase angle by ω_1. Thereby, ω_1 can be regarded as the shift angle from the global to local decomposition, which represents both the geometry and material mismatches along the interface as shown in Eq. (50). Based on the numerical results of Davidson et al. [7], the phase shift angle ω_1 can be easily obtained. It is interesting to point out that ω_1 and the mode mix parameter Ω defined by Davidson et al. [7] were recently proved to be identical [17].

3.2.2 Fracture Parameters for Interface Fracture Specimens. Based on the above-presented analysis, fracture parameters of several typical fracture test specimens are derived and presented in this section. Comparisons have been conducted in this section, which show the improvement of accuracy offered by the present study.

3.2.2.1 Asymmetric Double Cantilever Beam. An asymmetric double cantilever beam specimen (Fig. 4) is a simple but effective specimen for measurement of polymer/polymer and polymer/non-polymer bi-material interface fracture toughness. The ERRs from the global decomposition are expressed as

$$G_{\mathrm{I}} = \frac{1}{2}\delta_Q(1+ka)^2P^2, \quad G_{\mathrm{II}} = 2\delta_N\xi^2\frac{P^2a^2}{(h_1\xi+2\eta)^2}, \tag{52}$$

$$G = \frac{1}{2}\delta_Q(1+ka)^2P^2 + 2\delta_N\xi^2\frac{P^2a^2}{(h_1\xi+2\eta)^2} \tag{53}$$

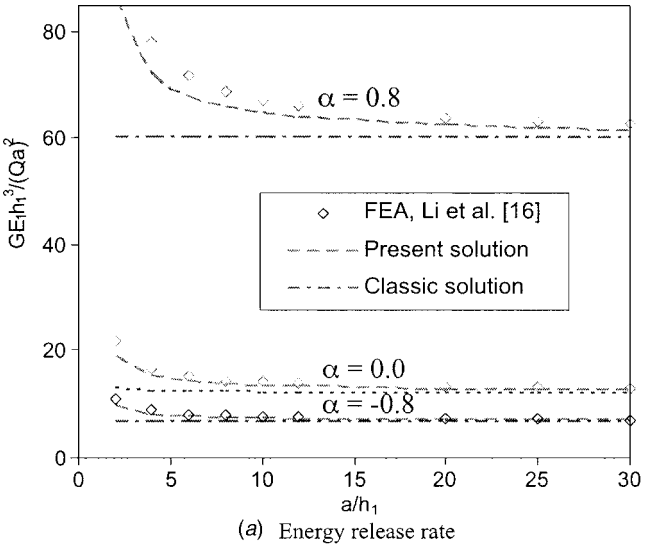
The SIFs from the local decomposition are expressed as

$$K_{\mathrm{I}} = -\frac{p}{\sqrt{2}}\left(\sqrt{C_M}a\sin(\omega+\gamma_1) + \sqrt{C_Q}\sin(\omega+\gamma_2)\right)P, \tag{54}$$

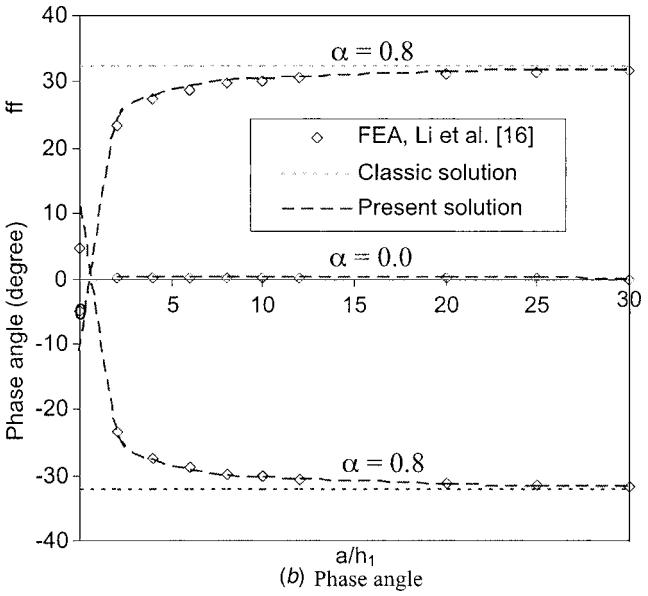
$$K_{\mathrm{II}} = \frac{p}{\sqrt{2}}\left(\sqrt{C_M}a\cos(\omega+\gamma_1) + \sqrt{C_Q}\cos(\omega+\gamma_2)\right)P \tag{55}$$

$$\psi = -\tan^{-1}\left(\frac{\sqrt{C_M}a\cos(\omega+\gamma_1) + \sqrt{C_Q}\cos(\omega+\gamma_2)}{\sqrt{C_M}a\sin(\omega+\gamma_1) + \sqrt{C_Q}\sin(\omega+\gamma_2)}\right) \tag{56}$$

Improvements both in the ERR and phase angle provided by the present solution Eqs. (52) and (56) are demonstrated in Fig. 9. Figure 9 compares the classical solution and the present solution [i.e., Eqs. (53) and (56)] with the finite element analysis (FEA) results of Li et al. [16] which is assumed as exact solution in this study. A value of 0.3 for Poisson's ratio is chosen for both the materials, and an excellent agreement is achieved when compared to the FEA results. Classic solution underestimates the total ERR significantly when the crack length is small due to the reason of neglecting the transverse shear effect; while the present solution is much closer to the FEA due to consideration of transverse shear in the calculation [Fig. 9(*a*)]. As discussed in Sec. 2.5, the present bi-layer beam model underestimates the actual root rotation at the crack tip. Consequently, the present solution of ERR is a bit lower than the exact solution (finite element solution) as demonstrated by Fig. 9. It can also be observed that the transverse shear has a significant effect on the phase angle when the crack length is small. The present method can capture this feature while the clas-



(*a*) Energy release rate



(*b*) Phase angle

Fig. 9 Comparison of fracture parameters for ADCB specimen ($h_1 = h_2$)

sic solution simply ignores it.

If the top and bottom beams have same geometry and material, the specimen becomes the widely used double cantilever beam (DCB) specimen. In this case $\xi=0$ and therefore:

$$G_{\mathrm{I}} = \frac{1}{2}\delta_Q(1+ka)^2P^2 = \frac{P^2a^2}{D_1} + \frac{P^2}{B_1} + \frac{2P^2a}{\sqrt{B_1D_1}}, \tag{57}$$

$$G_{\mathrm{II}} = 0 \tag{58}$$

The above equation is the same as the solution obtained by Bruno and Greco [9]. Compared with the conventional first-order shear deformation solution of G_{I}, the third term in Eq. (57) is new, and it represents the contribution of transverse shear in the uncracked portion of beam to the total mode-I ERR of the specimen.

3.2.2.2 Single Leg Bending (SLB) Specimen. A comprehensive study of this test was performed for a bi-material specimen by Davidson and Sundararaman [14] in which a closed-form solution of ERR was obtained by the crack tip element (CTE) analysis for isotropic materials. When testing orthotropic materials, however, the existing closed-form solution is not quite as accurate as it is for isotropic one due to the more pronounced effects of transverse shear deformation. Therefore, a need exists for more accurate evaluation of fracture parameters of the test specimen.

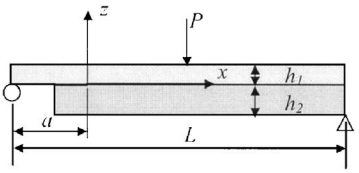


Fig. 10 Single leg bending (SLB) specimen

By applying the results of this study to the single leg bending (SLB) or modified ENF specimen (Fig. 10), the ERRs, SIF are given as

$$G_{\mathrm{I}} = \frac{1}{2}\delta_Q\left(1+\frac{k}{a}\right)^2\left(1-\left(\frac{\eta}{\xi}+\frac{h_1}{2}\right)A_M+\frac{1}{\xi}\frac{h_2}{2D_2}\right)^2(Pa/2)^2, \tag{59}$$

$$G_{\mathrm{II}} = \frac{1}{2}\frac{1}{\delta_N}\left(\frac{h_1}{2D_1}Pa/2\right)^2, \tag{60}$$

$$K_{\mathrm{I}} = \frac{p}{2\sqrt{2}}\left(\sqrt{C_N}A_M\cos(\omega)+\sqrt{C_M}\left(-1+\frac{\eta}{\xi}A_M-\frac{1}{\xi}\frac{h_2}{2D_2}\right)\sin(\omega+\gamma_1)+\sqrt{C_Q}\frac{1}{a}\left(-1+\left(\frac{\eta}{\xi}+\frac{\eta}{\xi}\right)A_M-\frac{1}{\xi}\frac{h_2}{2D_2}\right)\sin(\omega+\gamma_2)\right)Pa, \tag{61}$$

$$K_{\mathrm{II}} = \frac{p}{2\sqrt{2}}\left(\sqrt{C_N}N\sin(\omega)-\sqrt{C_M}\left(-1+\frac{\eta}{\xi}A_M-\frac{1}{\xi}\frac{h_2}{2D_2}\right)\cos(\omega+\gamma_1)-\sqrt{C_Q}\left(-1+\left(\frac{\eta}{\xi}+\frac{h_1}{2}\right)A_M-\frac{1}{\xi}\frac{h_2}{2D_2}\right)\cos(\omega+\gamma_2)\right)Pa \tag{62}$$

Figure 11 compares the present solution of ERR with classic solution (CTE) and finite element solution obtained by Davidson and Sundararaman [14] for three types of bi-material interface homogeneous, aluminum/niobium and glass/epoxy, which essentially "span" the range of the material property mismatch ratios which one would expect to encounter in practical applications. Details of material properties and specimen geometry are given in Davidson and Sundararaman's paper [14] and therefore omitted here for brevity. The ERRs calculated by the present solution and CTE are normalized by the finite element solution which is assumed as exact solution in this study. Figure 11 shows that CTE underestimates a maximum 13% of ERR by neglecting the transverse shear deformation. A significant improvement of accuracy has been achieved by present solution in which the ERR is only underestimated by a maximum of 3% due to the same reason aforementioned.

4 Conclusions

By modeling a bi-material split composite beam under plane strain deformation subjected to general loading as two separate shear deformable beams bonded perfectly along their interface, a novel split beam model which accounts for the local effects is presented in this study. The relatively close comparison of the present model with finite element analysis demonstrates that the proposed split beam model can capture the local deformation at the crack tip of bi-material interface, and therefore, it is more accurate than the conventional composite beam theory. This unique feature of the new split beam model is then applied to solve the local delamination buckling and interface fracture problems of bi-layer beams. It has been shown that the local buckling load considering the root rotation at the delamination tip is lower

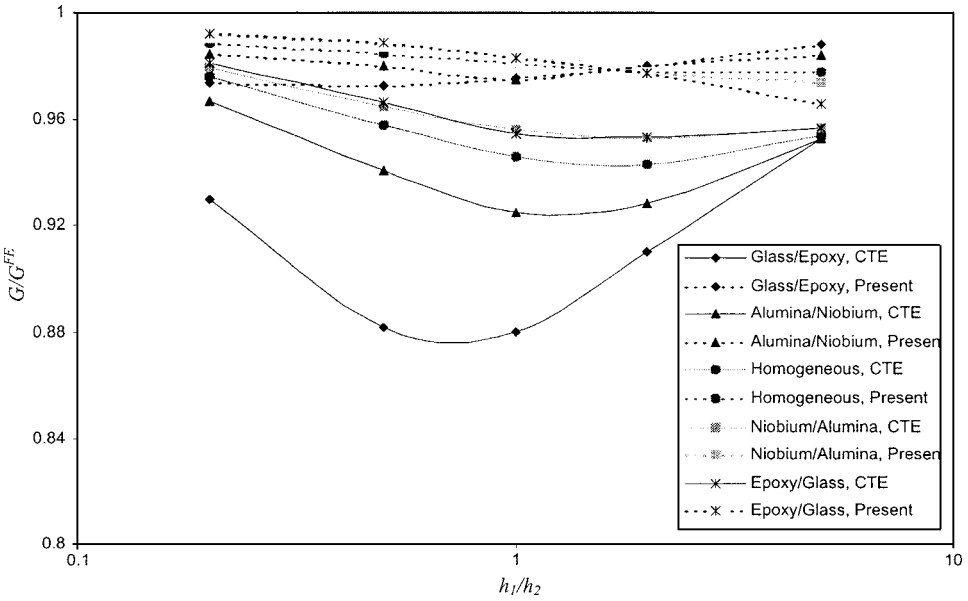


Fig. 11 Effect of shear deformation on ERR of SLB for $a/(h_1+h_2)=8.33$ and $a/L=0.5$

than the existing solution of which the root rotation is ignored. The transverse shear effect is successfully incorporated into the energy release rate (ERR) and stress intensity factor (SIF) based on the novel split beam model. Two conventional interface fracture specimens are analyzed by the present interface fracture solution. New ERR and SIF are obtained, of which the transverse shear deformation is taken into consideration. Compared with the existing classical solution, the present explicit solution shows a better agreement with finite element solution due to the inclusion of the transverse shear effect. This suggests that the contribution of transverse shear deformation is significant and therefore should be included in the refined analysis, especially for the specimens with relatively lower transverse shear modulus and moderate thickness.

Acknowledgment

This study was partially supported by the National Science Foundation (CMS-0002829 under program director Dr. Ken P. Chong).

References

[1] Chai, H., Babcock, C. D., and Knauss, W. G., 1981, "One Dimensional Modeling of Failure in Laminated Plates by Delamination Buckling," Int. J. Solids Struct. **17**, pp. 1069–1083.
[2] Wang, J., and Qiao, P., 2003, "Novel Beam Analysis of End-Notched Flexure Specimen for Mode-II Fracture," Eng. Fract. Mech., **71**, pp. 219–231.
[3] Farris, T. N., and Doyle, J. F., 1993, "A Global-Local Approach to Length-wise Cracked Beams: Dynamic Analysis," Int. J. Fract., **60**, pp. 147–156.
[4] Brandinelli, L., and Massabo, R., 2003, "Free Vibration of Delaminated Beam-Type Structures With Crack Bridging," Compos. Struct., **61**, pp. 129–142.
[5] Sun, C. T., and Pandey, R. K., 1994, "Improved Method for Calculating Strain Energy Release Rate Based on Beam Theorem," AIAA J., **32**, pp. 184–189.
[6] Shu, D., and Mai, Y., 1993, "Buckling of Delaminated Composites Re-Examined," Compos. Sci. Technol., **47**, pp. 35–41.
[7] Davidson, B. D., Hu, H., and Schapery, R. A., 1995, "An Analytical Crack-Tip Element for Layered Elastic Structures," J. Appl. Mech., **62**, pp. 294–305.
[8] Point, N., and Sacco, E., 1996, "Delamination of Beams: An Application to the DCB Specimen," Int. J. Fract., **79**, pp. 225–247.
[9] Bruno, D., and Greco F., 2001, "Mixed Mode Delamination in Plates: A Refined Approach," Int. J. Solids Struct., **38**, pp. 9149–9177.
[10] Wang, J., and Qiao P., 2004, "Interface Crack between Two Shear Deformable Elastic Layers," J. Mech. Phys. Solids, **52**, pp. 891–905.
[11] Vizzini, A. J., and Lagace, P. A., 1987, "The Buckling of a Delaminated Sublaminate on an Elastic Foundation," J. Compos. Mater., **21**, pp. 1106–1117.
[12] Suo, Z., and Hutchinson, J. W., 1990, "Interface Crack between Two Elastic Layers," Int. J. Fract., **43**, pp. 1–18.
[13] Schapery, R. A., and Davidson, B. D., 1990, "Prediction of Energy Release Rate for Mixed-mode Delamination Using Classical Plate Theory," Appl. Mech. Rev., **43**, pp. S281–S287.
[14] Davidson, B. D., and Sundararaman, V., 1996, "A Single Leg Bending Test for Interfacial Fracture Toughness Determination," Int. J. Fract., **78**, pp. 193–210.
[15] Fraisse, R., and Schmit, F., 1993, "Use of J-Integral as Fracture Parameters in Simplified Analysis of Bonded Joints," Int. J. Fract., **63**, pp. 59-73.
[16] Li, S., Wang, J., and Thouless, M. D., 2004. "The Effects of Shear on Delamination in Layered Materials," J. Mech. Phys. Solids, **52**, pp. 193–214.
[17] Wang, J., and Qiao P., 2004, "On the Energy Release Rate and Mode Mix of Delaminated Shear Deformable Composite Plates," Int. J. Solids Struct., **41**, pp. 2757–2779.
[18] Rice, J. R., 1988, "Elastic Fracture Mechanics Concepts for Interfacial Cracks," J. Appl. Mech., **55**, pp. 98–103.

M. Ahmer Wadee
Department of Civil and Environmental Engineering,
Imperial College of Science, Technology and Medicine,
South Kensington Campus,
London SW7 2AZ, UK

L. A. P. Simões da Silva
Departamento de Engenharia Civil,
Universidade de Coimbra Polo II,
Pinhal de Marrocos,
3030-290 Coimbra, Portugal

Asymmetric Secondary Buckling in Monosymmetric Sandwich Struts

An interactive buckling model for sandwich struts accounting for buckle pattern localization is extended to cover such struts with differing face plate thicknesses. Although this does not affect the critical buckling characteristics of the structure, there is a significant change in the postbuckling behavior; formerly symmetric secondary buckling and imperfection sensitivity characteristics lose this quality as both become asymmetric. [DOI: 10.1115/1.1979513]

1 Introduction

Sandwich construction is a mass efficient structural form used extensively in astronautic [1], aeronautic [2,3], and marine applications [4]. They are used as members for general loading situations: bending, shear, and axial loading. In axial compression, however, there are serious structural integrity issues precisely because of their inherent efficiency; their susceptibility to highly unstable interactive buckling phenomena in practical situations is widely accepted [5,6]. Previous work developing a nonlinear variational model, accounting for the severe interaction between overall (Euler-type) and local modes of buckling that leads to highly unstable localized buckling [7], has identified the problems with using orthotropic core materials [8], and with having preexisting defects in terms of lack of straightness in the face plates [9] and face-core delamination [10,11]. These effects do not significantly affect the linearly evaluated critical load capacity, which have been historically well understood [12], but they can have severe implications on the postcritical load-carrying capacity of the sandwich strut concerned. Moreover, the nonlinear effects can seriously question the practical value of linear eigenvalue analysis for these cases—the maximum capacity of the structural component being well below the linearly evaluated critical load and therefore significant factors of safety would need to be applied in relevant design rules.

A further enhancement to this sandwich structure model is presented in this paper where the stiff face plates now have the possibility of having differing thicknesses, hence, introducing a degree of monosymmetry in the cross section [13], which is a common way of modeling corrugated sandwich panels in industrial buildings. Comparing the critical buckling behavior of an Euler strut with a doubly symmetric cross section against a monosymmetric one, it is well known from classical work [14] that the respective levels of the linear buckling load may differ but their postbuckling behavior would still be symmetric, assuming of course that there is no possibility of local buckling within the cross-section itself. However, and here lies the key point, if local buckling is possible then the situation changes significantly; any local buckling on one side of the cross-sectional neutral axis will differ markedly from the other side due to the different configuration of the section on either side of the neutral axis. For the sandwich structure, if the face plates have different thicknesses then this introduces the possibility of the local mode of one of the face plates interacting with the overall mode of buckling in differing degrees depending on the initial sign of the critical mode. The result of this is that although the primary postbuckling response is still symmetric—given that the overall mode of buckling is the first instability—further deformation introduces a second instability that leads to localized buckling in the face plate with greater compression; the magnitude of overall buckling displacement required to trigger the second instability relating to the thickness of the face plate in more compression. The result of this is an asymmetric postbuckling response following the second instability with the consequent imperfection sensitivity also becoming asymmetric.

The present paper begins with the development of the interactive buckling model with the new feature of cross-section monosymmetry. The formulation is based on the variational principle of minimum potential energy, the contributions being from the bending and compression in the face plates, shearing and transverse compression in the core, and the work done by the external load. The equilibrium equations are then developed using the calculus of variations and a linear eigenvalue analysis yields the critical load for overall buckling. As the equilibrium equations form a system of nonlinear ordinary differential equations, subject to integral and boundary conditions, these are solved with a powerful numerical continuation code for a series of different struts with different levels monosymmetry with an objective to determine the severity of the postbuckling response. A particularly severe case is then investigated in detail such that results from both the perfect and imperfect struts, along with a consideration of the worst case imperfection, are presented. Conclusions are then drawn.

2 Interactive Buckling Model

Sandwich panels have stiff face plates placed on a lightweight and softer core material. Figure 1 shows the layout of the sandwich strut to be studied. Face plates are assumed to be thin, which in the present case means that through thickness shear is negligible and that the depth of the core b is large in comparison, isotropic and have a Young's modulus E and Poisson's ratio ν. The core can be constructed from balsa wood, two-dimensional cellular materials such as, aluminium honeycombs, or three-dimensional cellular materials such as polyurethane foams [15–17]. The constitutive law for these materials are commonly assumed to be either isotropic, orthotropic, or transversely isotropic. In the present paper, for the least complexity, the core will be assumed to be homogeneous and isotropic with Young's modulus E_c and Poisson's ratio ν_c, even though previous interactive localized buckling studies have investigated the effect of core orthotropy [8]. The principal difference from previous interactive buck-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division March 19, 2004; final revision, December 3, 2004. Associate Editor: E. M. Arruda. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

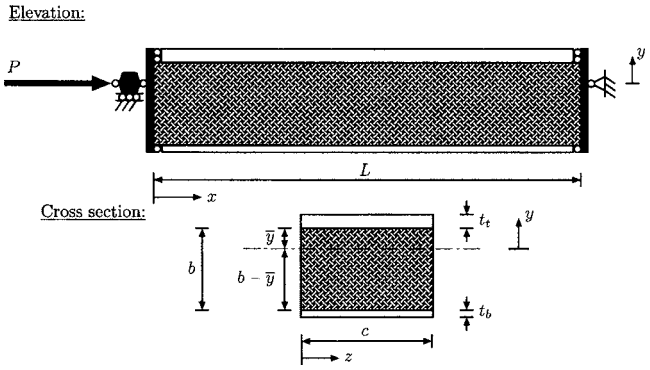


Fig. 1 The sandwich strut and its cross section. The face plates can have different thicknesses (t_b and t_t) and that the distance from the top face plate to the neutral axis is $\bar{y}$. Note that the load P is applied at the neutral axis of the strut.

ling formulations is that here the face plates have differing thicknesses introducing the opportunity for asymmetry of response particularly in the postbuckling range; the symmetry of the critical buckling response is basically unchanged but the relative location of the resulting secondary instability depends on the thickness configuration of the face plates and the initial orientation of the critical buckling displacement. It is also worth noting that owing to face plate anisotropy, monosymmetry can be introduced to the sandwich panel even if face plates have equal thicknesses [18].

The expected behavior of such a strut in compression is for it to follow the pure squash fundamental path, Fig. 2(*a*), and then to buckle in the overall mode—Fig. 2(*b*)—causing differential compression in the faces. This is likely to be followed closely by a secondary bifurcation, in which the face under the greater compression buckles in a second *localized* mode, Fig. 2(*c*). Any initial geometric imperfection in the strut (shown as an initial value of end-shortening $\mathcal{E}_i$) smears out the nonsmooth nature of the perfect equilibrium path shown in Fig. 2(*d*).

2.1 Cross-section Monosymmetry. Figure 1 shows the monosymmetric configuration of the sandwich strut with differing face plate thicknesses t_t and t_b representing the "top" face and "bottom" face, respectively. When undergoing overall buckling, the bottom face plate is deemed always to have the greater compression; the extra secondary buckling displacement is thus always confined to that face plate. With these definitions, the distance of the cross-section neutral axis from the top face-core interface $\bar{y}$ can be found in the conventional way for a section with different material elements and is a key quantity to the formulation of the model

$$\bar{y} = \frac{E_c b^2 + E t_b [2b + t_b(1-\tau^2)]}{2[E_c b + E t_b (1+\tau)]}, \tag{1}$$

where the ratio of the face plate thicknesses τ is defined

$$\tau = \frac{t_t}{t_b}, \tag{2}$$

which at times is referred to as the monosymmetry parameter. Of course if the load were to be applied elsewhere from the neutral axis then the structure would combine bending and compression, a more complicated modeling problem which is beyond the scope of the present study and is left for future work.

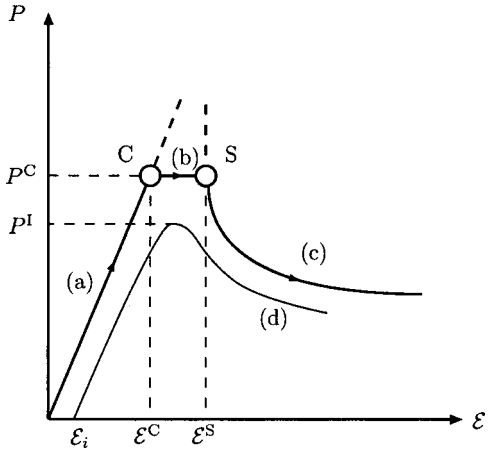


Fig. 2 Typical load P vs end-shortening $\mathcal{E}$ equilibrium diagram for sandwich struts: (a) fundamental path; (b) critical path of overall buckling; (c) secondary path of localized buckling; (d) typical imperfect structure equilibrium path

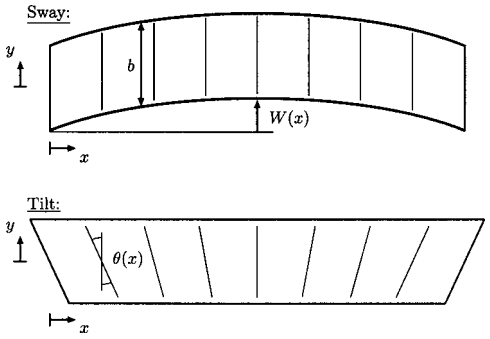


Fig. 3 Sway and tilt components of overall mode

2.2 Face Plate Displacements. An earlier theoretical model [7] had overall buckling represented by sway and tilt components of a long wave mode (Fig. 3) reflecting the possibility of shearing deformations in the core material where

$$W(x) = q_s L \sin\frac{\pi x}{L}, \tag{3}$$

$$\theta(x) = q_t \pi \cos\frac{\pi x}{L}, \tag{4}$$

and interactive localized buckling defined by two functions $w(x)$ and $u(x)$ representing, respectively, lateral and in-plane displacements of a single face plate (Fig. 4). There is also a pure squash strain component of both faces, Δ, which is introduced as a degree of freedom.

3 Variational Formulation

The principle of minimum potential energy is the basis for the following model's formulation. The total potential energy function V is defined as the sum of the internal strain energy stored in the structure (U) minus the work done by the external loads ($P\mathcal{E}$). The accumulated energy has a functional form which is then analyzed using the calculus of variations to determine the equilibrium equations.

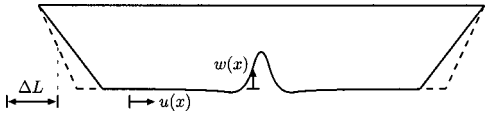


Fig. 4 Displacement functions used to model localized buckling

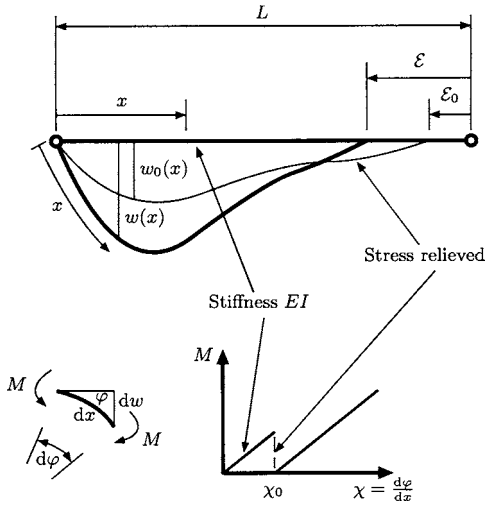


Fig. 5 Stress-relieved state of the strut, after Thompson and Hunt (see Ref. [20])

3.1 Strain Energy. In a general state of deflection there are three components of strain energy U: pure bending in the faces alone (U_b), membrane action in stretching or compressing the faces (U_m), and core energy (U_c) that comprises transverse and shearing strains.

3.1.1 Bending Energy. The bending energy components arise from the overall bending of both faces, coupled with the local bending of the lower face plate. Linear curvature expressions suffice giving the following expression for the contribution from the overall mode:

$$U_{bo} = \frac{1}{2}E\int_0^L \left[(I_t + I_b) q_s^2 \frac{\pi^4}{L^2} \sin^2 \frac{\pi x}{L} \right] dx, \tag{5}$$

where E is the Young's modulus of the face plate with an associated Poisson's ratio ν. The quantities of I_t and I_b are the local minor axis second moments of area of the top and bottom face plates, respectively,

$$I_t = \frac{ct_t^3}{12(1-\nu^2)} \tag{6}$$

$$I_b = \frac{ct_b^3}{12(1-\nu^2)} \tag{7}$$

The model assumes that the bottom face plate has the greater compression once overall buckling has occurred.

The sandwich panel is also modeled with a generalized imperfection that represents the initial deformation of the more compressed face plate w_0, which is formulated from an energy principle. The form of the imperfection closely matches that of the localized buckling mode for the strut on a softening foundation—derived from a first order approximation of a multiple scale perturbation analysis [19]

$$w_0(x) = A_0 \operatorname{sech}[\alpha(x - L/2)] \cos[\beta\pi(x - L/2)/L], \tag{8}$$

where A_0 is the amplitude of the imperfection, with α and β defining the shape of it; α governs the degree of localization and β governs the number of sinusoidal waves in the imperfection. The imperfection is introduced by supposing an initially deformed shape of the more compressed face plate $w_0(x)$ is stress relieved, such that the elemental bending moment, M, and thus stored strain energy of bending U_{bl} drop to zero as represented in Fig. 5 [20,9]:

$$dU_{bl} = \tfrac{1}{2}M d\varphi = \tfrac{1}{2}EI_b(\chi - \chi_0)^2 dx, \tag{9}$$

where χ is the curvature of the strut due to w, χ_0 is the curvature of the strut due to w_0 and EI_b is the flexural rigidity of the imperfect face plate. This also assumes that the strain energy in the foundation is nonzero in the initial state. Assuming a small deflection curvature relationship, i.e., $\chi = w''$, the strain energy of bending becomes

$$U_{bl} = \tfrac{1}{2}EI_b \int_0^L (w'' - w_0'')^2 dx, \tag{10}$$

where primes represent differentiation with respect to the spatial variable x. Therefore the total strain energy stored from bending U_b is the sum of U_{bo} and U_{bl}, thus

$$U_b = \frac{1}{2}E\int_0^L \left[(I_t + I_b) q_s^2 \frac{\pi^4}{L^2} \sin^2 \frac{\pi x}{L} + I_b(w'' - w_0'')^2 \right] dx. \tag{11}$$

3.1.2 Membrane Energy. The membrane strain energy U_m accounts for the axial tension and compression in the face plates when the structure bends after overall buckling. In the "tilt" configuration shown in Fig. 3, the upper face plate simply contributes the tensile equivalent of the compressive strain resulting from overall buckling along with the component of pure squash Δ:

$$\epsilon_{xt} = q_t \frac{\bar{y}\pi^2}{L} \sin \frac{\pi x}{L} - \Delta, \tag{12}$$

while the lower face plate adds as extra contributions the corresponding strains from Von Kármán large-deflection plate theory

$$\epsilon_{xb} = u' + \frac{1}{2}w'^2 - q_t(b - \bar{y}) \frac{\pi^2}{L} \sin \frac{\pi x}{L} - \Delta. \tag{13}$$

The membrane strain energy is therefore ($D = Et_b c/2$):

$$\begin{aligned} U_m &= \frac{Et_t c}{2}\int_0^L \epsilon_{xt}^2 dx + \frac{Et_b c}{2}\int_0^L \epsilon_{xb}^2 dx \\ &= D\int_0^L \left\{ q_t^2[\tau\bar{y}^2 + (b - \bar{y})^2] \frac{\pi^4}{L^2} \sin^2 \frac{\pi x}{L} \right. \\ &\quad + 2q_t[b - \bar{y}(1+\tau)] \frac{\pi^2}{L} \sin \frac{\pi x}{L} \Delta + (1+\tau)\Delta^2 \\ &\quad + u'^2 + \frac{1}{4}w'^4 + u'w'^2 - \left[2\Delta + 2q_t(b - \bar{y}) \frac{\pi^2}{L} \sin \frac{\pi x}{L} \right] \\ &\quad \left. \times \left(u' + \frac{1}{2}w'^2 \right) \right\} dx. \end{aligned} \tag{14}$$

3.1.3 Core Energy. In general, the core provides all the transverse and shear resistance but only some of the longitudinal resistance, most of which comes from the faces. Analysis is somewhat simplified if it is assumed that the faces provide all of the latter; a usual assumption if the core is assumed to be *soft* [21]. Plane stress expressions are readily derivable which include a longitudinal strain component in the core, but little is lost conceptually if this is ignored [8]. To match the assumed displacements $w(x)$ and $u(x)$ on the lower face and zero on the upper face of Fig. 4, displacements $w_c(x,y)$ and $u_c(x,y)$ must vary through the core. If x and y are defined as in Fig. 3 and linear variations with y are assumed the following expressions are obtained:

$$u_c(x,y) = \left(\frac{\bar{y} - y}{b} \right) u(x), \tag{15}$$

$$w_c(x,y) = \left(\frac{\bar{y} - y}{b} \right) w(x). \tag{16}$$

It is worth noting that the validity of this assumption has been confirmed in earlier work in a comparative study on a simplified strut on a nonlinear elastic foundation with a simulated overall

buckling mode [7]. Under the assumption that ϵ_x is zero in the core, the remaining core strains can be written

$$\epsilon_y = \left(\frac{\partial w_c}{\partial y}\right) = -\left(\frac{w}{b}\right), \tag{17}$$

$$\gamma_{xy} = \frac{\partial W}{\partial x} - \theta + \frac{\partial w_c}{\partial x} + \frac{\partial u_c}{\partial y} = (q_s - q_t)\pi \cos\frac{\pi x}{L} + \left(\frac{\bar{y}-y}{b}\right)w' - \frac{u}{b}, \tag{18}$$

where G_c is the core shear modulus and can be related to the core's Young's modulus E_c for an isotropic material

$$G_c = \frac{E_c}{2(1+\nu)}. \tag{19}$$

At this point it is worth re-emphasizing that although soft-core materials are usually modeled constitutively as orthotropic or transversely isotropic, the present study takes the simplest case of isotropy such that the effect of differing face plate thicknesses is highlighted more than the constitutive law of the core material; comparisons of core orthotropy versus isotropy have been studied at length in earlier work [8]. Therefore, the general expression for the strain energy stored in the core is

$$U_c = \frac{E_c c}{2(1-\nu_c^2)}\int_0^L \int_{\bar{y}-b}^{\bar{y}} \epsilon_y^2 dy dx + \frac{G_c c}{2}\int_0^L \int_{\bar{y}-b}^{\bar{y}} \gamma_{xy}^2 dy dx, \tag{20}$$

and substituting the appropriate expressions and integrating with respect to y, we obtain ($G=G_c cb/2$):

$$U_c = \int_0^L \left\{ \left(\frac{1}{2}kw^2 - \frac{1}{3}k_1 w^3 + \frac{1}{4}k_2 w^4\right) + G\left[(q_s - q_t)^2 \pi^2 \cos^2\frac{\pi x}{L} + (q_s - q_t)\pi \cos\frac{\pi x}{L}\left(w' - \frac{2u}{b}\right) + \left(\frac{1}{3}w'^2 + \frac{u^2}{b^2} - \frac{1}{b}uw'\right)\right]\right\} dx, \tag{21}$$

where k is introduced as the linear component of the transverse core stiffness

$$k = \frac{E_c c}{(1-\nu_c^2)b}. \tag{22}$$

Note the introduction of nonlinearities in the core's transverse stiffness (k_1 and k_2) that account for the general cellular core material response under direct compression. The force versus displacement profile of a cellular core material in direct compression has a distinctive shape in that it is initially linear followed by a plateau where the stiffness can actually drop below zero ($k_1>0$) as the microscopic cell walls in the material matrix buckle elastically. Subsequently the material restiffens ($k_2>0$) as the cells densify and the material becomes more resistant to further cellular compression [17,22]. For the purposes of the present study, which is to examine the location of the secondary instability and the initial postbuckling response, the core is assumed to behave in a linearly elastic fashion ($k_1=k_2=0$).

3.2 Work Done By Load. The final component of energy to be identified is the work done by the load. The overall tilt mode contributes nothing to the corresponding deflection $\mathcal{E}$ of a load P applied along the neutral axis of the cross section (Fig. 1), while the contribution from the local buckling of a single face scales the total of the end shortening of the face itself by the ratio of the distance to the neutral axis from the top face plate $\bar{y}$ to the depth of the core b. Together with the contributions from pure squash and sway from overall buckling (Fig. 3) this gives the work done as

$$P\mathcal{E} = P\int_0^L \left[\frac{q_s^2\pi^2}{2}\cos^2\frac{\pi x}{L} - \left(\frac{\bar{y}}{b}\right)u' + \Delta\right] dx. \tag{23}$$

3.3 Potential Energy Functional. The total potential energy V of the complete structure is given by the summation of all the strain energy terms: U_c, U_m, and U_b, minus the work done by the load, $P\mathcal{E}$ [23]. This hybrid form of the potential energy involves three degrees of freedom: q_s, q_t, and Δ; and a functional involving the two functions $w(x)$ and $u(x)$:

$$V = \int_0^L \left\{ \frac{1}{2}E\left[(I_t + I_b)q_s^2\frac{\pi^4}{L^2}\sin^2\frac{\pi x}{L} + I_b(w'' - w_0'')^2\right] + D\left\{ q_t^2[\tau\bar{y}^2 + (b-\bar{y})^2]\frac{\pi^4}{L^2}\sin^2\frac{\pi x}{L} + 2q_t[b - \bar{y}(1+\tau)]\frac{\pi^2}{L}\sin\frac{\pi x}{L}\Delta + u'^2 + (1+\tau)\Delta^2 + \frac{1}{4}w'^4 + u'w'^2 - \left[2\Delta + 2q_t(b-\bar{y})\frac{\pi^2}{L}\sin\frac{\pi x}{L}\right]\left(u' + \frac{1}{2}w'^2\right)\right\} + \frac{1}{2}kw^2 + G\left[(q_s - q_t)^2\pi^2\cos^2\frac{\pi x}{L} + \left(\frac{1}{3}w'^2 + \frac{u^2}{b^2} - \frac{1}{b}uw'\right) + (q_s - q_t)\pi\cos\frac{\pi x}{L}\left(w' - \frac{2u}{b}\right)\right] - P\left[\frac{q_s^2\pi^2}{2}\cos^2\frac{\pi x}{L} - \left(\frac{\bar{y}}{b}\right)u' + \Delta\right]\right\} dx \tag{24}$$

The strut width c is a common factor which can be completely eliminated from the model. It is therefore taken as unity for the purposes of numerical solution.

3.4 Linear Eigenvalue Analysis for Overall Buckling. Linear eigenvalue analysis yields the critical load for overall buckling P^C, arising on the pure squash fundamental path F at which $q_s = q_t = w = u = 0$. The potential energy is nondiagonal with respect to q_s and q_t, and so this critical load occurs when the matrix

$$\mathbf{V}_{ij}^F = \begin{pmatrix} V_{ss}^F & V_{st}^F \\ V_{ts}^F & V_{tt}^F \end{pmatrix} \tag{25}$$

is singular, where

$$V_{ij}^F = \left.\frac{\partial^2 V}{\partial q_i \partial q_j}\right|^F. \tag{26}$$

Therefore the critical load for overall buckling is obtained

$$P^C = 2G + \frac{\pi^2 EI_b}{L^2}(1+\tau^3) - \left[\frac{2G^2}{\frac{D\pi^2}{\tau L^2}[\tau\bar{y}^2 + (b-\bar{y})^2] + G}\right], \tag{27}$$

and this expression holds for both positive and negative values of q_s.

3.5 Equilibrium Equations. The integral in Eq. (24) representing the total potential energy of the physical system which must be stationary at equilibrium; the calculus of variations is applied to find this condition. The analysis that follows is a summary of this application of the calculus of variations. Consider the Lagrangian ($\mathcal{L}$) of the form

$$V = \int_0^L \mathcal{L}(w'', w', w, u', u, x) dx. \tag{28}$$

For the system to satisfy equilibrium V has to be stationary. This requires the first variation of V, which is given by the following expression [24]:

$$\delta V = \int_0^L \left[\frac{\partial \mathcal{L}}{\partial w''}\delta w'' + \frac{\partial \mathcal{L}}{\partial w'}\delta w' + \frac{\partial \mathcal{L}}{\partial w}\delta w + \frac{\partial \mathcal{L}}{\partial u'}\delta u' + \frac{\partial \mathcal{L}}{\partial u}\delta u \right] dx. \tag{29}$$

By satisfying the condition $\delta V=0$, the Euler–Lagrange equations give a system of nonlinear ordinary differential equations

$$EI_b w'''' + D\left[2\Delta w'' + 2q_t(b-\bar{y})\frac{\pi^2}{L}\left(\sin\frac{\pi x}{L}w'' + \frac{\pi}{L}\cos\frac{\pi x}{L}w' \right) - 3w'^2 w'' - 2(u''w' + u'w'') \right] + G\left[\frac{u'}{b} - \frac{2}{3}w'' + (q_s - q_t)\frac{\pi^2}{L}\sin\frac{\pi x}{L} \right] + kw = EI_b w_0'''', \tag{30}$$

$$D\left[u'' + w'w'' - q_t(b-\bar{y})\frac{\pi^3}{L^2}\cos\frac{\pi x}{L} \right] + \frac{G}{b}\left[\frac{1}{2}w' - \frac{u}{b} + (q_s - q_t)\pi\cos\frac{\pi x}{L} \right] = 0. \tag{31}$$

Other equilibrium equations are found by differentiating V with respect to the degrees of freedom Δ, q_s and q_t, respectively,

$$2q_t\pi[b-\bar{y}(1+\tau)] + (1+\tau)\Delta L - \frac{PL}{2D} = \int_0^L \left(u' + \frac{1}{2}w'^2 \right) dx, \tag{32}$$

$$P = E(I_t + I_b)\frac{\pi^2}{L^2} + 2G\left(\frac{q_s - q_t}{q_s} \right) + \frac{2G}{q_s \pi L}\int_0^L \cos\frac{\pi x}{L}\left(w' - \frac{2u}{b} \right) dx, \tag{33}$$

$$\frac{D}{G}\left\{ q_t[\tau\bar{y}^2 + (b-\bar{y})^2]\frac{\pi^2}{L^2} + \frac{4}{\pi L}[b-\bar{y}(1+\tau)]\Delta \right\} - (q_s - q_t) = \int_0^L \left[\frac{2D(b-\bar{y})}{GL^2}\sin\frac{\pi x}{L}\left(u' + \frac{1}{2}w'^2 \right) + \frac{1}{\pi L}\cos\frac{\pi x}{L}\left(w' - \frac{2u}{b} \right) \right] dx. \tag{34}$$

The system of equilibrium equations are also subject to three boundary conditions at each end, these are determined from integrating V by parts and yield the following [7]:

$$w(0) = w''(0) = 0, \tag{35}$$

$$w(L) = w''(L) = 0, \tag{36}$$

$$u'(0) + \frac{1}{2}w'^2(0) - \Delta = \frac{P\bar{y}}{Db}, \tag{37}$$

$$u'(L) + \frac{1}{2}w'^2(L) - \Delta = \frac{P\bar{y}}{Db}. \tag{38}$$

However, owing to the symmetry of the structure lengthwise it is easier to solve the equations numerically between $x=0$ and midspan $x=L/2$. The following symmetry condition is therefore imposed at midspan in the numerical code:

$$w'(L/2) = w'''(L/2) = u(L/2) = 0, \tag{39}$$

and because of this the conditions at $x=L$, i.e., conditions given in (36) and (38) are automatically satisfied.

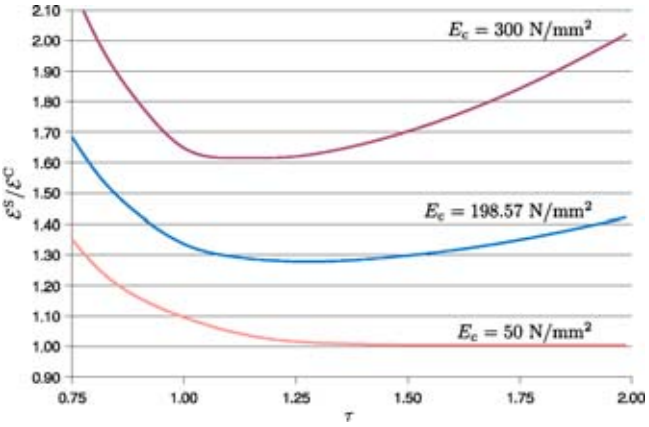


Fig. 6 Relative proximity of secondary and critical bifurcations for struts with different cores and monosymmetries

4 Numerical Experiments

The full system of equilibrium Eqs. (30)–(34) are discretized and solved subject to the conditions (35), (37), and (39) using the numerical continuation package AUTO97 [25]. This allows the evaluation of the physical postbuckling modes and the equilibrium response of a variety of different strut configurations that are selected initially from earlier studies.

In the present study, the initial sign of the overall mode amplitude q_s is of paramount importance; the overall mode is triggered at the same load for both positive and negative values of q_s putting the thicker and thinner face plates into extra compression in turn. Therefore, with a face plate configuration with differing thicknesses, a source of asymmetry in the postcritical buckling response is introduced; the secondary bifurcation that triggers localized buckling occurs at different magnitudes of q_s depending on its initial sign. This leads to the conclusion that the degree of imperfection sensitivity of the strut is asymmetric.

The cases are presented such that a symmetric struts with different core properties are selected from which their relative face plate thicknesses are varied. The struts are compared against each other by measuring the proximity of the secondary instability leading to localized buckling relative to the initial instability that causes overall (Euler-type) buckling. The study then focuses in on a particular strut with a highly pronounced overall-local mode interaction and presents its postbuckling characteristics together with a study of its behavioral sensitivity to initial geometric imperfections.

4.1 Proximity of Secondary Bifurcation. The basic strut used in the numerical study has the following material properties taken from the literature [26,5,7]. The strut length and core depth are kept constant in the present study: $L=508$ mm and $b=50.8$ mm, respectively. The face plate material properties are also kept constant: Young's modulus $E=68{,}947.57$ N/mm^2 and Poisson's ratio $\nu=0.3$. The face plate thicknesses and the Young's modulus of the core are varied in the numerical study.

Figure 6 shows the effect of changing the monosymmetry parameter τ for the strut on the relative gap between the critical bifurcation for overall buckling and the secondary bifurcation for localized buckling ($\mathcal{E}^S/\mathcal{E}^C$). Three different cases of core modulus are shown and it can be seen that the degree of monosymmetry can bring together the separate bifurcations, which in turn results in a less stable structure in the postbuckling range. For the cases of $E_c=300$ and 198.57 N/mm^2, there is clearly a worst case monosymmetry with a minimum value $\mathcal{E}^S/\mathcal{E}^C$ being achieved for a certain value of τ, as shown in Table 1. For the third case, $E_c=50$ N/mm^2, the bifurcations become all but simultaneous for a relatively large range of τ, this is most likely due to the local mode becoming nearly critical for those values.

Table 1 Worst case monosymmetry parameter τ for different core moduli E_c. Simultaneous (compound) bifurcation points would be represented by $\mathcal{E}^S/\mathcal{E}^C=1$.

E_c (N/mm^2)	τ	$\mathcal{E}^S/\mathcal{E}^C$
300	1.15	1.610
198	1.27	1.278
50	1.40–2.00	1.005

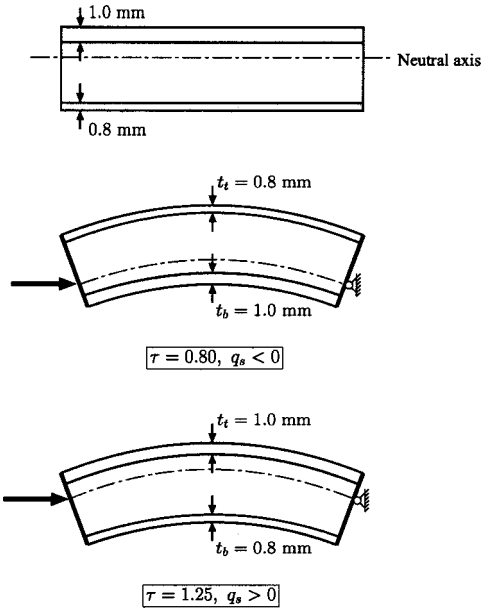


Fig. 7 The selected strut: definitions of the thicknesses t_t, t_b, monosymmetry parameter τ, and the sign of overall buckling q_s

4.2 Full Postbuckling Results. Leading on from the previous section, postbuckling results are now presented for a strut that has quite differing relative secondary bifurcation properties depending on the sign of the overall buckling. Figures 7–9 show a schematic representation of the selected strut, buckling displacements, and equilibrium diagrams for the geometrically perfect strut, respectively. The selected strut has the following dimensions and properties: face thicknesses 1.0 and 0.8 mm; length $L=508$ mm; core depth and properties $b=50.8$ mm, $E_c=50$ N/mm^2, and $\nu_c=0.2$. As shown in Fig. 7, if the thicker face plate is in extra compression after overall buckling is triggered then q_s is negative and if the thinner face plate is in extra compression after overall buckling then q_s is positive. For $q_s<0$ the value of the monosymmetry parameter $\tau=0.8$ with the bifurcation proximity: $\mathcal{E}^S/\mathcal{E}^C=1.267$, and for $q_s>0$ the value of the monosymmetry parameter $\tau=1.25$ with the bifurcation proximity: $\mathcal{E}^S/\mathcal{E}^C=1.016$.

It can be seen in the graphs that the secondary (localized) postbuckling modes of w, shown in Figs. 8(a) and 8(c), have differing localized wavelengths for the face plates. Moreover, the important result from this is shown in Fig. 9 where the proximity between the critical and secondary bifurcations is significant only for $q_s<0$; for $q_s>0$ the gap is much smaller. This implies that if the strut has imperfections forcing $q_s>0$, the strut is much less likely to reach the linear eigenvalue critical load than if $q_s<0$ were forced. This asymmetry in the potential sensitivity to imperfections is now quantified in the following section.

4.2.1 Imperfection Sensitivity. The geometric imperfection w_0, shown in Eq. (8), is introduced for the strut that was selected in the previous section. For each face the imperfection sensitivity for the periodic mode was determined using the following technique: keeping the localization parameter α at zero and the wave number parameter β at the linear eigenvalue solution, the value of the imperfection magnitude $\mathcal{E}_0$, where

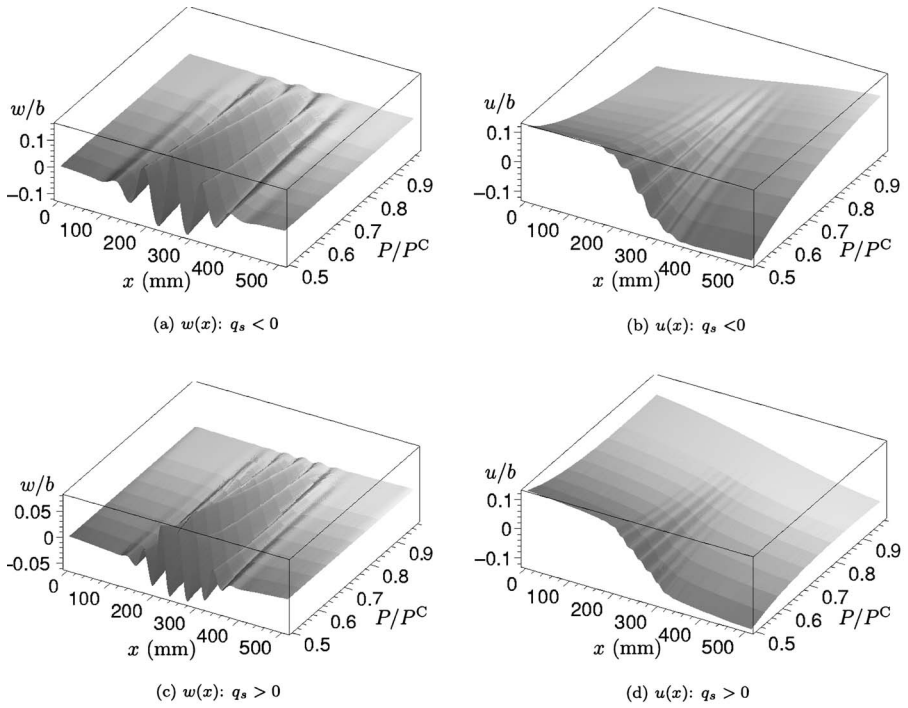
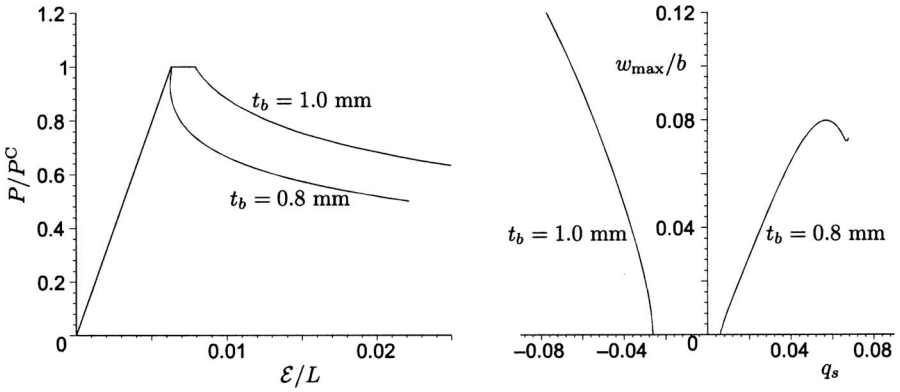


Fig. 8 Postbuckling profiles of monosymmetric strut with face thicknesses 1.0 and 0.8 mm. Other dimensions and properties: $L=508$ mm, $b=50.8$ mm, $E_c=50$ N/mm^2, and $\nu_c=0.2$.



(a) Load vs end-shortening

(b) Localized buckle amplitude vs overall buckle amplitude

Fig. 9 Postbuckling equilibrium diagrams of monosymmetric strut with face thicknesses 1.0 and 0.8 mm. Other dimensions and properties: L=508 mm, b=50.8 mm, E_c =50 N/mm^2, and ν_c=0.2.

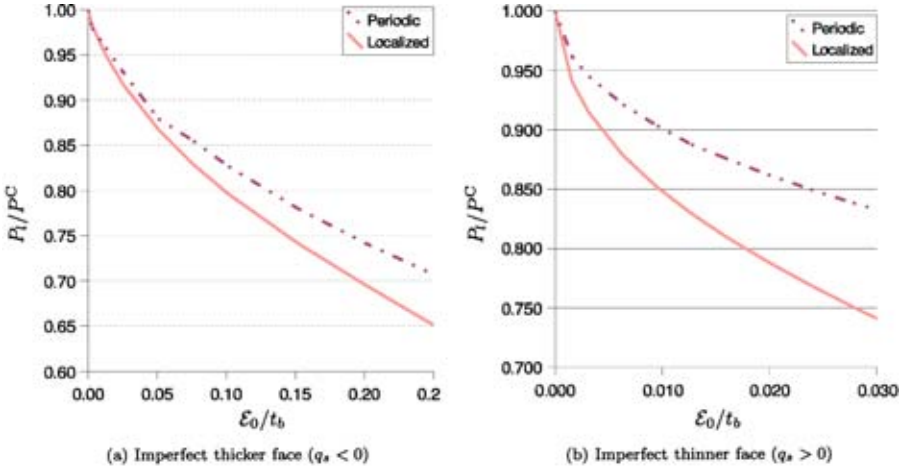
$$\mathcal{E}_0 = \int_0^L \tfrac{1}{2} w_0'^2 dx, \tag{40}$$

was varied; each variation of $\mathcal{E}_0$ gave an explicit value of A_0. The imperfection magnitude $\mathcal{E}_0$ is in fact a measure of the axial end-shortening introduced by the imperfection in the face plate prior to the commencement of external loading. The load P was then varied as a parameter in AUTO to find the corresponding limit load P_l/P^C (see Fig. 2). The locus of these values of the limit loads plotted against the size of the initial imperfection defines the so-called imperfection sensitivity curve.

For the localized imperfection at a particular value of $\mathcal{E}_0$, β was still kept at the linear eigenvalue solution but now both α and A_0 were varied. Again, the load P was then varied as a parameter to find the corresponding limit load P_l/P^C; the worst case was defined as the combination of A_0 and α that minimized the limit load for a given value of $\mathcal{E}_0$. Figures 10–12 show the imperfection sensitivity curves and the changing profile of the imperfection respectively for the cases where the sole imperfection is confined to either the thicker or thinner face plate. As expected from previous work [9], the localized imperfection gives the more severely unstable equilibrium response for nontrivial values of $\mathcal{E}_0$. Moreover, the monosymmetry of the sandwich panel means that the relative imperfection sensitivity is higher when the thinner face is imperfect rather than the thicker face; for example, when $\mathcal{E}_0/t_b$ =0.025, the limit load for localized imperfections for $q_s<0$ is P_l/P^C=0.917, the corresponding limit load level for $q_s>0$ being P_l/P^C=0.765, a drop in ultimate strength by 17% for an imperfection amplitude A_0 very much less than a third of the face plate thickness.

5 Conclusions

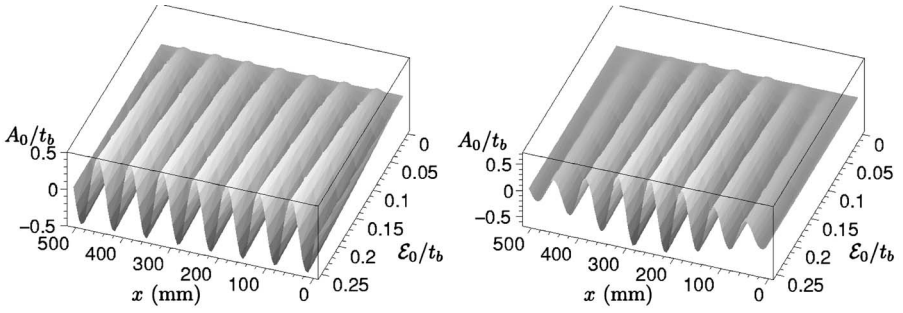
The present paper has formulated an enhanced variational interactive buckling model for sandwich struts that includes monosymmetry in the cross section, a component type that is relatively common in practical engineering situations. The principal finding is that this monosymmetry of construction introduces asymmetry in the nonlinear response, i.e., different localized postbuckling responses are found depending on the initial orientation of the critical buckling. An important issue here is that the relative proximity between the critical and secondary bifurcation can be significantly changed just by reversing the orientation of the critical buckling mode. The parametric study showed regimes where the degree of monosymmetry forced the critical and secondary bifurcations to be effectively coincident. Moreover, the detailed numerical study that focused in on a particular strut configuration showed a marked difference between the sensitivities to initial imperfections for the individual face plates; the thinner face being much more severely unstable after triggering the interactive (localized) mode than the thicker face.



(a) Imperfect thicker face ($q_s < 0$)

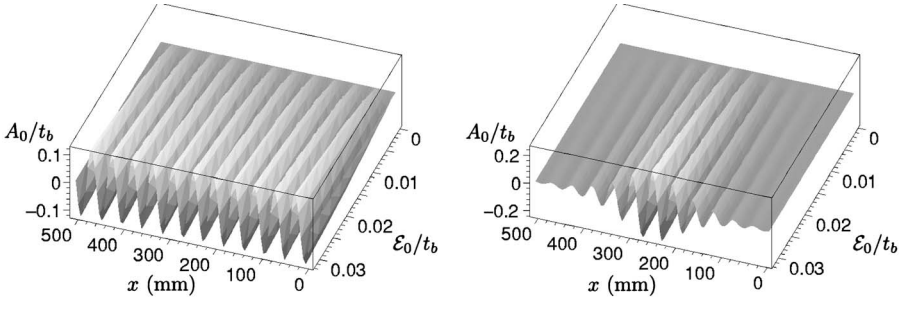
(b) Imperfect thinner face ($q_s > 0$)

Fig. 10 Imperfection sensitivity curves for periodic and localized geometric imperfections for the monosymmetric strut with thicknesses 1.0 and 0.8 mm; in this case the thinner face is much more sensitive



(a) Periodic imperfection (b) Localized imperfection

Fig. 11 Initial imperfection profiles for periodic and worst case localized geometric imperfections for the monosymmetric strut with t_t=0.8 mm and t_b=1.0 mm, i.e., the thicker face buckles (q_s<0)



(a) Periodic imperfection (b) Localized imperfection

Fig. 12 Initial imperfection profiles for periodic and worst case localized geometric imperfections for the monosymmetric strut with t_t=1.0 mm and t_b=0.8 mm, i.e., the thinner face buckles locally (q_s>0)

These findings are significant for design practice, as although the basic result of linear eigenvalue analysis is unchanged, the nonlinear analysis reveals a significant sensitivity to very small defects in the geometry such that the structure in certain configurations can only attain a fraction at its critical buckling load.

References

[1] Kodiyalam, S., Nagendra, S., and DeStefano, J., 1996, "Composite Sandwich Structure Optimization with Application to Satellite Components," AIAA J., **34**(3), pp. 614–621.

[2] Bannink, E., Hadcock, R., and Forsch, H., 1978, "Advanced Design Composite Material Aircraft Study," J. Aircr., **15**(10), pp. 661–669.

[3] Duthie, A. C., 1987, "Composite Materials Technology for Helicopters," Plastics and Rubber Int., **12**(1), pp. 20–24.

[4] Knox, E. M., Cowling, M. J., and Winkle, I. E., 1998, "Adhesively Bonded Steel Corrugated Core Sandwich Construction for Marine Applications," Mar. Struct., **11**(4–5), pp. 185–204.

[5] Hunt, G. W., Da Silva, L. S., and Manzocchi, G. M. E., 1988, "Interactive Buckling in Sandwich Structures," Proc. R. Soc. London, Ser. A, **417**, pp. 155–177.

[6] Sokolinsky, V., and Frostig, Y., 1999, "Nonlinear Behavior of Sandwich Panels with a Transversely Flexible Core," AIAA J., **37**, pp. 1474–1482.

[7] Hunt, G. W., and Wadee, M. A., 1998, "Localization and Mode Interaction in Sandwich Structures," Proc. R. Soc. London, Ser. A, **454**, pp. 1197–1216.

[8] Wadee, M. A., and Hunt, G. W., 1998, "Interactively Induced Localized Buckling in Sandwich Structures with Core Orthotropy," ASME J. Appl. Mech., **65**(2), pp. 523–528.

[9] Wadee, M. A., 2000, "Effects of Periodic and Localized Imperfections on Struts on Nonlinear Foundations and Compression Sandwich Panels," Int. J. Solids Struct., **37**(8), pp. 1191–1209.

[10] Wadee, M. A., and Blackmore, A., 2001, "Delamination from Localized Instabilities in Compression Sandwich Panels," J. Mech. Phys. Solids, **49**(6), pp. 1281–1299.

[11] Wadee, M. A., 2002, "Localized Buckling in Sandwich Struts with Pre-Existing Delaminations and Geometrical Imperfections," J. Mech. Phys. Solids, **50**(8), pp. 1767–1787.

[12] Allen, H. G., 1969, *Analysis and Design of Structural Sandwich Panels*, Pergamon, Oxford.

[13] Simões da Silva, L., 1991, "Asymmetric Post-Buckling Behaviour of Two and Three-Layered Sandwich Structures," *Proceedings of the Second Pan American Congress of Applied Mechanics*, Valparaiso, Chile, pp. 494–497.

[14] Timoshenko, S. P., and Gere, J. M., 1961, *Theory of Elastic Stability*, McGraw-Hill, New York.

[15] Easterling, E. J., Harrysson, R., Gibson, L. J., and Ashby, M. F., 1982, "On the Mechanics of Balsa and Other Woods," Proc. R. Soc. London, Ser. A, **383**(1784), pp. 31–41.

[16] Gibson, L. J., Ashby, M. F., Schjaer, G. S., and Robertson, C. I., 1982, "The Mechanics of Two-Dimensional Cellular Materials," Proc. R. Soc. London, Ser. A, **382**, pp. 25–42.

[17] Gibson, L. J., and Ashby, M. F., 1982, "The Mechanics of Three-Dimensional Cellular Materials," Proc. R. Soc. London, Ser. A, **382**, pp. 43–59.

[18] Hohe, J., and Librescu, L., 2004, "Core and Face-Sheet Anisotropy in Deformation and Buckling of Sandwich Panels," AIAA J., **42**(1), pp. 149–158.

[19] Wadee, M. K., Hunt, G. W., and Whiting, A. I. M., 1997, "Asymptotic and Rayleigh-Ritz Routes to Localized Buckling Solutions in an Elastic Instability Problem," Proc. R. Soc. London, Ser. A, **453**(1965), pp. 2085–2107.

[20] Thompson, J. M. T., and Hunt, G. W., 1984, *Elastic Instability Phenomena*, Wiley, Chichester.

[21] Librescu, L., Hause, T., and Camarda, C. J., 1997, "Geometrically Nonlinear Theory of Initially Imperfect Sandwich Curved Panels Incorporating Nonclassical Effects," AIAA J., **35**(8), pp. 1393–1403.

[22] Gibson, L. J., and Ashby, M. F., 1999, *Cellular Solids: Structure and Properties* (Cambridge Solid State Science Series), 2nd ed., Cambridge University Press, Cambridge.

[23] Thompson, J. M. T., and Hunt, G. W., 1973, *A General Theory of Elastic Stability*, Wiley, London.

[24] Fox, C., 1987, *An Introduction to the Calculus of Variations*, Dover, New York.

[25] Doedel, E. J., Champneys, A. R., Fairgrieve, T. F., Kuznetsov, Y. A., Sandstede, B., and Wang, X.-J., 1997, "AUTO97: Continuation and bifurcation software for ordinary differential equations," Technical report, Department of Computer Science, Concordia University, Montreal, Canada; available by FTP from ftp.cs.concordia.cain/pub/doedel/auto.

[26] Brush, D. O., and Almroth, B. O., 1975, *Buckling of Bars, Plates and Shells*, McGraw-Hill, New York.

Harmonic Shapes in Finite Elasticity Under Nonuniform Loading

G. F. Wang

P. Schiavone[1]
e-mail: p.schiavone@ualberta.ca

C.-Q. Ru

Department of Mechanical Engineering,
University of Alberta,
4-9 Mechanical Engineering Building,
Edmonton, Alberta T6G 2G8, Canada

We investigate the classic (inverse) problem concerned with the design of so-called harmonic shapes for an elastic material undergoing finite plane deformations. In particular, we show how to identify such shapes for a particular class of compressible hyperelastic materials of harmonic type. The "harmonic condition," in which the sum of the normal stresses in the original stress field remains unchanged everywhere after the introduction of the harmonic hole or inclusion, is imposed on the final stress field. Using complex variable techniques, we identify particular harmonic shapes arising when the material is subjected nonuniform (remote) loading and discuss conditions for the existence of such shapes. [DOI: 10.1115/1.1979514]

1 Introduction

The design of optimum structural shapes leads inevitably to an inverse problem in which we seek to determine the geometry of a boundary from prescribed conditions that must be satisfied by the final stress field. The approach used to solve such problems depends significantly on the pre-selected design criteria. Among these proposed criteria (see, for example, [1–3]), the "harmonic condition" advanced by Richards and Bjorkman [4] for linear elasticity is extremely attractive, in that it lends itself to the possibility of extension to more interesting and practical cases, for example, to cases involving *finite* elastic deformations. In the case of plane elastic deformations, the "harmonic condition" itself essentially requires that the sum of normal stresses remains unperturbed everywhere in the stress field following the introduction of a so-called "harmonic hole or inclusion" (harmonic shape).

In [5] the harmonic condition of Richards and Bjorkman [4] was extended to the case of *finite* plane elastic deformations and shapes of corresponding harmonic holes and inclusions were identified in a particular class of compressible hyperelastic materials of harmonic type. In fact, under uniform biaxial loading, it was shown that, as in the case of the linearly elastic material ([3,4]), the harmonic shape is elliptical, but, in the case of finite deformations, the material nonlinearity significantly affects both the geometry of the elliptical shape and the condition for its existence.

The harmonic shapes identified in [5], however, correspond only to the simple case of a *uniform* (remote) stress field imposed in the material. For a linearly elastic material subjected to a *nonuniform* (remote) stress field, Bjorkman and Richards [6] have identified harmonic holes significantly different from those obtained under the assumption of uniform loading, namely harmonic holes in the shape of deloids and cardeloids.

In the present paper, we continue the analysis begun in [5] and identify harmonic shapes (holes and inclusions) in a particular class of compressible hyperelastic materials of harmonic type under the more general assumption of *nonuniform* (remote) loading.

For completeness, in Sects. 1–3, we include a brief review of the model describing deformations of the harmonic material and the corresponding equations used to determine harmonic shapes. The main results are presented in Sec. 4.

2 Basic Equations For a Harmonic Material

The model of a compressible hyperelastic material of harmonic type was first advanced by John [7] and subsequently studied by various authors (see, for example, [8–11]). In these papers, it was found that theoretical results based on this model were in good agreement with experimental data. In this section, we present only a brief review of the equations governing finite plane (strain) deformations of a harmonic material. Further details can be found in [8–11].

Let the complex variable $z=x_1+ix_2$ represent the initial coordinates of a material particle in the undeformed configuration, and $w(z)=y_1(z)+iy_2(z)$ the corresponding spatial coordinates in the deformed configuration. Define the deformation gradient tensor as

$$F_{ij}=\frac{\partial y_i}{\partial x_j}. \tag{1}$$

For a particular class of harmonic materials, the strain energy density defined with respect to the undeformed unit area can be expressed by [9–11]:

$$W=2\mu[F(I)-J],\quad F'(I)=\frac{1}{4\alpha}\left[I+\sqrt{I^2-16\alpha\beta}\right], \tag{2}$$

where I and J are the scalar invariants of FF^T given by

$$I=\sqrt{F_{ij}F_{ij}+2J},\quad J=\det[F], \tag{3}$$

μ is the shear modulus and $1/2\leqslant\alpha<1$, $\beta>0$ are two material constants. The importance of this class of materials has been shown in [9,10], where experimental data show good agreement with the theoretical model corresponding to (2).

According to the formulation given by Ru [11], the deformation $w(z)$ can be written in terms of two analytic functions φ and ψ as

$$iw(z)=\alpha\varphi(z)+i\overline{\psi(z)}+\frac{\beta z}{\varphi'(z)}, \tag{4}$$

and the complex Piola stress function $\chi(z)$ is given by

$$\chi(z)=2\mu i\left[(\alpha-1)\varphi(z)+i\overline{\psi(z)}+\frac{\beta z}{\varphi'(z)}\right]. \tag{5}$$

The stress components can be written in terms of the Piola stress function as

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, July 9, 2004; final revision, December 3, 2004. Associate Editor: E. M. Arruda. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

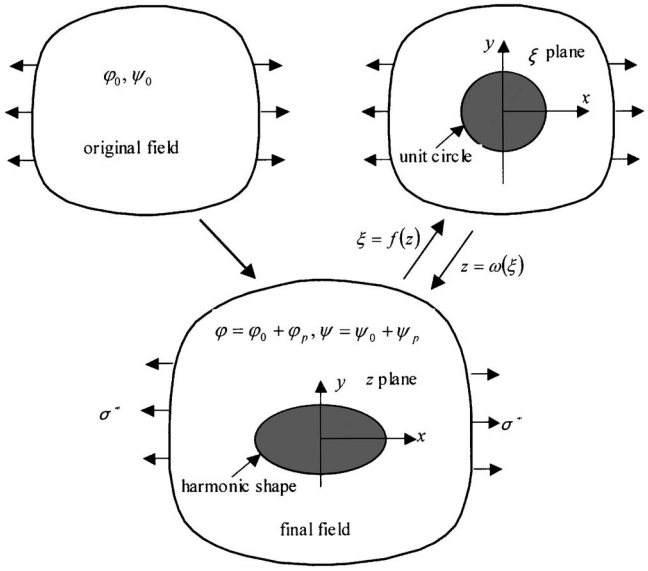


Fig. 1 Schematic of the strategy used to identify a harmonic shape

$$\sigma_{11} + i\sigma_{21} = -i\chi_{,2} \quad \text{and} \quad -i\sigma_{12} + \sigma_{22} = \chi_{,1}. \tag{6}$$

Using Eqs. (5) and (6), the sum of the two normal stresses is given by

$$\sigma_{11} + \sigma_{22} = \frac{\partial \chi}{\partial z} + \overline{\left(\frac{\partial \chi}{\partial z}\right)} = 2\mu i \left\{ (\alpha - 1)\left[\varphi'(z) - \overline{\varphi'(z)}\right] + \beta\left[\frac{1}{\overline{\varphi'(z)}} - \frac{1}{\varphi'(z)}\right]\right\}. \tag{7}$$

The displacements and stresses can be completely determined from the boundary conditions. That is, from the force boundary condition

$$2\mu i\left[(\alpha - 1)\varphi(z) + i\overline{\psi(z)} + \frac{\beta z}{\varphi'(z)}\right] = H(z), \tag{8}$$

and the deformation boundary condition

$$\alpha\varphi(z) + i\overline{\psi(z)} + \frac{\beta z}{\varphi'(z)} = iG(z). \tag{9}$$

Here, H and G are prescribed functions.

Our objective in the present paper is to solve the inverse problem concerned with the determination of the shape of a harmonic inclusion or hole (in the undeformed configuration) in this particular class of nonlinearly elastic material when the remote loading is *nonuniform*.

We proceed by first using the mapping function $z=\omega(\xi)$ to map (conformally) the infinite region outside an "arbitrary" harmonic shape in the z plane to the infinite region outside the unit circle in the ξ plane (see Fig. 1). The corresponding harmonic shape can then be identified if the mapping function $z=\omega(\xi)$ can be determined. We begin by transforming the necessary relations in the z plane to the ξ plane. For convenience, we write,

$$\varphi(z) = \varphi(\omega(\xi)) = \varphi(\xi), \quad \psi(z) = \psi(\omega(\xi)) = \psi(\xi). \tag{10}$$

The sum of the two normal stresses is then expressed in the ξ plane as

$$\sigma_{11} + \sigma_{22} = 2\mu i\left\{(\alpha - 1)\left[\frac{\varphi'(\xi)}{\omega'(\xi)} - \frac{\overline{\varphi'(\xi)}}{\overline{\omega'(\xi)}}\right] + \beta\left[\frac{\overline{\omega'(\xi)}}{\overline{\varphi'(\xi)}} - \frac{\omega'(\xi)}{\varphi'(\xi)}\right]\right\}. \tag{11}$$

Similarly, the boundary conditions are given in the ξ plane as

$$2\mu i\left[(\alpha - 1)\varphi(\xi) + i\overline{\psi(\xi)} + \frac{\beta\omega(\xi)\overline{\omega'(\xi)}}{\overline{\varphi'(\xi)}} = H(\xi)\right], \tag{12}$$

and

$$\alpha\varphi(\xi) + i\overline{\psi(\xi)} + \frac{\beta\omega(\xi)\omega'(\xi)}{\varphi'(\xi)} = iG(\xi). \tag{13}$$

3 General Equations For the Identification of Harmonic Shapes

The schematic used to identify the harmonic shape is depicted in Fig. 1. In a two-dimensional infinite sheet without any hole or inclusion, the original displacement and stress fields are determined by two analytic functions, φ_o and ψ_o. When a hole or inclusion is introduced into the sheet, the original stress field is locally distorted near the hole or inclusion. Suppose the final stress field is expressed by the two analytic functions $\varphi(\xi)$ and $\psi(\xi)$ through relations of the form

$$\varphi(\xi) = \varphi_o(\xi) + \varphi_p(\xi), \quad \psi(\xi) = \psi_o(\xi) + \psi_p(\xi). \tag{14}$$

Here, the subscript "p" refers to the perturbed terms resulting from the introduction of the hole or inclusion and the subscript "o" indicates the original field.

Asymptotically, it is clear that we require $\varphi(\xi)$ and $\psi(\xi)$ to approach $\varphi_o(\xi)$ and $\psi_o(\xi)$, respectively, as $|\xi| \to \infty$. To determine the geometry of the harmonic shape, we require that the final stress field satisfies a type of harmonic condition similar to that proposed by Bjorkman and Richards [3], in which the sum of normal stresses remains unperturbed everywhere in the field.

Consequently, in the ξ plane, substituting Eq. (14) into Eq. (11), we obtain that

$$\varphi_p(\xi) = 0. \tag{15}$$

In addition, it is clear that, at infinity, the perturbed function ψ_p must satisfy

$$\psi_p(\xi) = 0, \quad |\xi| \to \infty. \tag{16}$$

Since $\psi_p(\xi)$ is analytic outside the unit circle $\gamma: |\xi| = 1$, it possesses the following property

$$\int_\gamma \frac{\overline{\psi_p(\sigma)}}{\sigma - \xi} d\sigma = 0 \quad \text{for any} \quad |\xi| > 1. \tag{17}$$

Multiplying each term in Eq. (12) or (13) by $d\sigma/(\sigma - \xi)$ and integrating over γ, we obtain, in each case, the following equations for the mapping function ω and hence the harmonic shape:

$$(\alpha - 1)\int_\gamma \frac{\varphi_o(\sigma)}{\sigma - \xi} d\sigma + \int_\gamma \frac{i\overline{\psi_o(\sigma)}}{\sigma - \xi} d\sigma + \beta\int_\gamma \frac{\omega(\sigma)\overline{\omega'(\sigma)}}{\varphi_o'(\sigma)(\sigma - \xi)} d\sigma = \frac{1}{2\mu i}\int_\gamma \frac{H(\sigma)}{\sigma - \xi} d\sigma, \tag{18}$$

or

$$\alpha\int_\gamma \frac{\varphi_o(\sigma)}{\sigma - \xi} s\sigma + \int_\gamma \frac{i\overline{\psi_o(\sigma)}}{\sigma - \xi} d\sigma + \beta\int_\gamma \frac{\omega(\sigma)\overline{\omega'(\sigma)}}{\varphi_o'(\sigma)(\sigma - \xi)} d\sigma = i\int_\gamma \frac{G(\sigma)}{\sigma - \xi} d\sigma. \tag{19}$$

For a given loading at infinity and prescribed conditions along the interior boundary, the mapping function $z=\omega(\xi)$ can be obtained from Eq. (18) or Eq. (19). Substitution of the mapping function into the interior boundary condition described by Eq. (12) or Eq. (13) then gives the boundary value of the analytic function $\psi_p(\xi)$ on the unit circle γ. Next, taking into account the asymptotic condition, Eq. (16), we can determine the analytic function $\psi_p(\xi)$ in the entire domain. Finally, the displacement and stress fields are obtained using Eqs. (4)–(6).

4 Examples: Harmonic Shapes in the Case of Nonuniform Loading

Under a uniform biaxial loading, the harmonic shape is an ellipse that depends on the ratio of the biaxial stresses and the mean normal stress [5]. Consider now the case when the loading is nonuniform. It is assumed that the remote stress field is given by two analytic functions φ_o and ψ_o as

$$\varphi_o(z)=Az, \quad \psi_o(z)=Bz+Cz^2. \tag{20}$$

The corresponding remote stresses are given by

$$\sigma_{11}^{\infty}=N_1-Ky, \quad \sigma_{22}^{\infty}=N_2+Ky, \quad \sigma_{21}^{\infty}=\sigma_{12}^{\infty}=-kx, \tag{21}$$

where N_1, N_2, and K are real constants related by the equations

$$\frac{N_2+N_1}{2}=2\mu i\left[(\alpha-1)A+\frac{\beta}{\bar{A}}\right], \quad \frac{N_2-N_1}{2}=2\mu i(i\bar{B}), \quad C=\frac{K}{4\mu}i. \tag{22}$$

As in [6], the fact that the considered stress field is unbounded at infinity causes no particular difficulty, since in any (finite) structure, one need only consider a sufficiently large (finite) region in which all stresses are then bounded.

For convenience, in what follows, we denote by Γ and Γ' the expressions

$$\Gamma=\frac{N_2+N_1}{2}, \quad \Gamma'=\frac{N_2-N_1}{2}. \tag{23}$$

For a rigid inclusion, the displacements on the interface between the surrounding material and the inclusion must be zero, thus $G(\sigma)=\omega(\sigma)$. Thus Eq. (19) simplifies to

$$-\eta\Gamma\int_{\gamma}\frac{\omega(\sigma)}{\sigma-\xi}d\sigma+\Gamma'\int_{\gamma}\frac{\overline{\omega(\sigma)}}{\sigma-\xi}d\sigma+\frac{K}{2}i\int_{\gamma}\frac{\overline{[\omega(\sigma)]^2}}{\sigma-\xi}d\sigma=0, \tag{24}$$

where

$$\eta=\frac{1+\sqrt{1+4(1-\alpha)\beta(2\mu/\Gamma)^2}}{(1-\alpha)}-1-\frac{2\mu}{\Gamma}, \tag{25}$$

and the sign of η has been determined by the condition that the deformation must have the same sign as the stress under equal biaxial loading.

In this case, it is found that mapping function $\omega(\xi)$ can be expanded in a Laurent series outside the unit circle as

$$\omega(\xi)=R\left(\xi+\frac{a_1}{\xi}+\frac{a_2}{\xi^2}+\cdots.\right), \tag{26}$$

where R is a real scaling constant and a_i, $i=1,\ldots$, are constants.

Substituting Eq. (26) into Eq. (24), evaluating the integral and comparing like powers of ξ, we obtain

$$a_1=\frac{\Gamma'}{\eta\Gamma}, \quad a_2=\lambda i, \quad a_3=a_4=\cdots=0 \tag{27}$$

with

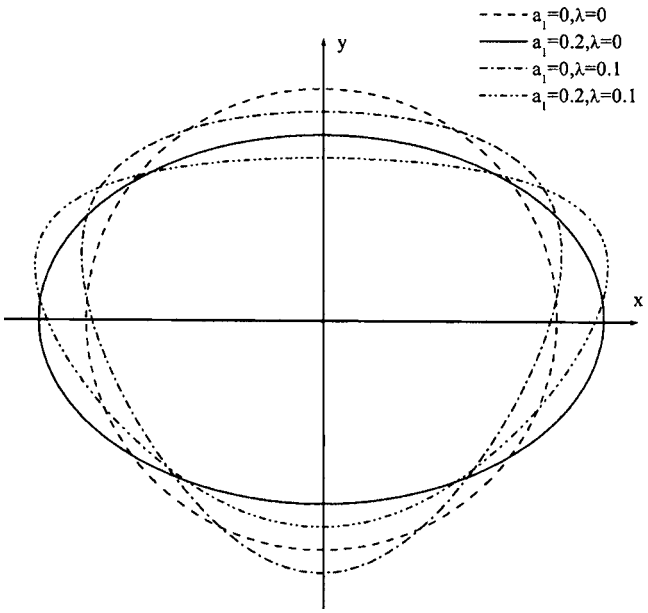


Fig. 2 Harmonic shape for different loading scenarios

$$\lambda=\frac{KR}{2\eta\Gamma}. \tag{28}$$

Thus the mapping function is finally given by

$$z=\omega(\xi)=R\left(\xi+\frac{a_1}{\xi}+\frac{\lambda i}{\xi^2}\right). \tag{29}$$

From Sec. 3, the stresses are then obtained as

$$\sigma_{11}+i\sigma_{21}=\Gamma-\left(\Gamma'+KiR\bar{\xi}-\Gamma\eta\frac{1}{\bar{\xi}^2}\right)\left(1-\frac{a_1}{\bar{\xi}^2}+2\frac{\lambda i}{\bar{\xi}^3}\right)^{-1}, \tag{30}$$

$$-i\sigma_{12}+\sigma_{22}=\Gamma+\left(\Gamma'+KiR\bar{\xi}-\frac{\Gamma\eta}{\bar{\xi}^2}\right)\left(1-\frac{a_1}{\bar{\xi}^2}+2\frac{\lambda i}{\bar{\xi}^3}\right)^{-1}. \tag{31}$$

When $K=0$, the results reduce to those obtained under uniform biaxial loading, namely that the harmonic inclusion is elliptical [5]. However, for a nonuniform loading described by the second-order term of ψ_o, the shape of the harmonic inclusion is somewhat different: the influence of the nonuniformity [characterized by λ in Eq. (28)] is shown clearly in Fig. 2.

When a higher-order term in ψ_o is considered, for example, $\psi_o(z)=Bz+Cz''$, one can obtain the corresponding harmonic inclusion using the same procedure described above with the mapping function described by parameters a_1 to a_n.

In the case of free holes, we can obtain the corresponding harmonic shape by setting $\eta=-1$ in Eqs. (27) and (28).

To ensure the existence of these harmonic shapes, the derivative of the mapping function $\omega'(\xi)$ must be nonzero in the region $|\xi|\geqslant 1$. By solving the equation $\omega'(\xi)=0$, and then requiring that all the roots are located inside the unit circle, we can obtain that a_1 and λ must lie within the region enclosed by the curve shown in Fig. 3. As in the case of uniform biaxial loading, the choice of a_1 still falls within the range $[-1,1]$. The range of permissible λ decreases from 0.5 to 0 as the absolute value of a_1 increases from 0 to 1. In the case of uniform remote loading ($\lambda=0$), the harmonic shape exists only if $|a_1|<1$, which depends only on the remote loading and material properties. However, in the case of nonuniform remote loading (of any order), the harmonic shape is also

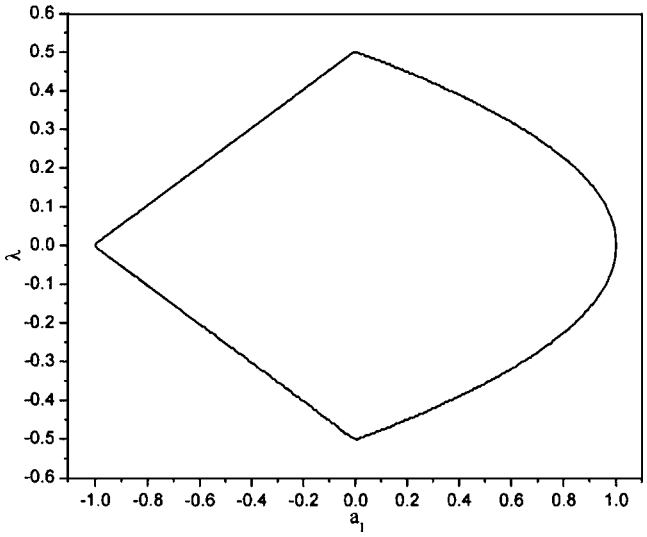


Fig. 3 Condition for the existence of the harmonic shape

dependent on the absolute size of the shape, (inclusion or hole) characterized by the parameter R, as indicated in Eq. (28). This particular result, which has been shown to be true also in linear elasticity [6], has significant implications for engineering design.

5 Conclusions

We consider the design of harmonic shapes (holes or inclusions) in a particular class of compressible hyperelastic materials of harmonic type undergoing finite plane deformations while subjected to *nonuniform* remote stress. Harmonic shapes are characterized by a ("harmonic condition") imposed on the final stress field. The ("harmonic condition") used here is a generalization of the original condition used in the corresponding problems of linear elasticity: specifically, that the sum of the normal stresses in the original stress field remains unchanged everywhere after the introduction of the harmonic hole or inclusion.

It has been shown that the influence of the nonuniformity of the remote loading has a significant effect on the design of the harmonic shape and on the condition for its existence.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada through Grant No. NSERC OGP 115112.

References

[1] Cherepanov, G. P., 1974, "Inverse Problem of the Plane Theory of Elasticity," Prikl. Mat. Mekh. **38**, pp. 963–979.
[2] Mansfield, E. H., 1953, "Neutral Holes in Plane Sheet—Reinforced Holes Which Are Elastically Equivalent to the Uncut Sheet," Q. J. Mech. Appl. Math. **4**, pp. 370–378.
[3] Bjorkman, G. S., Jr., and Richards, R., Jr., 1976, "Harmonic Holes—An Inverse Problem in Elasticity," ASME J. Appl. Mech. **43**, pp. 414–418.
[4] Richards, R., Jr., and Bjorkman, G. S., Jr., 1980, "Harmonic Shapes and Optimum Design," J. Eng. Mech. Div., Proc. ASCE, **106**(EM6), pp. 1125-1134.
[5] Wang, G. F., Schiavone, P., and Ru, C-Q., 2004, "Harmonic Shapes in Finite Elasticity" (submitted).
[6] Bjorkman, G. S., Jr., and Richards, R., Jr., 1979, "Harmonic Holes for Nonconstant Fields," ASME J. Appl. Mech. **46**, pp. 573–576.
[7] John, F., 1960, "Plane Strain Problems For a Perfectly Elastic Material of Harmonic Shape." Commun. Pure Appl. Math. **XIII**, pp. 239–290.
[8] Knowles, J. K., and Sternberg, E., 1975, "On the Singularity Induced by Certain Mixed Boundary Conditions in Linearized and Nonlinear Elastostatics," Int. J. Solids Struct. **11**, pp. 1173–1201.
[9] Varley, E., and Cumberbatch, E., 1980, "Finite Deformation of Elastic Material Surrounding Cylindrical Holes," J. Elast. **10**, pp. 341–405.
[10] Li, X., and Steigmann, D. J., 1993, "Finite Plane Twist of an Annular Membrane," Q. J. Mech. Appl. Math. **46**, pp. 601–625.
[11] Ru, C. Q., 2002, "On Complex-Variable Formulation For Finite Plane Elastostatics of Harmonic Materials," Acta Mech. **156**, pp. 219-234.

The Pseudo-Rigid Rolling Coin

Marcelo Epstein
e-mail: epstein@enme.ucalgary.ca

R. Ivan Defaz
e-mail: rdefaz@yahoo.com

Department of Mechanical and Manufacturing Engineering,
The University of Calgary,
Calgary, Alberta T2N 1N4, Canada

A pseudo-rigid coin is a thin disk that can deform only to the extent of undergoing an arbitrary affine deformation in its own plane. The coupling of the classical rolling problem with this deformability, albeit limited, may shed light on such phenomena as the production of noise by a twirling dish. From the point of view of analytical dynamics, one of the interesting features of this problem is that the rolling constraint turns out to be nonholonomic even in the case of motion on a straight line in a vertical plane. After the analytical formulation of the general problem, explicit solutions are obtained for special shape-preserving motions. For more general motions, numerical studies are carried out for various initial conditions. [DOI: 10.1115/1.1979515]

1 Introduction

The theory of pseudo-rigid bodies [1–3] may be seen as emerging from the following simple idea: if a deformable body is constrained to undergo only homogeneous affine deformations (namely, with a space-wise constant deformation gradient), its degrees of freedom are so severely curtailed that its motion is governed no longer by partial differential equations but by a system of ordinary differential equations. In other words, the problem falls from the realm of Continuum Mechanics into the domain of Classical Analytical Mechanics. In fact, the limitation to affine deformations, apart from having the merit of simplicity, is not the essential point: any restriction of the motion to a family governed by a finite number of parameters will do. The methods of Rayleigh-Ritz and Galerkin and their heir, the method of finite elements, are typical examples. Nevertheless, the particular case of an affine deformation (which from the point of view of the finite-element method would be equivalent to a single constant-strain element) lends itself to a rather explicit dynamical analysis by means of Lagrangian or Hamiltonian mechanics. In the case of a thin structure, such as the coin that occupies us, another interesting possibility would be to allow for a *bending* deformation with a prescribed shape. In fact, for the study of the noise produced by rolling wheels, this should be considered as the most important degree of freedom. In this paper, however, we shall focus only on the in-plane deformability of the disk, while our studies on the bending deformation are underway. Possible applications include the dynamics of tires and other flexible rolling devices. Rings, in particular, are likely to be more flexible in their main plane than in transverse bending. Another interesting potential application is the analysis of the noise produced by the rolling (and sliding) of a disk on a horizontal table. Interest in this classical problem has been recently revived by a number of publications [4–7] and by the commercialization of a toy known as Euler's disk. For specificity, we develop (and, in some cases, solve) the equations that correspond to a solid incompressible neo-Hookean disk.

There has been considerable attention devoted to pseudo-rigid bodies, also called *Cosserat points* [8], in the recent past [9–13], ranging from the theoretical to the applied realms (impact problems, vehicle dynamics, granular media, etc.). Our own interest in this topic does not emanate from any position that would seem to suggest that pseudo-rigid bodies can shed light on fundamental problems of Continuum Mechanics. Quite to the contrary, we believe that pseudo-rigid bodies, belonging as they do to the realm of classical mechanics, offer a wealth of interesting problems not treated in the traditional literature which are worth exploring in their own right. To the best of our knowledge, the general problem of rolling, with its attendant nonholonomic constraints, has not been treated to date. The particular case of planar (vertical) rolling on a horizontal straight line has been tackled by Cohen and Sun [14]. Although this problem can be regarded as a particular case of the more general problem of rolling arbitrarily on a horizontal plane, we will deal with it separately in Sec. 2. (A direct numerical comparison with the results of [14] has not been carried out, since the constitutive equations used for the elastic material are different.) One reason for doing so, apart from the gain in simplicity, is the somewhat unexpected fact that, even in this highly simplified context, the rolling constraint turns out to be nonholonomic, in sharp contradistinction with the rigid-body counterpart. Only after we gather some specific knowledge from this case will we move on to the formulation of the general problem (Sec. 3). A detailed study is carried out of special *shape-preserving* motions, for which an explicit analytical solution can be obtained. A perhaps surprising by-product of the analysis is the discovery that the most general family of force-free motions of a free pseudo-rigid body does not include any nontrivial shape-preserving motion. Finally, numerical examples for more general motions with various initial conditions are presented and discussed in Sec. 4.

2 Straight Rolling in a Vertical Plane

2.1 Kinematic Analysis. Following the continuum mechanics paradigm, we adopt a *reference configuration* in a Euclidean space with Cartesian coordinates X, Y, Z. We purposely avoid the use of indexed quantities to deal with as explicit equations as possible, both analytically and numerically. In this reference space lies a perfectly circular disk (the coin) with radius R, thickness h (with $h/R \ll 1$), and center at the origin of coordinates. The normal to the main plane of the coin is aligned with the Z-axis. Its mass density ρ is uniform. The material properties are assumed to be uniform and the coin to be in a homogenenous reference configuration. Moreover, we assume the material to be fully isotropic. We adopt a system of Cartesian coordinates x, y, z in physical space, placing the y-axis vertically. For the purpose of visualization, it is convenient to think of both coordinate systems as being immersed in a single space and being coincident (or at least parallel). We are interested in studying configurations in physical space whereby the coin retains its planarity while undergoing only affine (homogeneous) deformations while the main plane remains vertical as the coin rolls on the x-axis. An affine deformation consists of a deformation with constant gradient, namely,

$$\mathbf{r}(t) = \mathbf{r}_0(t) + \mathbf{F}(t)\mathbf{R}, \tag{1}$$

where $\mathbf{R}$ and $\mathbf{r}$ are the position vectors in the reference and spatial configurations, respectively, $\mathbf{r}_0$ is the spatial position of the center

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, July 7, 2004; final revision, December 21, 2004. Associate Editor: O. M. O'Reilly. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

of the coin, **F** is the deformation gradient, and t is time. According to the assumptions listed above, this formula has the following component form:

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} x_0 \\ y_0 \\ 0 \end{Bmatrix} + \begin{bmatrix} x_1 & x_2 & 0 \\ y_1 & y_2 & 0 \\ 0 & 0 & z_3 \end{bmatrix} \begin{Bmatrix} X \\ Y \\ Z \end{Bmatrix}, \tag{2}$$

where $(x_0, y_0, 0)$ is the position of the center. The entries of the square matrix are all nondimensional. In particular, z_3 represents the relative change in thickness. All the indexed x's, y's and z's are functions of time. Without loss of generality, we may assume the material to be incompressible, so that $\det \mathbf{F} = 1$. Denoting:

$$\Delta = \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} = x_1 y_2 - x_2 y_1, \tag{3}$$

the incompressibility condition can be expressed as

$$z_3 = 1/\Delta. \tag{4}$$

Differential equations of motion will be formulated in terms of the x's and the y's, and the time evolution of z_3 will follow from Eq. (4). In other words, due to the assumed extreme thinness of the coin, there is no dynamical contribution from the z-direction.

To complete the kinematic description of the motion, we need to introduce the constraint of no-sliding contact. In fact, this constraint results in two equations: the first equation represents the fact that at all times the lowest point of the deformed coin has zero height above the ground, while the second equation ensures that the speed of the point of contact vanishes. To find the lowest point of the coin, we introduce an angular parameter γ in the reference configuration (see Fig. 1) and write the parametric equation of the perimeter as

$$\begin{aligned} X &= R \sin \gamma, \\ Y &= -R \cos \gamma. \end{aligned} \tag{5}$$

Introducing these formulas in Eq. (3), we obtain the corresponding equations for the deformed perimeter as

$$\begin{aligned} x &= x_0 + x_1 R \sin \gamma - x_2 R \cos \gamma, \\ y &= y_0 + y_1 R \sin \gamma - y_2 R \cos \gamma, \end{aligned} \tag{6}$$

which, as expected, are the parametric equations of an ellipse. The lowest point of this ellipse occurs at a value γ_c satisfying

$$\frac{dy}{d\gamma} = y_1 R \cos \gamma_c + y_2 R \sin \gamma_c = 0. \tag{7}$$

Using this result in Eq. (6) via standard trigonometric identities, and setting $y=0$ (for the point of contact), we obtain the following constraint:

$$y_0^2 = R^2 (y_1^2 + y_2^2), \tag{8}$$

where we have incorporated the ambiguity of the trigonometric functions (or the square root) by squaring. Obviously, we should consider always $y_0 > 0$, the negative value corresponding to having the uppermost point (rather than the lowest) in contact with the ground.

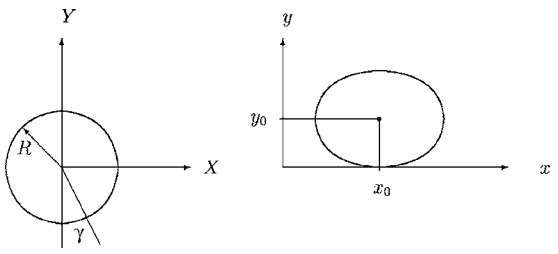


Fig. 1 Geometry and deformation

To obtain the second constraint equation, we differentiate Eq. (6) with respect to time and obtain the velocity of the point of contact as

$$\begin{aligned} \dot{x} &= \dot{x}_0 + \dot{x}_1 R \sin \gamma_c - \dot{x}_2 R \cos \gamma_c, \\ \dot{y} &= \dot{y}_0 + \dot{y}_1 R \sin \gamma_c - \dot{y}_2 R \cos \gamma_c. \end{aligned} \tag{9}$$

It is clear that, by virtue of Eq. (8), the y-component of the velocity already vanishes identically. We are left with the imposition of the vanishing of the x-component, yielding the following constraint:

$$\dot{x}_0 y_0 = R^2 (\dot{x}_1 y_1 + \dot{x}_2 y_2), \tag{10}$$

where we have taken Eq. (8) into account. Unlike its counterpart for the case of the rigid coin, this constraint is *nonintegrable*.

2.2 Equations of Motion. The total kinetic energy of the coin is given by

$$K = \int\int\int \tfrac{1}{2} \rho (\dot{x}^2 + \dot{y}^2) dX dY dZ, \tag{11}$$

where, invoking the assumption that $h/R \ll 1$, we have neglected the contribution of the thickness-wise velocity components. Substituting the derivatives of Eq. (2), Eq. (11) yields

$$K = \frac{\rho h \pi R^2}{2} (\dot{x}_0^2 + \dot{y}_0^2) + \frac{\rho h \pi R^4}{8} (\dot{x}_1^2 + \dot{x}_2^2 + \dot{y}_1^2 + \dot{y}_2^2). \tag{12}$$

The potential energy of gravitation is

$$U_g = \rho h \pi R^2 g y_0, \tag{13}$$

where g is the acceleration of gravity. The energy associated with the deformation depends, of course, on the material. For the sake of specificity, we will assume a hyperelastic neo-Hookean material, with the following stored-energy function per unit volume:

$$w = \frac{G}{2} (\mathrm{trace}(\mathbf{F}^{\mathbf{T}} \mathbf{F}) - 3), \tag{14}$$

where G is the *shear modulus*. This equation presupposes the material to be incompressible, a constraint that we have already imposed by the *a posteriori* calculation of z_3, according to Eq. (4). Taking due consideration of Eqs. (3) and (4), we can write Eq. (14) more explicitly as

$$w = \frac{G}{2} \left(x_1^2 + x_2^2 + y_1^2 + y_2^2 + \frac{1}{\Delta^2} - 3 \right). \tag{15}$$

The total elastic stored energy

$$U_e = \int\int\int w dX dY dZ, \tag{16}$$

is evaluated as

$$U_e = \frac{h \pi R^2 G}{2} \left(x_1^2 + x_2^2 + y_1^2 + y_2^2 + \frac{1}{\Delta^2} - 3 \right). \tag{17}$$

Forming the Lagrangian $L = K - (U_g + U_e)$ and introducing the Lagrange multipliers λ and μ, respectively, for the constraints given in Eqs. (8) and (10), for whose treatment we adopt the standard Chetaev rule, we obtain the following equations of motion for the unknown functions x_0, y_0, x_1, y_1, x_2, y_2, λ, and μ

$$\ddot{x}_0 = \mu y_0, \tag{18}$$

$$\ddot{y}_0 = -g + \lambda y_0, \tag{19}$$

$$\ddot{x}_1 = -k \left(x_1 - \frac{y_2}{\Delta^3} \right) - 4 \mu y_1, \tag{20}$$

$$\ddot{y}_1 = -k\left(y_1 + \frac{x_2}{\Delta^3}\right) - 4\lambda y_1, \tag{21}$$

$$\ddot{x}_2 = -k\left(x_2 + \frac{y_1}{\Delta^3}\right) - 4\mu y_2, \tag{22}$$

$$\ddot{y}_2 = -k\left(y_2 - \frac{x_1}{\Delta^3}\right) - 4\lambda y_2, \tag{23}$$

where

$$k = \frac{4G}{\rho R^2}, \tag{24}$$

and where Δ is given by Eq. (3). These equations of motion must, of course, be supplemented with the constraint equations (8) and (10) themselves, for a total of 8 equations for the 8 unknown functions. For numerical purposes, it is convenient to further differentiate the constraints and eliminate algebraically the Lagrange multipliers. Such details are not discussed here, since they are standard computational techniques.

2.3 Driven Motions and the Conveyor-Belt Problem. We will call *driven motions* those that are induced by a prescribed motion of the center of the disk, just as an axle would. We start by assuming that only the translations x_0 and y_0 are prescribed, while the rotation is left to its own device. Equations (18) and (19) are then discarded, while the prescribed motion is plugged in, wherever else it appears. Of all possible driven motions there is one kind that deserves special attention, namely, the class of motions that are *shape preserving*. In particular, we are interested in a *stationary* shape-preserving motion, in which not only the shape is preserved, but it also appears to glide as if in translation, as opposed to one in which the shape is preserved but it appears to rotate. A practical instance of a stationary motion is that of a conveyor belt being driven by friction by a rubber tire. We start by specifying the driving motion as

$$x_0(t) = At, \tag{25}$$

and

$$y_0(t) = B, \tag{26}$$

where A and B are constants. The exact mathematical definition of a stationary motion might seem quite perplexing, except for the fact that the *polar decomposition theorem* of algebra comes to our help. In particular, the left polar decomposition of the deformation gradient:

$$\mathbf{F} = \mathbf{V}\mathbf{Q}, \tag{27}$$

where $\mathbf{V}$ is symmetric and positive definite and where $\mathbf{Q}$ is (proper) orthogonal, leads immediately to the conclusion that: a motion is stationary if, and only if, the matrix $\mathbf{V}$ is constant in time. Let, therefore:

$$\mathbf{V} = \begin{bmatrix} a & b & 0 \\ b & c & 0 \\ 0 & 0 & z_3 \end{bmatrix}, \tag{28}$$

where a, b, c are constants with $a+c>0$ and $ac-b^2>0$. For our particular choice of driving motion it is clear also that the matrix $\mathbf{Q}$ must represent a rotation at a constant angular speed ω_d, namely,

$$\mathbf{Q} = \begin{bmatrix} \cos\omega_d t & \sin\omega_d t & 0 \\ -\sin\omega_d t & \cos\omega_d t & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{29}$$

Substituting the above expressions into the equations of motion (20)–(23) and the constraints (8) and (10), it turns out that the time dependence cancels out and we obtain a system of 6 nonlinear algebraic equations involving A, B, ω_d, a, b, c, λ, and μ. Fortunately, it is possible to solve this system explicitly and obtain

$$a^2 = \frac{A^2}{2kR^2} + \sqrt{\left(\frac{A^2}{2kR^2}\right)^2 + \left(\frac{R}{B}\right)^2}, \tag{30}$$

$$b = 0, \tag{31}$$

$$c = \frac{B}{R}, \tag{32}$$

$$\omega_d = \frac{A}{Ra}, \tag{33}$$

$$4\lambda = (\omega_d^2 - k) + \frac{k}{a^2 c^4}, \tag{34}$$

$$\mu = 0. \tag{35}$$

The fact that $b=0$ means that the eigenvectors of $\mathbf{V}$ are aligned with the coordinate axes, so that a stationary solution of this kind is possible only with a "horizontal" ellipse. The fact that $\mu=0$ was to be expected, since μ represents the horizontal component of the ground reaction, which should vanish for a motion at constant speed. The value $c=B/R$ is also predictable, since c represents the minor axis of the ellipse, which is obviously dictated directly by the imposed value of y_0. To have a positive ground reaction (i.e., upward), we impose $\lambda>0$ and get the following restriction between driving height (B) and speed (A):

$$A^2 > kB^2\left(1 - \left(\frac{R}{B}\right)^6\right). \tag{36}$$

Naturally, this restriction is automatically satisfied if $B<R$, but higher values of B are permitted provided the driving speed is sufficiently fast so that the "centrifugal" expansion of the disk makes it enter in contact with the ground. [According to Eq. (19), if it so happens that the calculated λ is exactly equal to g/B, then the steady motion is not driven, but happens on its own, the shape of the ellipse being now prescribed by the weight and the speed A alone. An explicit formula to find the value of the speed A necessary to achieve a prescribed height B with no driving force can be found without great difficulty, but is not shown here. On the other hand, one can always interpret a driven motion as spontaneous in a gravitational environment whose intensity is adjusted to the value λB. For example, the values $R=1$, $k=1$, $A=1$, $B=\sqrt{6}/6$, yield the solution $a=\sqrt{3}$, $b=0$, $c=\sqrt{6}/6$, $\lambda=17/6$, $\mu=0$, $\omega_d=\sqrt{3}/3$. If we specify, with the right initial conditions, an acceleration of gravity of $17\sqrt{6}/36$, this stationary-shape motion will endure without the need of being driven.] The particular case in which λ is exactly zero corresponds to the free rotation of the disk, without constraints. For this case we obtain $a=c$ (circular shape). If we assume in addition that the ground is absent and suppose that what can be controlled is actually the angular speed ω_d, it follows from Eq. (34) that this speed is related to the radius of the deformed disk ($R'=cR$) by the formula

$$\omega_d = \sqrt{k\left(1 - \frac{1}{c^6}\right)}. \tag{37}$$

Surprisingly, we conclude that the driving angular speed cannot exceed the limiting value $\omega_{\max}=\sqrt{k}$, but that this value can be approximated from below very closely with relatively small deformed radii. For example, for $c=1.2$ we obtain $\omega_d \approx 0.816\ \omega_{\max}$, and for $c=2$ the value is $\omega_d \approx 0.993\ \omega_{\max}$. Thereafter, the energy needed to obtain a marginal increase in angular speed becomes very large. If the initial conditions specify a radius slightly different from the one corresponding to stationary shape, oscillations take place. When the driving angular speed is equal to the maxi-

mum value, the amplitude begins to increase with time, a growth that becomes asymptotically linear.

2.4 Small Oscillations. The phenomena just described are analogous to what would be observed when placing a spring-mass system radially on a rotating table. If the spring is linear, resonance will occur as the rotational speed of the table reaches the natural frequency of the system. The "centrifugal force" will then drive the mass to infinity at a constant acceleration. For angular speeds lower than the resonant one, there exists a position of dynamic equilibrium of the mass, and any small disturbance will cause oscillations to develop with a frequency that becomes smaller the closer the angular velocity is to the resonant value. Although, as we have discussed, the rotation of the free neo-Hookean disk is qualitatively similar, it is important to verify that the natural frequencies of the disk at rest are *higher* than the value $\omega_{\max}=\sqrt{k}$ calculated above. What this means is that there is no danger that some resonance will be triggered by the driving rotation before it reaches that value.

To verify that this is indeed the case, we set $\lambda=\mu=0$ and linearize the equations of motion (20)–(23) around the static equilibrium and obtain

$$\ddot{u}_1 = -k(u_1 - v_2 + 3(u_1 + v_2)), \tag{38}$$

$$\ddot{v}_1 = -k(v_1 + u_2), \tag{39}$$

$$\ddot{u}_2 = -k(u_2 + v_1), \tag{40}$$

$$\ddot{v}_2 = -k(v_2 - u_1 + 3(u_1 + v_2)), \tag{41}$$

where u_1, v_1, u_2, and v_2 are small increments of x_1, y_1, x_2, and y_2 around the equilibrium values 1,0,0, and 1, respectively. In matrix form we obtain

$$\begin{Bmatrix} \ddot{u}_1 \\ \ddot{v}_1 \\ \ddot{u}_2 \\ \ddot{v}_2 \end{Bmatrix} + k \begin{bmatrix} 4 & 0 & 0 & 2 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 2 & 0 & 0 & 4 \end{bmatrix} \begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}. \tag{42}$$

The corresponding natural frequencies are: $\omega_1=0$, $\omega_2=\omega_3=\sqrt{2k}$, and $\omega_4=\sqrt{6k}$. The zero frequency corresponds to the eigenvector

$$\begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{Bmatrix}, \tag{43}$$

which represents a rigid rotation (irrelevant when the rotation is being driven). The largest eigenvalue (ω_4) corresponds to the eigenvector

$$\begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{Bmatrix}, \tag{44}$$

representing a uniform expansion of the disk. Finally, the repeated eigenvalue corresponds to the following eigenspace:

$$\begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \end{Bmatrix} = C_1 \begin{Bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{Bmatrix} + C_2 \begin{Bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{Bmatrix}, \tag{45}$$

where C_1 and C_2 are arbitrary real numbers. This eigenspace represents all area-preserving ellipses in the plane (in the linear approximation, of course). Without pursuing this analytical study further, these results seem to indicate that no matter how the coin is initially propelled, there is no danger that a critical initial speed might exist such that resonance will be triggered.

3 General Rolling on a Horizontal Plane

3.1 Kinematics. Adopting the same notation and general scheme as before, we start from the most general affine deformation of our coin, namely,

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} x_0 \\ y_0 \\ z_0 \end{Bmatrix} + \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} \begin{Bmatrix} X \\ Y \\ Z \end{Bmatrix}, \tag{46}$$

where the only difference with the previous case is that there are no automatically vanishing entries. Nevertheless, even in the more general case, we do not allow the coin to deform arbitrarily. Indeed, neglecting the effects of transverse shear, we will require that the material lines initially perpendicular to the main plane of the coin remain so at all times. In other words, the last column of **F** must be orthogonal to the first two. The entries of the third column must, therefore, be proportional to the third row of the inverse of **F**, namely

$$x_3 = (y_1 z_2 - y_2 z_1)/\Gamma, \quad y_3 = -(x_1 z_2 - x_2 z_1)/\Gamma, \quad z_3 = (x_1 y_2 - x_2 y_1)/\Gamma. \tag{47}$$

Since we will again require that the total volume be preserved, the proportionality factor Γ must be adjusted to the value:

$$\Gamma = (y_1 z_2 - y_2 z_1)^2 + (x_1 z_2 - x_2 z_1)^2 + (x_1 y_2 - y_2 x_1)^2. \tag{48}$$

The no-sliding condition will now involve not two but three equations of constraint, since the velocity tangential to the plane of rolling has now two possible components, both of which need to vanish. By exactly the same procedure as before, we obtain the following three equations:

$$y_0^2 = R^2(y_1^2 + y_2^2), \tag{49}$$

$$\dot{x}_0 y_0 = R^2(\dot{x}_1 y_1 + \dot{x}_2 y_2), \tag{50}$$

$$\dot{z}_0 y_0 = R^2(\dot{z}_1 y_1 + \dot{z}_2 y_2), \tag{51}$$

the first two of which are identical with the ones for the two-dimensional case. The last two equations constitute nonholonomic constraints. [We note that taking the derivative of the holonomic constraint (49) one gets the formula $\dot{y}_0 y_0 = R^2(\dot{y}_1 y_1 + \dot{y}_2 y_2)$, which has an obvious similarity to the other two. This aesthetic notational point, although noticed, has not been exploited. It may serve to explore symmetries of solutions.]

3.2 Dynamic Equations. The kinetic energy can be obtained, as before, by a straightforward integration, yielding

$$K = \frac{\rho h \pi R^2}{2}(\dot{x}_0^2 + \dot{y}_0^2 + \dot{z}_0^2) + \frac{\rho h \pi R^4}{8}(\dot{x}_1^2 + \dot{x}_2^2 + \dot{y}_1^2 + \dot{y}_2^2 + \dot{z}_1^2 + \dot{z}_2^2). \tag{52}$$

The gravitational potential energy expression (13) remains unchanged. Finally, for the neo-Hookean material, the stored elastic energy is given by

$$U_e = \frac{h \pi R^2 G}{2}(x_1^2 + y_1^2 + z_1^2 + x_2^2 + y_2^2 + z_2^2 + x_3^2 + y_3^2 + z_3^2 - 3), \tag{53}$$

which, in view of Eqs. (47) and (48), can be written as

$$U_e = \frac{h \pi R^2 G}{2}\left(x_1^2 + y_1^2 + z_1^2 + x_2^2 + y_2^2 + z_2^2 + \frac{1}{\Gamma} - 3\right), \tag{54}$$

an expression that reduces to Eq. (17) in the purely two-dimensional case.

As before, we form the Lagrangian $L=K-(U_g+U_e)$ and intro-

duce the Lagrange multipliers λ, μ, and ν, respectively, for the constraints (49)–(51), we obtain the following equations of motion for the unknown functions x_0, y_0, z_0, x_1, y_1, z_1, x_2, y_2, z_2, λ, μ, and ν:

$$\ddot{x}_0 = \mu y_0, \tag{55}$$

$$\ddot{y}_0 = -g + \lambda y_0, \tag{56}$$

$$\ddot{z}_0 = \nu y_0, \tag{57}$$

$$\ddot{x}_1 = -k\left(x_1 - \frac{y_2(x_1y_2 - x_2y_1) + z_2(x_1z_2 - x_2z_1)}{\Gamma^2}\right) - 4\mu y_1, \tag{58}$$

$$\ddot{y}_1 = -k\left(y_1 - \frac{-x_2(x_1y_2 - x_2y_1) + z_2(y_1z_2 - y_2z_1)}{\Gamma^2}\right) - 4\lambda y_1, \tag{59}$$

$$\ddot{z}_1 = -k\left(z_1 - \frac{-x_2(x_1z_2 - x_2z_1) - y_2(y_1z_2 - y_2z_1)}{\Gamma^2}\right) - 4\nu y_1, \tag{60}$$

$$\ddot{x}_2 = -k\left(x_2 - \frac{-y_1(x_1y_2 - x_2y_1) - z_1(x_1z_2 - x_2z_1)}{\Gamma^2}\right) - 4\mu y_2, \tag{61}$$

$$\ddot{y}_2 = -k\left(y_2 - \frac{x_1(x_1y_2 - x_2y_1) - z_1(y_1z_2 - y_2z_1)}{\Gamma^2}\right) - 4\lambda y_2, \tag{62}$$

$$\ddot{z}_2 = -k\left(z_2 - \frac{x_1(x_1z_2 - x_2z_1) + y_1(y_1z_2 - y_2z_1)}{\Gamma^2}\right) - 4\nu y_2, \tag{63}$$

where Γ is given by Eq. (48). When supplemented with the constraints (49)–(51), these equations constitute an algebraic-differential system to solve for the 12 unknown functions.

3.3 Steady Motions: Pivoting. The complicated figure-skatinglike patterns traced by the disk as it rotates and deforms should be left perhaps for numerical investigation, and we will show some examples in the next sections. On the other hand, the possibility of inducing cylindrical and conical motions, whether by driving the movement or by specifying the right initial conditions and letting the system move of its own accord, is worthy of analytical study. The easiest example is that of vertical *pivoting* about a fixed point of contact. As before, we will concentrate on shape-preserving motions. Noting that in this case (due to the assumed fixity of the point of contact) the motion is actually a *material* rotation, we use the *right* polar decomposition to express the deformation gradient as

$$\mathbf{F} = \mathbf{RU} = \begin{bmatrix} \cos\omega t & 0 & \sin\omega t \\ 0 & 1 & 0 \\ -\sin\omega t & 0 & \cos\omega t \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & \dfrac{1}{ac} \end{bmatrix}. \tag{64}$$

The diagonal nature of $\mathbf{U}$ follows in this case by an argument of symmetry. Explicitly, we have

$$x_1 = a\cos\omega t, \quad x_2 = 0, \quad y_1 = 0, \quad y_2 = c, \qquad z_1 = -a\sin\omega t, \quad z_2 = 0. \tag{65}$$

Assuming the pivot to be at the origin, we complete the description of the motion by setting $x_0 = z_0 = 0$ and $y_0 =$ constant. A direct substitution into Eq. (48) yields

$$\Gamma = a^2c^2. \tag{66}$$

Finally, a substitution of all these values into the equations of motion and the constraints, yields the following two relations:

$$\omega^2 = k\left(1 - \frac{1}{a^4c^2}\right), \tag{67}$$

$$\left(\frac{1}{a^2c^4} - 1\right) = \frac{4g}{kRc}. \tag{68}$$

It follows from Eq. (67) that there is an upper limit ($\sqrt{k}$) for the angular velocity. This fact, consistent with Eq. (37) obtained for the in-plane spinning, is due to the assumed neo-Hookean behavior. For any given ω, the explicit expressions for a and c involve the solution of a cubic equation, which will not be displayed. On the other hand, one can easily show that $c<1$ and $a>1$, as expected. For $\omega=0$ we recover the same equations for the static deformation that would emerge from the system (30)–(35) with $A=0$, etc. Naturally, in the three-dimensional context, the static solution is unstable. Taking into consideration the static values c_s and a_s, one can refine the above bounds to obtain $c \leqslant c_s$ and $a \geqslant a_s$.

3.4 Steady Motions: General Case. So far, we have considered only very particular cases of steady (i.e., shape-preserving) motions, and we found that they could be expressed with the help of one of the polar decompositions of the deformation gradient. We ask now the question: What is the most general form of a shape preserving motion of an object?

Suppose that an affinely deformable object is given in a stress-free reference configuration, in which, moreover, the material symmetry group is the full orthogonal group $\mathcal{O}$. The reason for imposing these restrictions (which we have already imposed on our pseudo-rigid coin, anyway) is that we want to avoid the question of the interplay between material and geometrical symmetry. In other words, we want the motion to be congruent not only geometrically, but also materially at all times. Let the geometrical symmetry group of the object in this reference configuration be $\mathcal{G} \subset \mathcal{O}$, namely, a group consisting of some rotations and reflections. This group controls the possible shape preserving motions. Let $\mathbf{G}(t) \in \mathcal{G}$ and $\mathbf{R}(t) \in \mathcal{O}$ be time-dependent orthogonal transformations, and let $\mathbf{U}$ be a time independent symmetric positive-definite referential tensor. Then, the most general shape-preserving motion of the object is given by

$$\mathbf{r}(t) = \mathbf{r}_0(t) + \mathbf{R}(t)\mathbf{U}\mathbf{G}(t). \tag{69}$$

If the object has no geometrical symmetries, then the only way to achieve a shape preserving motion (according to this formula) is to deform it arbitrarily by means of a fixed $\mathbf{U}$ and follow this by a time-dependent rotation and a translation. If, on the other hand, geometrical symmetries are available, we are allowed to use them first, without altering the resulting shape. In the case of the coin, the symmetry group $\mathcal{G}$ is the group of all rotations around the Z-axis, reflections about the X-Y plane and rotation of magnitude π around the X-axis. If we assume the function $\mathbf{G}(t)$ to be connected, then the only part that matters is the subgroup of all rotations about Z.

We will now apply these concepts to describe and analyze a cylindrical steady motion of the pseudo-rigid coin. The coin shall remain always vertical while revolving around the spatial y-axis in a circular path of radius r. Because of the assumed verticality, this radius can be measured either to the center of the coin or to the point of contact with the ground (x-z plane). Let ϕ denote the angle from z to the line between the origin and the point of contact, measured positively according to the right-hand rule. We want the main plane of the coin to be constantly perpendicular to this last line (so that the motion can be called "cylindrical"). We assume the motion to be actually driven by a rigid horizontal arm, at a given height B from the ground, connecting the center of the coin with the y-axis. The connection at the center of the coin is assumed to be pin-jonted, so as to allow free rotations. Finally, we

assume the arm to be rotating at a constant rate $\dot{\phi}=\alpha$, while the applied time-dependent geometrical symmetry is a rotation ψ about Z, also at a constant rate $\dot{\psi}=\beta$.

Because the motion is driven, we disregard the first three equations of motion, Eqs. (55)–(57). The deformation gradient is now given by

$$\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & \dfrac{1}{ac} \end{bmatrix} \times \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{70}$$

or

$$\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} = \begin{bmatrix} a\cos\phi\cos\psi & -a\cos\phi\sin\psi & \dfrac{\sin\phi}{ac} \\ c\sin\psi & c\cos\psi & 0 \\ -a\sin\phi\cos\psi & a\sin\phi\sin\psi & \dfrac{\cos\phi}{ac} \end{bmatrix}. \tag{71}$$

Substitution into Eq. (48) yields:

$$\Gamma = a^2c^2. \tag{72}$$

Plugging these results into the equations of motion, one can observe that they can be fulfilled identically for all times if, and only if, the following relations hold:

$$y_0 = Rc, \tag{73}$$

$$r\alpha = -R\beta a, \tag{74}$$

$$\alpha^2 + \beta^2 = k\left(1 - \frac{1}{a^4c^2}\right), \tag{75}$$

$$4\lambda = \beta^2 - k\left(1 - \frac{1}{a^2c^4}\right), \tag{76}$$

$$2\mu c = -a\alpha\beta\sin\phi, \tag{77}$$

$$2\nu c = -a\alpha\beta\cos\phi. \tag{78}$$

Fixing, for example, the values of the height y_0, the arm length r, and the angular velocity α of the arm, we can solve for the shape of the ellipse (a and c), the spin β, and the ground reaction λ, μ, and ν. We observe that the horizontal ground reaction is radial (centrifugal), as expected. Notice that Eq. (74) is the pseudo-rigid version of the no-sliding condition $r\alpha=R\beta$ for the rigid counterpart.

Intuitively, we realize that the cylindrical motion just discussed cannot occur without it being driven, since the weight does not have a moment-arm to produce the necessary change of angular momentum. This intuitive observation is confirmed by checking that Eqs. (55) and (57) could not be satisfied by any solution of Eqs. (73)–(78). To have the possibility of a self-sustaining steady motion, we need to consider a *conical* movement, just like a twirling dish. Let θ be the angle of tilt, which we will assume positive if the top of the dish is closer than the bottom to the y-axis. To locate the dish at an arbitrary position of its conical motion, we proceed in 4 steps, 3 of them identical to the cylindrical motion: (i) Rotation of the circular dish about the (material) Z-axis by an amount ψ; (ii) Deformation to a time-independent "horizontal" ellipse; (iii) Tilting by a time-independent angle θ (clockwise) around the spatial x-axis; (iv) Rotation by an angle ϕ about the spatial y-axis. These steps represent just the rotations. The translation is given, as before, by a circular motion of the center of the dish consistent with the ϕ-rotation. The deformation gradient is now given by

$$\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \times \begin{bmatrix} a & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & \dfrac{1}{ac} \end{bmatrix} \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{79}$$

By direct substitution into Eq. (48) we obtain

$$\Gamma = a^2c^2, \tag{80}$$

which is the same result as for cylindrical motion. A detailed substitution into the equations of motion leads to the conclusion that they can be satisfied identically if, and only if, the following relations between the parameters hold:

$$y_0 = Rc\cos\theta, \tag{81}$$

$$r\alpha = -R(\beta a + c\alpha\sin\theta), \tag{82}$$

$$\alpha^2 + \beta^2 + \frac{2c}{a}\alpha\beta\sin\theta = k\left(1 - \frac{1}{a^4c^2}\right), \tag{83}$$

$$4\lambda = \beta^2 - k\left(1 - \frac{1}{a^2c^4}\right), \tag{84}$$

$$4\mu\cos\theta = \left(-\frac{2a}{c}\alpha\beta - (\alpha^2+\beta^2)\sin\theta + k\sin\theta\left(1 - \frac{1}{a^2c^4}\right)\right)\sin\phi, \tag{85}$$

$$4\nu\cos\theta = \left(-\frac{2a}{c}\alpha\beta - (\alpha^2+\beta^2)\sin\theta + k\sin\theta\left(1 - \frac{1}{a^2c^4}\right)\right)\cos\phi. \tag{86}$$

Unlike the case of cylindrical motion, these equations do not involve a contradiction with Eqs. (55) and (57) and, therefore, the possibility exists that conical motion may be self-sustaining. A case of particular interest is that for which $r=0$, namely, with the center of the dish constantly above the origin. In the case of cylindrical motion, the movement would degenerate into pivoting, but in the case of conical motion we obtain the everyday experience of the wobbling of a dish. For this case we obtain that the ground reaction is perfectly vertical ($\mu=\nu=0$) and that the rates of rotation are related by

$$\beta = -\frac{c}{a}\alpha\sin\theta. \tag{87}$$

Suppose that the tilt θ and the angular speed α are given. Then, the deformed steady shape is obtained by solving the following two simultaneous equations:

$$\frac{1}{a^2c^4} + \frac{1}{a^4c^2} = 2, \tag{88}$$

and

$$\frac{\dfrac{1}{a^4c^2} - 1}{\dfrac{c^2}{a^2}\sin^2\theta - 1} = \frac{\alpha^2}{k}. \tag{89}$$

We then calculate the Lagrange multiplier λ from Eq. (84). This motion will be self-sustaining for an acceleration of gravity g

$=\lambda Rc\cos\theta$.

Setting the acceleration of gravity to zero, we should be able to obtain the most general shape-preserving motion of a pseudo-rigid dish free in space. We expect in this case (as in its classical rigid counterpart) that we will no longer be at freedom to specify both the tilt and the angular speed. The extra condition is obtained by setting λ in Eq. (84) to zero. Surprisingly, the equation obtained is in contradiction with Eqs. (87)–(89), thus implying that there do not exist shape-preserving free conical motions, except pivoting and rotation in a plane. (It should be pointed out that these two extreme cases may elude the general formulas, since they were obtained by simplifying by trigonometric functions that may vanish at 0 or $\pi/2$.)

In closing this section we note that, although the calculations were carried out for a neo-Hookean material only, similar solutions exist for any other hyperelastic material as long as it is fully isotropic. Lack of isotropy will in general preclude these shape-preserving solutions. (For instance, the quantity Γ will no longer be constant.)

4 Numerical Studies

4.1 Nondimensionalization. For the purpose of numerical studies, it is best to obtain a nondimensionalized version of the dynamical equations, so as to minimize and rationalize the number and type of examples. Define the nondimensional space and time variables by means of the formulas

$$\tilde{x}_0 = x_0/R, \quad \tilde{y}_0 = y_0/R, \quad \tilde{z}_0 = z_0/R, \tag{90}$$

and

$$\tilde{t} = t\sqrt{k}. \tag{91}$$

We note that the variables $x_1, y_1, \ldots, z_3$ are already nondimensional by nature. Denoting:

$$(\,)' := \frac{d}{d\tilde{t}}(\,), \tag{92}$$

we can rewrite the dynamic equations (55)–(63) and the constraints (49)–(51) in the following nondimensional form:

$$\tilde{x}_0'' = \tilde{\mu}\tilde{y}_0, \tag{93}$$

$$\tilde{y}_0'' = -\tilde{g} + \tilde{\lambda}\tilde{y}_0, \tag{94}$$

$$\tilde{z}_0'' = \tilde{\nu}\tilde{y}_0, \tag{95}$$

$$x_1'' = -\left(x_1 - \frac{y_2(x_1y_2 - x_2y_1) + z_2(x_1z_2 - x_2z_1)}{\Gamma^2}\right) - 4\tilde{\mu}y_1, \tag{96}$$

$$y_1'' = -\left(y_1 - \frac{-x_2(x_1y_2 - x_2y_1) + z_2(y_1z_2 - y_2z_1)}{\Gamma^2}\right) - 4\tilde{\lambda}y_1, \tag{97}$$

$$z_1'' = -\left(z_1 - \frac{-x_2(x_1z_2 - x_2z_1) - y_2(y_1z_2 - y_2z_1)}{\Gamma^2}\right) - 4\tilde{\nu}y_1, \tag{98}$$

$$x_2'' = -\left(x_2 - \frac{-y_1(x_1y_2 - x_2y_1) - z_1(x_1z_2 - x_2z_1)}{\Gamma^2}\right) - 4\tilde{\mu}y_2, \tag{99}$$

$$y_2'' = -\left(y_2 - \frac{x_1(x_1y_2 - x_2y_1) - z_1(y_1z_2 - y_2z_1)}{\Gamma^2}\right) - 4\tilde{\lambda}y_2, \tag{100}$$

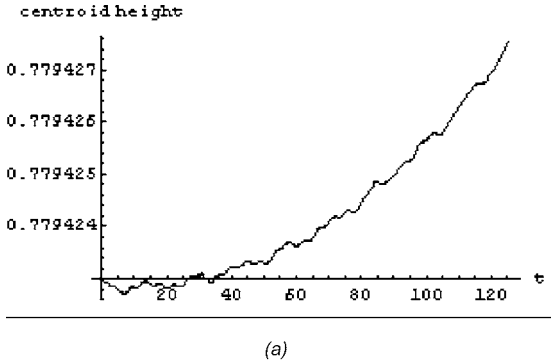
$$z_2'' = -\left(z_2 - \frac{x_1(x_1z_2 - x_2z_1) + y_1(y_1z_2 - y_2z_1)}{\Gamma^2}\right) - 4\tilde{\nu}y_2, \tag{101}$$

$$\tilde{y}_0^2 = y_1^2 + y_2^2, \tag{102}$$

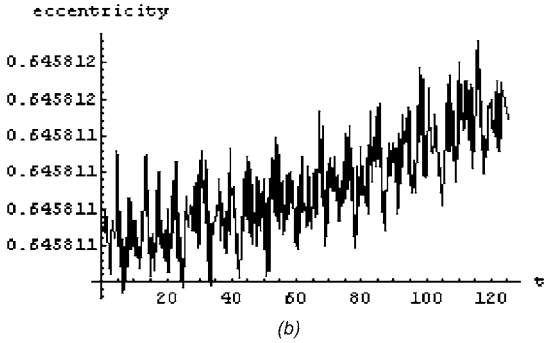
$$\tilde{x}_0'y_0 = x_1'y_1 + x_2'y_2, \tag{103}$$

$$\tilde{z}_0'y_0 = z_1'y_1 + z_2'y_2, \tag{104}$$

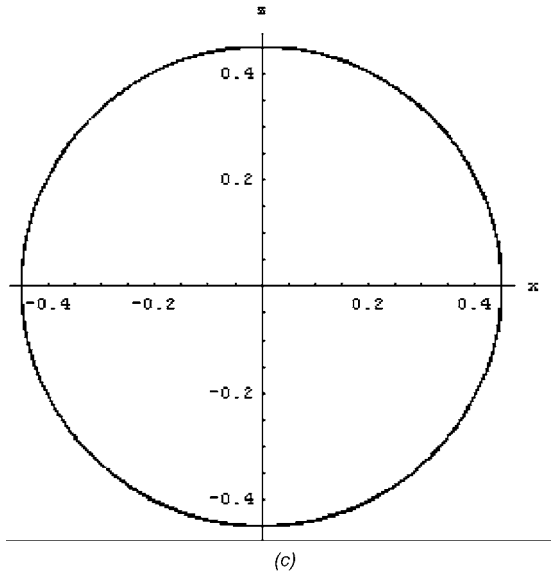
where



(a)



(b)



(c)

Fig. 2 Centroid height (*a*), eccentricity (*b*) and trajectory of point of contact (*c*), for $\tilde{g}$=0.05008

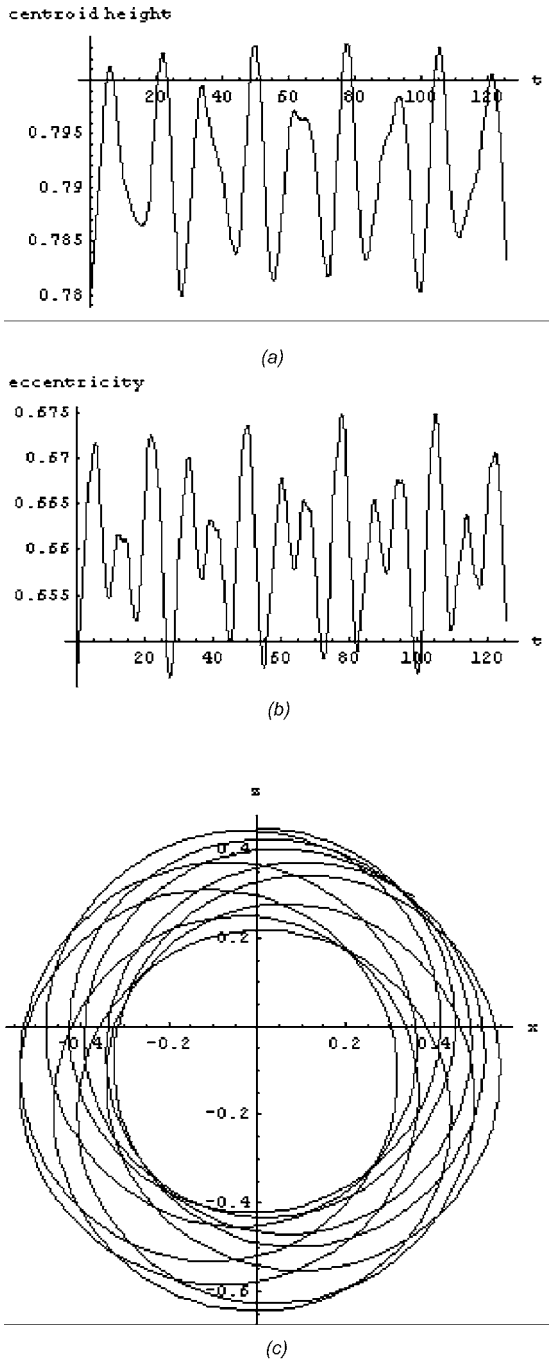


Fig. 3 Centroid height (*a*), eccentricity (*b*) and trajectory of point of contact (*c*), for $\tilde{g}=0.045$

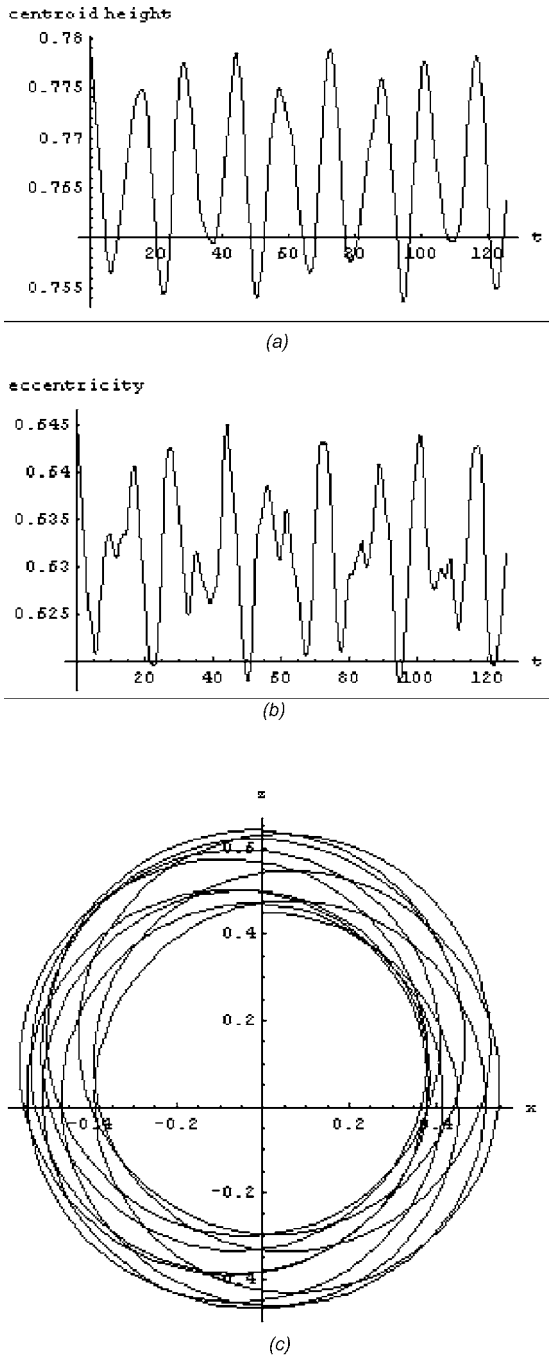


Fig. 4 Centroid height (*a*), eccentricity (*b*) and trajectory of point of contact (*c*), for $\tilde{g}=0.055$

$$\tilde{\lambda}=\lambda/k, \quad \tilde{\mu}=\mu/k, \quad \tilde{\nu}=\nu/k, \tag{105}$$

and

$$\tilde{g}=\frac{g}{kR}. \tag{106}$$

This last nondimensional parameter, $\tilde{g}$, remains as the only vestige of the actual properties of the coin. It combines the acceleration of gravity g, with the size R and the stiffness-density ratio k. In other words, examples must differ either in the value of $\tilde{g}$, or in the initial conditions, or in both.

4.2 The Limit $\tilde{g}\rightarrow 0$. For a given radius, the surviving physically meaningful nondimensional parameter $\tilde{g}$ may tend to zero for two reasons: either the gravity field disappears or the stiffness of the material tends to infinity. Although the effect on the solution of the system is identical for both cases, the interpretation of the results is not. In the first case ($\tilde{g}=0$) we obtain (as we have done in several examples already) the actual physical situation of lack of gravity. The general solutions obtained will, therefore, provide a meaningful result under those conditions. For example, in the case of pivoting, Eq. (68) reduces to

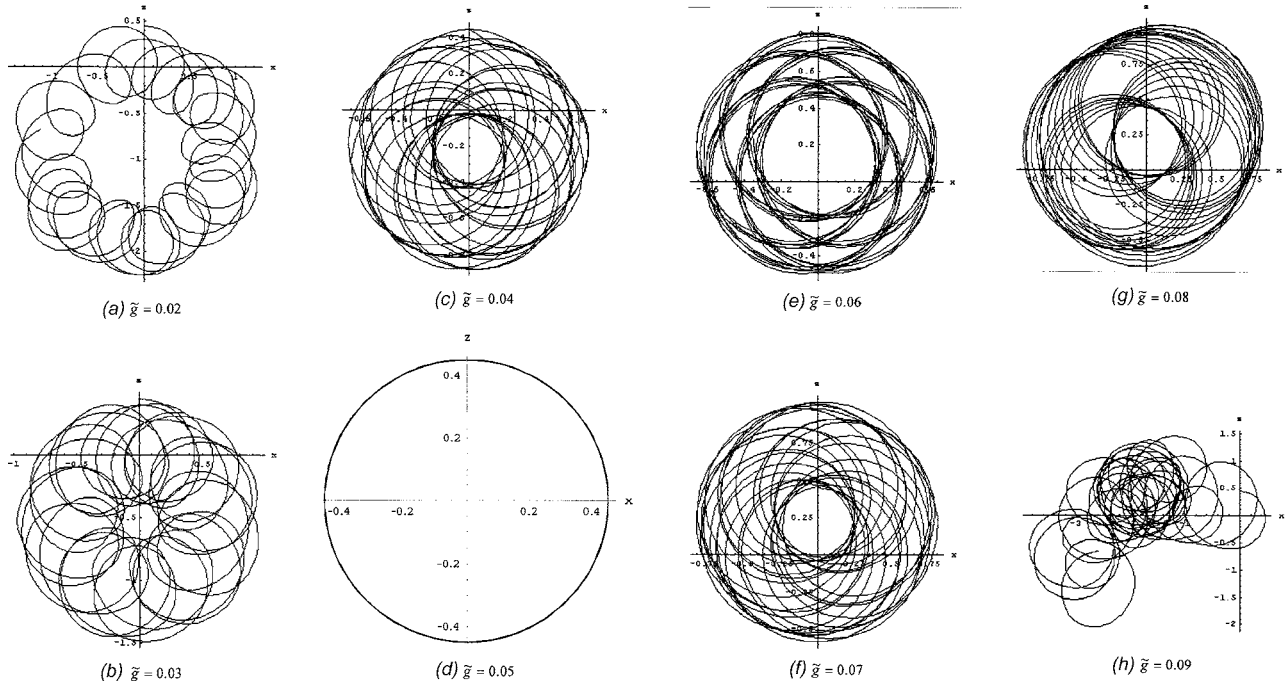


Fig. 5 Trajectory of point of contact for various values of $\tilde{g}$

$$a^2 = c^4, \tag{107}$$

which, after substitution into Eq. (67), yields the following relation between the (nondimensional) angular velocity and the (nondimensional) minor semiaxis of the ellipse:

$$\tilde{\omega}^2 = 1 - c^6, \tag{108}$$

a perfectly legitimate result. On the other hand, if k were to become very large, we would have to be more careful in interpreting this result. Indeed, in trying to redimensionalize the angular velocity [namely, $\omega^2 = k\tilde{\omega}^2 = k(1-c^6)$], we would conclude that the right-hand side of Eq. (108) would have to vanish, so that $c=1$, as it should in a rigid-body deformation. The velocity of pivoting then becomes indeterminate, as it should, since a rigid coin can be made to pivot at any speed. Similar arguments apply to all other instances of passing to the limit: the vanishing of the gravity g causes no problem of interpretation, but the unboundedness of k, although having the same mathematical effect, does.

4.3 Example. According to Eqs. (87)–(89), the wobbling of a pseudo-rigid dish right above the origin of coordinates may take place at constant shape if the deformed semiaxes, the angle of tilt, the angular speeds of "precession" and "spin" and the stiffness ratio are properly proportioned. On nondimensionalization, however, the stiffness ratio is completely absorbed into the angular speeds. Always using the convention that tildes denote nondimensionalized quantities, Eq. (89) becomes

$$\frac{\dfrac{1}{a^4c^2} - 1}{\dfrac{c^2}{a^2}\sin^2\theta - 1} = \tilde{\alpha}^2, \tag{109}$$

while Eq. (88) remains unchanged, since it is already nondimensional. Setting, for the sake of a numerical experiment, $c=0.9$ and the tilt $\theta=30$, we obtain $\tilde{\alpha}=0.506599$, $\tilde{\beta}=-0.203557$, and $a=1.119928$. Finally, Eq. (84) yields $\tilde{\lambda}=0.064161$. As pointed out after Eq. (89), this value of the Lagrange multiplier can be generated automatically by the weight of the dish provided the parameter $\tilde{g}$ is set to the value: $\tilde{g}=\tilde{\lambda}c\cos\theta=0.050008$. A consistent set of initial conditions (corresponding to starting the motion when the point of contact lies on the z-axis) is $\tilde{x}_0=0$, $\tilde{y}_0=c\cos\theta=0.779423$, $\tilde{z}_0=0$, $x_1=a=1.119928$, $x_2=0$, $y_1=0$, $y_2=c\cos\theta=0.779423$, $z_1=0$, $z_2=-c\sin\theta=-0.45$, $x_1'=0$, $x_2'=-a\tilde{\beta}-c\tilde{\alpha}\sin\theta=0$, $y_1'=c\tilde{\beta}\cos\theta=-0.158657$, $y_2'=0$, $z_1'=-a\tilde{\alpha}-c\tilde{\beta}\sin\theta=-0.475753$, $z_2'=0$. The equations of motion have been programmed in Mathematica^R. Figure 2 shows the numerical solution for the height of the center of mass y_0 [Fig. 2(*a*)], the ratio between the semiaxes of the deformed dish [Fig. 2(*b*)], and the trace of the point of contact on the horizontal plane. As expected, and with a great degree of accuracy, it is observed that after 4 complete revolutions ($\tilde{t}=40\pi$) the height of the center of mass is still at its initial value (to 5 significant digits), the ellipse eccentricity has remained constant (to the same accuracy), and the trace of the point of contact is a perfect circle. The next two figures (Figs. 3 and 4) show the same quantities when the parameter $\tilde{g}$ is first reduced (to 0.045) and then increased (to 0.055) without changing the initial conditions. The resulting incompatibility with our conditions for constant shape is reflected in corresponding oscillations of the height of the center of mass and the eccentricity, while the point of contact exhibits a more erratic behavior. This behavior seems, however, to abide by the regularity of being confined between two concentric circles. In Fig. 5 we show the trajectory for a wider range of values of $\tilde{g}$ and for a longer period of time ($\tilde{t}=100\pi$). For values of $\tilde{g}$ higher than 0.09, the behavior at large times seems to become chaotic. It is clear that a detailed analytical study, beyond the scope of this paper, will serve to provide a predictive scheme and to define zones of instability.

Acknowledgment

This work has been partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Slavianowski, J. J., 1974, "Analytical Mechanics of Finite Homogeneous Strains," Arch. Mech., **26**, pp. 569–587.
[2] Slavianowski, J. J., 1975, "Newtonian Mechanics of Homogeneous Strains," Arch. Mech., **27**, pp. 93–102.

[3] Cohen, H., and Muncaster, R. G., 1988, *The Theory of Pseudo-Rigid Bodies*, Springer-Verlag, Berlin.
[4] Cushman, R., Hermans, J., and Kemppainnen, D., 1996, "The Rolling Disc," *Progr. Nonl. Diff. Eqs. Appl.*, H. W. Broer, S. A. van Gils, I. Hoveijn, and F. Takens, eds., Birkhauser, Boston, **19**, pp. 21–60.
[5] O'Reilly, O. M., 1996, "The Dynamics of Rolling Disks and Sliding Disks," Nonlinear Dyn., **10**, pp. 287–305.
[6] Moffatt, H. K., 2000, "Euler's Disk and its Finite-time Singularity," Nature (London), **404**, pp. 833–834.
[7] Kessler, P., and O'Reilly, O. M., 2002, "The Ringing of Euler's Disk," Regular Chaotic Dyn., **7**, pp. 49–60.
[8] Rubin, M. B., 1985, "On the Theory of a Cosserat Point and its Application to the Numerical Solution of Continuum Problems," J. Appl. Mech., **52**, pp. 368–372.
[9] Solberg, J. M., and Papadopoulos, P., 1999, "A Simple Finite-Element based Framework for the Analysis of Elastic Pseudo-Rigid Bodies," Int. J. Numer. Methods Eng., **45**, pp. 1297–1314.
[10] Varadi, P. C., Lo, G. J., O'Reilly, O. M., and Papadopoulos, P., 1999, "A Novel Approach to Vehicle Dynamics using the Theory of a Cosserat Point and its Application to Collision Analyses of Platooning Vehicles," Veh. Syst. Dyn., **32**, pp. 85–108.
[11] Solberg, J. M., and Papadopoulos, P., 2000, "Impact of an Elastic Pseudo-Rigid Body on a Rigid Foundation," Int. J. Eng. Sci., **38**, pp. 589–603.
[12] Papadopoulos, P., 2001, "On a Class of Higher-Order Pseudo-Rigid Bodies," Math. Mech. Solids, **6**, pp. 631–640.
[13] Casey, J., 2004, "Pseudo-Rigid Continua: Basic Theory and a Geometrical Derivation of Lagrange's Equations," Proc. R. Soc. London, Ser. A, **460**, pp. 2021–2049.
[14] Cohen, H., and Sun, Q. X., 1990, "Rocking, Rolling and Oscillation of Elastic Pseudo-Rigid Membranes," J. Sound Vib., **143**, pp. 423–441.

Daniel J. Segalman
Fellow ASME
e-mail: djsegal@sandia.gov

Michael J. Starr

Martin W. Heinstein

Sandia National Laboratories,
P.O. Box 5800,
Albuquerque, NM 87185-0847

New Approximations for Elastic Spheres Under an Oscillating Torsional Couple

The Lubkin solution for two spheres pressed together and then subjected to a monotonically increasing axial couple is examined numerically. The Deresiewicz asymptotic solution is compared to the full solution and its utility is evaluated. Alternative approximations for the Lubkin solution are suggested and compared. One approximation is a Padé rational function which matches the analytic solution over all rotations. The other is an exponential approximation that reproduces the asymptotic values of the analytic solution at infinitesimal and infinite rotations. Finally, finite element solutions for the Lubkin problem are compared with the exact and approximate solutions.
[DOI: 10.1115/1.1985430]

1 Introduction

Lubkin [1] considered two spheres pressed together and then subjected to a monotonically increasing couple aligned with the force of contact (see Fig. 1). He derived analytic expressions for the applied couple M in terms of $k'=c/a$, where a is the radius of the contact patch and c is the radius of the region that remains stuck at torque amplitude M. He also derived a corresponding expression for the relative rotation β associated with that value of radius ratio k'. (Actually, Lubkin defined the relative rotation to be one half of the difference of the absolute axial rotations of the spheres and that convention is continued here.) Both of these expressions involved complete elliptic integrals and other singular integrals, but with the aid of tables of special functions he was able to obtain plots of torque versus angle (M versus β).

Deresiewicz [2] later looked at this problem and examined the same geometry where the torque amplitude is oscillated in a positive and negative sense. In this work Deresiewicz employed Taylor series for $M(k')$ and $\beta(k')$, retaining only two or three terms in those series. He inverted the first series to obtain k' in terms of M and substituted the result into the second series to obtain β in terms of M. He similarly obtained expressions for the case of reversal of applied torque and finally obtained his well-known expression for energy dissipation as a function of torque.

Unlike the analogous derivation for energy dissipation under oscillatory tangential forces (Mindlin et al. [3]), the Deresiewicz expressions were derived from ruthlessly truncated Taylor series and are valid only for cases of very small torque magnitude. Because the rotational motion of contacting surfaces is a significant source of fretting fatigue (Chateauminois and Briscoe [4] and Briscoe et al. [5]) and also because the dissipation of such motion is often a source of vibration damping or of whirl in rotational machinery, it is important to obtain better approximate models of this process.

The work presented here compares the Deresiewicz equations for torque versus angle and for dissipation versus torque with numerical solutions of the original integral equations to define the regions of validity of those approximations and to present some other approximate but more globally applicable expressions for the quantities of interest. In the process, we discuss the Masing [6] properties of the rotating sphere problem and demonstrate how to calculate the energy dissipation from the torque curve associated with a monotonically applied couple. Additionally, finite element analysis is performed to support the new relations introduced here.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, July 23, 2004; final revision, December 2, 2004. Associate Editor: S. Mukherjee. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Monotonic Rotation

Lubkin derived the following equation (modulo the normalization) for the dimensionless torque as a function of k' (or its trigonometric complement $k=\sqrt{1-k'^2}$)

$$T(k)=\frac{1}{4\pi}\left\{\frac{3\pi^2}{4}+k'k^2[6K(k)+(4k'^2-3)D(k)]-3kK(k)\arcsin(k')-3k^2\left[K(k)\int_0^{\pi/2}\frac{\arcsin(k'\sin(\alpha))d\alpha}{(1-k'^2\sin^2(\alpha))^{3/2}}-D(k)\int_0^{\pi/2}\frac{\arcsin(k'\sin(\alpha))d\alpha}{(1-k'^2\sin^2(\alpha))^{1/2}}\right]\right\} \quad (1)$$

where $T=M/(\mu Na)$, K and E are complete elliptic integrals of the first and second kind, $D(k)=(K(k)-E(k))/k^2$, a is the radius of the contact patch, μ is the coefficient of friction between the spheres, and N is the force pushing the spheres together.

Lubkin also derived the following equation (again modulo a normalization) for the relative angle of twist between each sphere and the plane of contact:

$$\theta(k)=\frac{3}{4\pi}k^2D(k) \quad (2)$$

where $\theta=\beta Ga^2/(\mu N)$ and G is the shear modulus.

In presenting his version of Eq. (1), Deresiewicz points out that he is correcting a typographic error of Lubkin's. The complex form of the above equation has subsequently resulted in other typographic errors occurring in the literature. The problem is exacerbated by the inconsistent definitions for the elliptic integrals to be found in the literature. (Two well-known, but inconsistent sources are Lebedev [7] and Abramowitz and Stegun [8].)

Deresiewicz reformulated Eq. (1) and expanded it and Eq. (2) as Taylor series in k, retaining two terms in each series. He employed these truncated series to solve for relative rotation in terms of moment. When that equation is inverted, it takes the form

 Copyright © 2005 by ASME

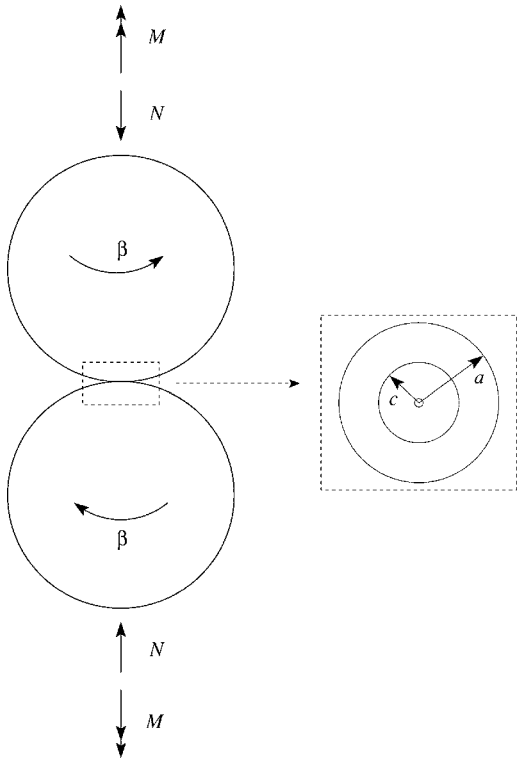


Fig. 1 Lubkin considered two spheres pressed together by forces *N* and then subjected to a monotonically applied torsion *M*. The resulting relative rotation (one half the difference in absolute rotation) is indicated by β. Indicated in the inset is the radius *a* of the contact patch and the radius *c* of the stuck region.

$$T = \tfrac{8}{3}\left[\sqrt{1+8\ \theta} - (1+2\theta)\right] \tag{3}$$

for monotonically applied twist.

Both Lubkin and Deresiewicz observe that the true moment-rotation curve should start from the origin with an initial slope of 16/3 and increase monotonically to an asymptotic value of $3\pi/16$. Deresiewicz demonstrates that his approximation satisfies the first of these conditions. A numerical evaluation of Eqs. (1) and (2) is compared to the approximation of Deresiewicz in Fig. 2. In that figure, the Deresiewicz approximation matches the numerical evaluation of the exact equation in only a very small zone near the origin and differs from it substantially away from the origin. In fact, the approximation is not even globally monotonic. It peaks above the analytic maximum value of $3\pi/16$. Further, examination of Eq. (3) shows that the torque of the approximation becomes negative as θ becomes large.

A far better approximation for the torque-rotation relation is obtained by fitting a fourth-order Padé rational function approximation to the results of numerical integration:

$$T_P(\theta) = \frac{a_0 + a_1\theta + a_2\theta^2 + a_3\theta^3 + a_4\theta^4}{b_0 + b_1\theta + b_2\theta^2 + b_3\theta^3 + b_4\theta^4} \tag{4}$$

where

$$\begin{aligned} a_0 &= 0 & b_0 &= 1 \\ a_1 &= 16/3 & b_1 &= 5.1193 \\ a_2 &= 6.0327 & b_2 &= 15.6833 \\ a_3 &= 19.6951 & b_3 &= 30.8099 \\ a_4 &= 42.5359 & b_4 &= 72.2111 \end{aligned} \tag{5}$$

This approximation agrees with the numerical evaluation of Eq. (1) to within $2\text{x}10^{-5}$ over the whole range ($0 < \theta < \infty$).

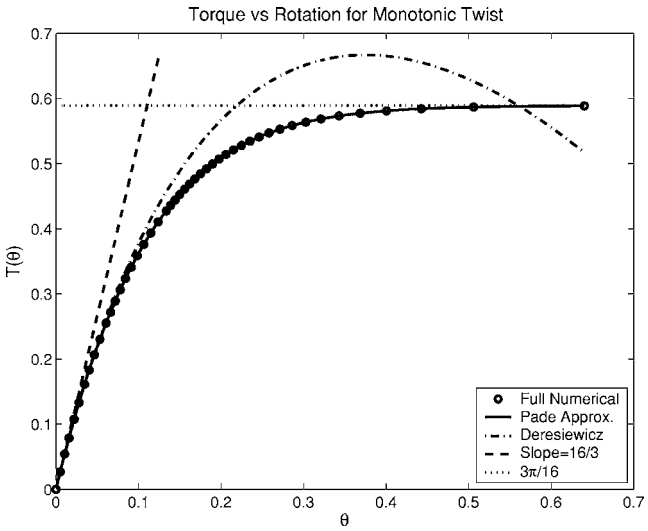


Fig. 2 The Deresiewicz approximation for $T(\theta)$ for monotonic rotation appears to be adequate only for the very small values of rotation. A rational function (Padé) approximation overlies the values calculated from Lubkin's integral equation and appears to be adequate over the whole range of values computed.

An alternative approximation is that of Cuttino and Dow [10] based on the results of the finite element analysis of relative rotation of contacting ellipsoids. For the case of circular contact treated here, their expression reduces to the following

$$T(\theta) = 0.62[1 - \exp(-10\ \theta)] \tag{6}$$

which has an incorrect limiting value of 0.62, about 5% higher than the analytic value. The finite element mesh used in the above calculation was very coarse by today's standards, but very respectable for the time.

A natural modification of Cuttino and Dow's approximation is the following, constructed to reproduce the Lubkin result at zero and infinity:

$$T(\theta) = (3\pi/16)\left[1 - \exp\left(-\left(\frac{\theta}{\pi}\right)\left(\frac{16}{3}\right)^2\right)\right] \tag{7}$$

The above approximation along with that of Cuttino and Dow are plotted in Fig. 3. Equation (7) is a substantially better approximation to the Lubkin results, though it does have some small visible error in the region $0.1 < \theta < 0.5$.

3 Oscillatory Rotation

The above observations about the limitations of Deresiewicz's approximation for monotonic rotation suggest a re-examination of the Deresiewicz approximation for energy dissipation in oscillatory rotation, which is similarly based on a severely truncated Taylor series.

Deresiewicz's approximate equation for dimensionless energy dissipation per cycle under oscillatory dimensionless torque of amplitude T is

$$\mathcal{D}(T) = 2\left\{\tfrac{8}{9}\left[1 - \left(1 - \tfrac{3}{2}T\right)^{3/2}\right] - T\left[1 - \left(1 - \tfrac{3}{2}T\right)^{1/2}\right]\right\} \tag{8}$$

where $\mathcal{D}$ is the energy dissipation normalized by $(\mu N)^2/(\text{Ga})$. Deresiewicz points out that the above expression behaves as $\mathcal{D}(T) = \left(\frac{3}{16}\right)T^3$ for small torques T.

We now introduce another approach to evaluating the energy dissipation associated with oscillatory torque. This approach exploits the observation that the torque/rotation hysteresis curves for oscillations between $-\theta_m$ and θ_m are consistent with the Masing [6] conditions:

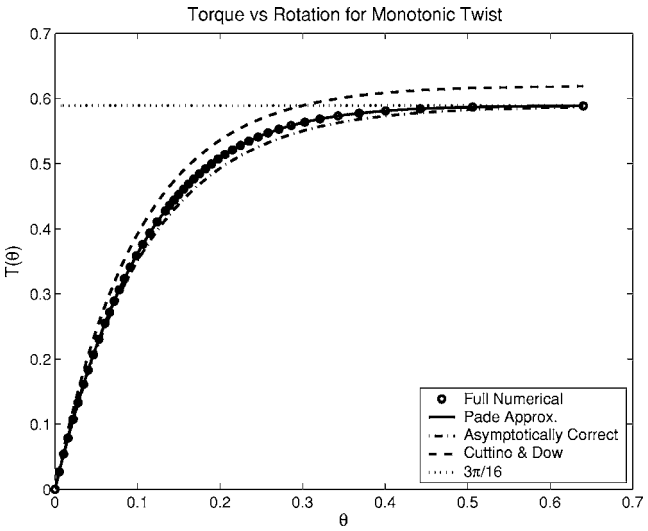


Fig. 3 The Cuttino and Dow approximation has the correct general form, but approaches incorrect asymptotic values. Another approximation of the same general form, but constructed to have the correct asymptotic values is a substantially better approximation though it does have some small visible error in the region $0.1<\theta<0.5$.

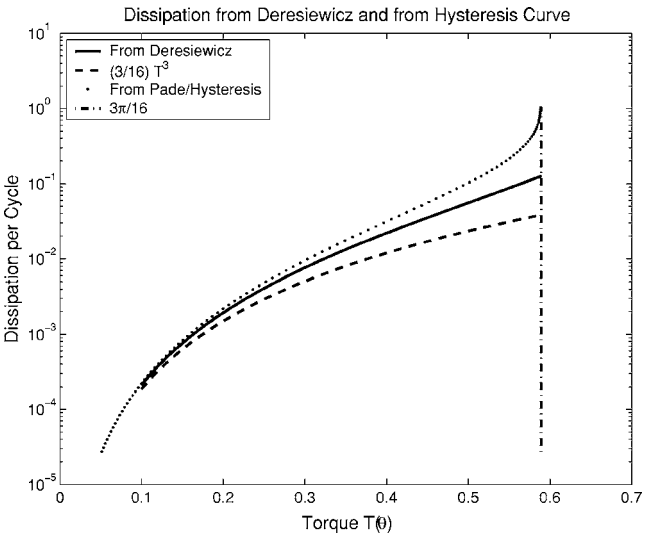


Fig. 5 The Deresiewicz approximation for dissipation per cycle appears to be adequate for only small values of applied torque. The dissipation at large torques calculated from hysteresis appears to go to infinity.

$$T_F(\theta,\theta_m) = 2T_0\left(\frac{\theta_m+\theta}{2}\right) - T_0(\theta_m) \tag{9}$$

$$T_B(\theta,\theta_m) = -2T_0\left(\frac{\theta_m-\theta}{2}\right) + T_0(\theta_m) \tag{10}$$

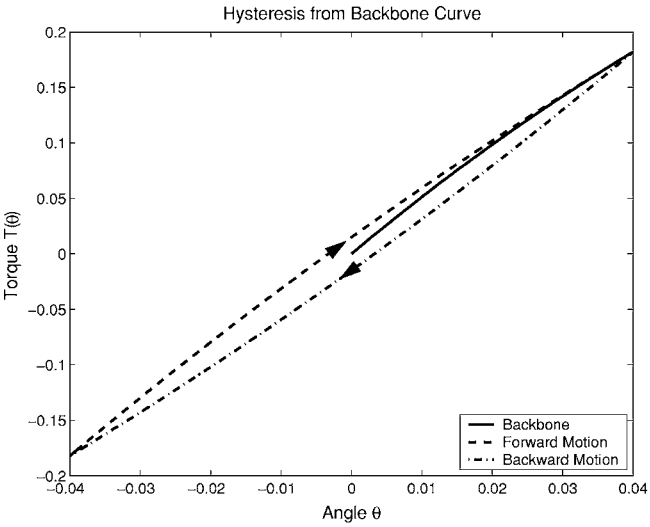
where T_F is the torque on the forward part of the hysteresis loop, T_B is the torque on the backward part of the hysteresis loop, and $T_0(\theta)$ is the torque curve associated with monotonic rotation (often called the "backbone" curve.) It was Deresiewicz who observed these symmetries between forward and back motion, though he did not connect these symmetries to Masing models. These relations are illustrated in Fig. 4 where the Padé approximation to the backbone curve is used to draw the hysteresis plots for rotation between $\theta=-0.04$ and $\theta=0.04$ and for the rotation between $\theta=-0.5$ and $\theta=0.5$.

The energy dissipation is the area enclosed by the T_F and T_B curves and with reference to Eqs. (9) and (10) this quantity is

$$\mathcal{D}(T(\theta)) = 8\left[\int_0^\theta T(\tau)d\tau - \tfrac{1}{2}\theta\, T(\theta)\right] \tag{11}$$

The Deresiewicz approximation for energy dissipation per cycle is compared in Fig. 5 with that obtained by the numerical evaluation of Eq. (11) where $T(\theta)$ is obtained from the Padé approximation. We see that the Deresiewicz approximation for dissipation per cycle goes as $\left(\frac{3}{16}\right)T^3$ for small values of T, as does the exact solution (Lubkin [1]). However, the Deresiewicz approximation remains bounded as torque T approaches its limiting value of $3\pi/16$. The numerically obtained dissipation values appear to grow to infinity as T approaches its limit.

An observation can be made about the nature of the singularity one may expect in the dissipation per cycle as T approaches its limiting value of $3\pi/16$. From Fig. 6, we see that the difference of



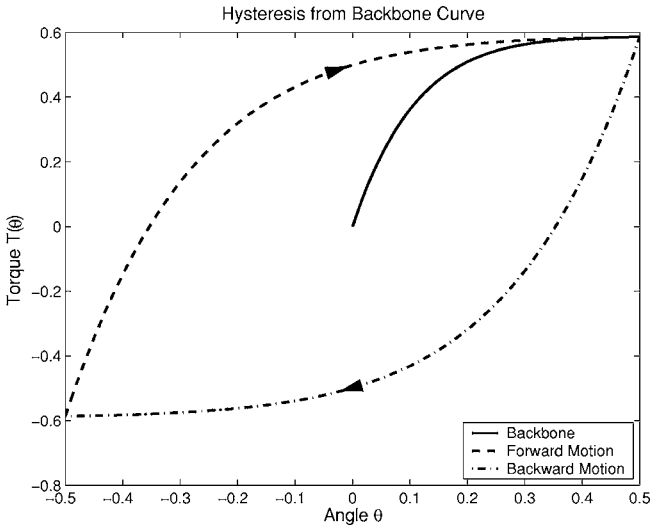


Fig. 4 The Padé approximation for the numerically obtained backbone curve is used to draw the hysteresis plots for $\theta=0.04$ and $\theta=0.5$. At large rotation angles, the slope of the hysteresis curve just before reversal becomes very small.

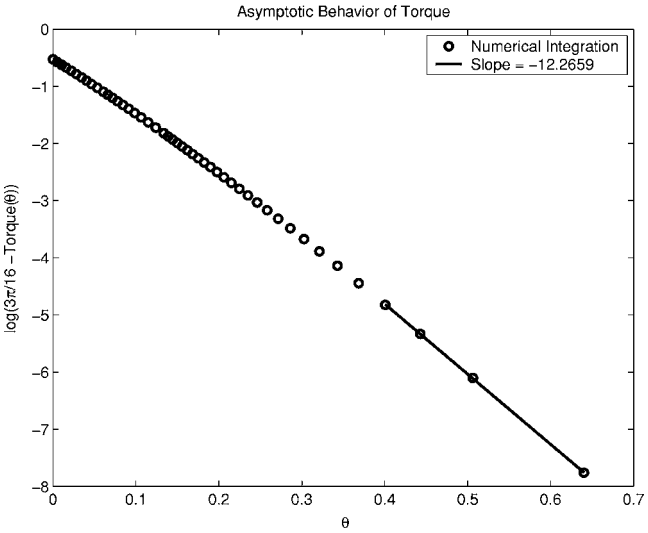


Fig. 6 The asymptotic behavior of the torque associated with large angles is shown in the plot of $\log(3\pi/16 - T(\theta))$, suggesting that the difference of the torque and its limiting value goes as $\theta^{-12.3}$. Note that the Padé approximation overlies the data points obtained through numerical evaluation of Lubkin's equation.

the torque and its limiting value appears to go as $\theta^{-12.3}$ for $\theta > 0.5$, suggesting the following asymptotic form for rotation angle in terms of torque for large rotations:

$$\theta - \theta_m = \frac{1}{A}\log\left(\frac{3\pi/16 - T(\theta)}{3\pi/16 - T(\theta_m)}\right) \tag{12}$$

where A is the slope of the asymptotic plot in Fig. 6 (approximately -12.3), θ_m a value greater than 0.5, and $T(\theta_m)$ is the torque corresponding to θ_m on the monotonic torque-rotation curve (Fig. 2).

For oscillatory rotation at similarly large angles of twist, the slopes of the hysteresis curves just before reversal (at the corners) are near zero (see the right hand plot in Fig. 4). Letting $\mathcal{D}(\theta_m)$ be the energy dissipation per cycle corresponding to θ_m, the dissipation per cycle at some larger rotation will be approximately that due to the additional rotation acting through the peak torque. Therefore,

$$\mathcal{D}(\theta) \approx \mathcal{D}(\theta_m) + 2\frac{3\pi}{8}(\theta - \theta_m) \tag{13}$$

as indicated in Fig. 7. This is the same approximate dissipation that would be obtained by expanding Eq. (11) about $(\theta_m, D(\theta_m))$ and noting that $dT(\theta)/d\theta|_{\theta_m}$ is approximately zero.

Substituting the above into Eq. (12), we obtain the following

$$\mathcal{D}(\theta) \approx \mathcal{D}(\theta_m) + 2\frac{3\pi}{8}\frac{1}{A}\log\left(\frac{3\pi/16 - T(\theta)}{3\pi/16 - T(\theta_m)}\right) \tag{14}$$

indicating a logarithmic singularity in the dissipation per cycle as a function of applied torque.

Having explored the anticipated asymptotic form of the dissipation as a function of torque amplitude, we observe that the dissipation of Eq. (11) can be evaluated in closed form for the case of the asymptotically correct approximation embodied in Eq. (7). Noting also that Eq. (7) can be inverted for θ in terms of T, we have

$$\mathcal{D}(T) = -4\pi\left(\frac{3}{16}\right)^2\left\{2T + \left[2\left(\frac{3\pi}{16}\right) - T\right]\ln\left[\frac{(3\pi/16) - T}{(3\pi/16)}\right]\right\} \tag{15}$$

which indeed contains the expected logarithmic singularity.

4 Finite Element Analysis

Since the unbounded levels of energy dissipation at high levels of oscillatory torque (Eq. (14)) are a new result, it would be comforting to support this observation by independent means. Finite element analysis provides that independent means.

An additional reason to submit the problem of elastic spheres under oscillatory torsional couple to finite element analysis is that this problem is a very sensitive measure of the accuracy of such analysis. Though the previously mentioned finite element study (Cuttino and Dow) employing a commercial finite element code, was heroic for the time, their analyses were insufficient to reproduce the Lubkin results to within 5%. Such errors are likely attributable to discretization error (overly coarse meshes), though the meshes they used reflected the state of code and computers available at that time.

The above considerations motivated numerical re-examination of the Lubkin problem via finite element analyses with much finer meshes. The finite element code JAS, developed at Sandia National Laboratories, was employed with the more appropriate hemispherical mesh shown in Fig. 8. This mesh employs approximately 67,000 8-node hex elements and 70,000 nodes. The element size is particularly small in the region of the contact patch where slip is anticipated. There the radial spacing of the elements is approximately 0.05. (The hemisphere radius is exactly 1.) Each element spans approximately 7.5 deg circumferentially. This mesh contrasts with that of Cuttino and Dow, which had fewer than 2500 elements.

Figure 9 shows the ramp-up torque (backbone curve) and the hysteresis curve associated with the normalized rotation angle $\theta = 0.3$. Also shown is the hysteresis plot that would be obtained by applying Eqs. (9) and (10) to the ramp-up torque. We see that the hysteresis loop predicted by the finite element analysis is nicely consistent with the Masing conditions.

The ramp-up torque calculated in the finite element analysis is plotted in Fig. 10 along with the Padé approximation to the nu-

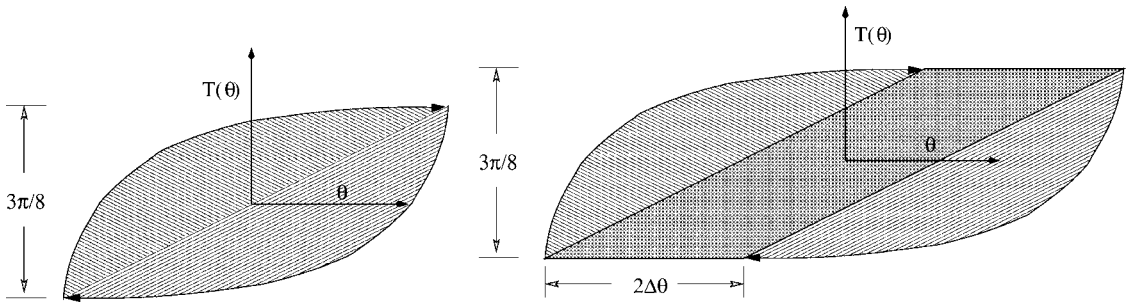


Fig. 7 Once the peaks of the hysteresis curve approach the limiting torque, additional rotations result in additional dissipation approximately equal to the area of the rhombus shown

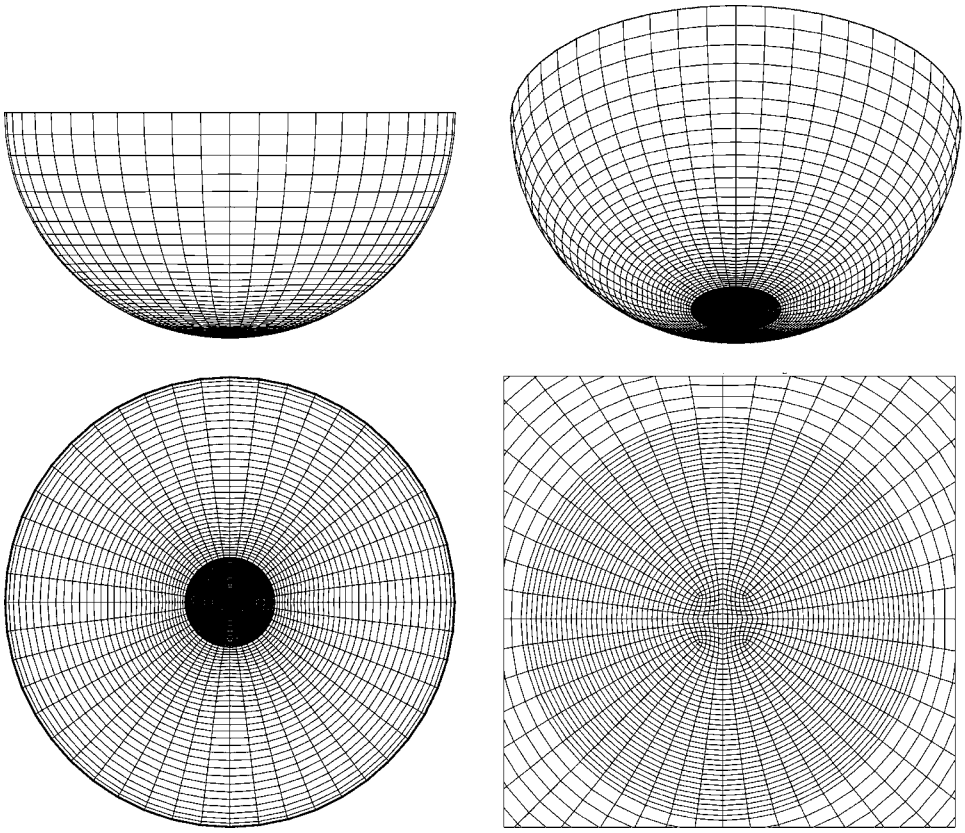


Fig. 8 A very fine mesh is employed to capture the Lubkin result with good precision. This mesh employs approximately 67 000 8-node hex elements and 70 000 nodes.

merical solution to Eqs. (1) and (2). We see that though there is some visible difference between the finite element predictions and the Padé approximation for $0.1<\theta<0.3$, there is overall reasonably good agreement, giving us confidence in the finite element results. This confidence permits us to trust the dissipation that is also calculated in the finite element analysis.

The dissipation per cycle calculated from the finite element analysis is shown in Fig. 11, along with the dissipation estimates calculated by other means. That finite element dissipation is shown to be neatly bounded by the predictions of the hysteresis calculations based on the Padé approximation and by the dissipation of Eq. (15). (Recall that Eq. (15) was derived from the asymptotically correct ramp-up approximation of Eq. (7).) Further, the finite element results are consistent with the logarithmically singular behavior of Eq. (14).

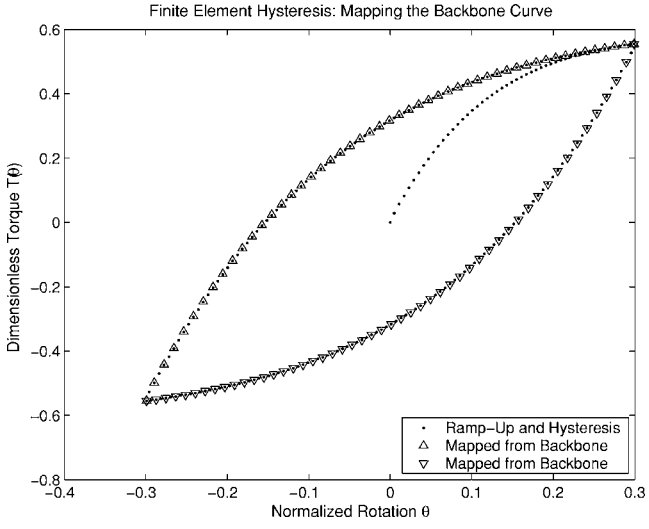


Fig. 9 The hysteresis curve calculated by the finite element code JAS for a normalized angle of $\theta=0.3$. This hysteresis curve manifests the symmetries associated with Masing models.

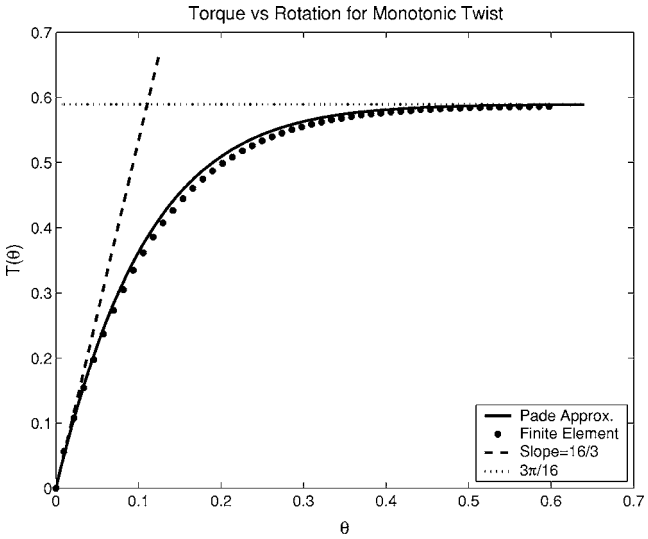


Fig. 10 The ramp-up torque calculated by the finite element code is reasonably consistent with the numerical evaluation of Eqs. (1) and (2), as represented by the Padé approximation

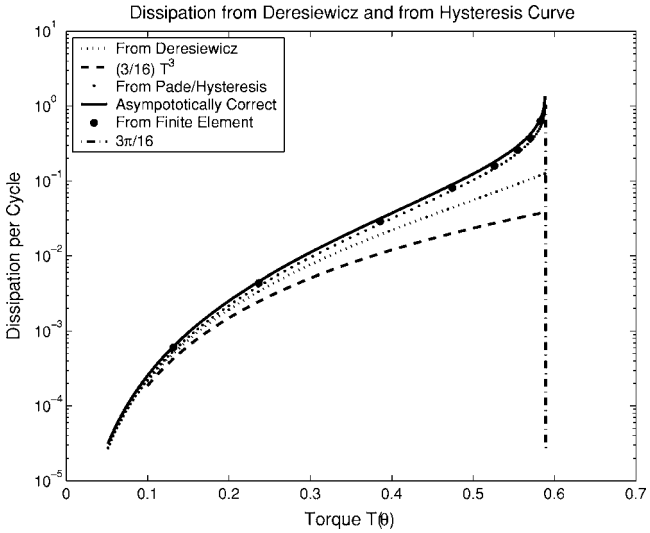


Fig. 11 The dissipation per cycle calculated from the finite element analysis is consistent with the logarithmically singular behavior predicted by Eq. (14)

5 Conclusions

The above analysis illuminates several points.

1. The Deresiewicz approximations for torque and for dissipation are best used only in regimes of very small rotation.
2. The torque-rotation relations of the rotating sphere problem manifest the properties of Masing models.
3. Two new approximations have been introduced. The first (Eq. (4)) is a rational function approximation to the Lubkin results that, though cumbersome, is very nearly exact over the entire range of θ. The second (Eq. (7)) has the same simple form as that of Cuttino and Dow, but was constructed to have the correct asymptotic behavior. This new approximation also has substantially less error overall than does that of Cuttino and Dow.
4. The energy dissipation grows logarithmically as the magnitude of torque amplitude approaches its limiting value. This logarithmic behavior is manifested by:
   (a) The hysteresis calculations based on the Padé approximation for the ramp-up torque;
   (b) the energy dissipation of Eq. (15) derived from the asymptotically correct approximation presented in Eq. (7);
   (c) the finite element results.

Acknowledgment

The authors thank Prof. Anthony Sackfield for going to substantial difficulty to clarify for us the correct derivation of Eq. (1). The authors also thank Prof. Dane Quinn of the University of Akron for introducing them into the world of the Padé approximation.

References

[1] Lubkin, L., 1951, "The Torsion of Elastic Spheres in Contact," ASME J. Appl. Mech. **18**, pp. 183–187.
[2] Deresiewicz, H., 1954 "Contact of Elastic Spheres Under an Oscillating Torsional Couple," ASME J. Appl. Mech. **21**, pp. 52–56.
[3] Mindlin, R. D., Mason, W. P., Osmer, T. F., and Deresiewicz, H., 1952, "Effects of an Oscillating Tangential Force on the Contact Surfaces of Elastic Spheres," *Proc. of 1st US National Congress of Applied Mechanics*, pp. 203–208.
[4] Briscoe, B. J., Chateauminois, A., Lindley, T. C., and Parsonage, D., 1998, "Fretting Wear Behaviour of Polymethylmethacrylate Under Linear Motions and Torsional Contact Conditions," Tribol. Int. **31**, pp. 701–711.
[5] Chateauminois, A., and Briscoe, B. J., 2002, "Measurements of Friction-Induced Surface Strains in a Steel/Polymer Contact," Tribol. Int. **35**, pp. 245–254.
[6] Masing, G., 1926, "Self-Stretching and Hardening in Brass," *Proc. 2nd Int. Cong. Applied Mechanics*, pp. 332-335.
[7] Lebedev, N. N., 1972, *Special Functions and Their Applications*, Translated and edited by R. A. Silverman, Dover Press, NY, p. 188.
[8] Abramowitz, M., and Stegun, I., 1972, *Handbook of Mathematical Functions*, Dover Press, NY, p. 590.
[9] Hills, D. A., and Sackfield, A., 1986, "The Stress Field Induced by a Twisting Sphere," ASME J. Appl. Mech. **53**, pp. 372–378.
[10] Cuttino, J. F., and Dow, T. A., 1997, "Contact Between Elastic Bodies With an Elliptic Contact Interface in Torsion," ASME J. Appl. Mech. **64**, pp. 144–148.

A Stabilized Mixed Finite Element Method for Nearly Incompressible Elasticity

Arif Masud[1]
Associate Professor of Mechanics and Materials
e-mail: amasud@uic.edu

Kaiming Xia[2]
Department of Civil and Materials Engineering,
The University of Illinois at Chicago,
Chicago, IL 60607-7023

We present a new multiscale/stabilized finite element method for compressible and incompressible elasticity. The multiscale method arises from a decomposition of the displacement field into coarse (resolved) and fine (unresolved) scales. The resulting stabilized-mixed form consistently represents the fine computational scales in the solution and thus possesses higher coarse mesh accuracy. The ensuing finite element formulation allows arbitrary combinations of interpolation functions for the displacement and stress fields. Specifically, equal order interpolations that are easy to implement but violate the celebrated Babushka-Brezzi inf-sup condition, become stable and convergent. Since the proposed framework is based on sound variational foundations, it provides a basis for a priori error analysis of the system. Numerical simulations pass various element patch tests and confirm optimal convergence in the norms considered.
[DOI: 10.1115/1.1985433]

1 Introduction

Application of the finite element method to the mixed variational formulations of elasticity has been an area of active research for over two decades. Attention has particularly been focused at incompressible elasticity [1] because of its fundamental place in solid mechanics in general and its ability to model a wide class of materials in particular. Volume preserving or the isochoric mode of deformation is an important kinematic constraint on the response of several materials [2]. Especially, in finite deformation elasto-plasticity, the plastic or the inelastic response of several metals and polymers is assumed volume preserving. Standard displacement based techniques for incompressible elasticity show an overly stiff response commonly termed as "locking" and require special treatment to yield engineering solutions, e.g., the use of mixed-methods [3–7], reduced and selective integration techniques [8–11], stress projection techniques [2,12,13], and B-bar methods [10]. Published literature on the treatment of locking phenomena is exhaustive [2,3,6–19], and the interested reader is referred to the standard texts by Brezzi et al. [20] and Hughes [21] for an overview of the various techniques.

Our objective in this work is to develop a stabilized formulation for incompressible elasticity such that the definition of the stability parameter appears naturally in the developments. The idea of using stabilized methods in computational solid mechanics is motivated by the success of stabilized methods in the area of computational fluid dynamics. Even though the similarity of the mixed $\boldsymbol{u}-p$ form of incompressible elasticity with the Stokes equations was pointed out in Hughes et al. [22], the remark largely went unnoticed and the application of stabilized methods to incompressible elasticity lagged behind its application to fluid dynamics. Relatively recently, attempts have been made to employ Galerkin/least-squares method for elastoplasticity [23], Petrov-Galerkin method for finite elasticity [17], and multiscale method for modeling weak discontinuities in solids [24].

In this paper we employ the variational multiscale method [25] and present a systematic procedure for the development of a stabilized formulation, here on termed as the Hughes variational multiscale (HVM) formulation, for compressible and nearly incompressible elasticity. Within the context of elasticity wherein the underlying continuum formulation does not possess any scales, the word multiscale is to be viewed as the computational scales in the solution to the Galerkin form of the problem. A novelty of the present method as compared to the celebrated stabilized methods, namely the streamline upwind Petrov-Galerkin (SUPG) and the Galerkin/least-squares (GLS) methods is that the present method is free of user defined or user designed stability parameters. In addition, a canonical expression for the stability parameter appears naturally in the developments presented in Sec. 2 and 3. A significant advantage of using stabilized formulations for incompressible elasticity is that the issues related to locking in the incompressible limit are completely avoided, and the need for special treatments, namely, the stress projection techniques, reduced integration techniques, and the use of special interpolation polynomials is completely bypassed. Another significant advantage of the stabilized methods is that arbitrary combinations of interpolation functions can be employed for the displacement and pressure fields. Specially, equal order interpolations that are easy to implement but are unstable within the classical Galerkin framework, become stable and convergent.

An outline of the paper is as follows: Section 2 presents a mixed form of elasticity and its corresponding standard Galerkin form. The multiscale computational framework and the mathematical steps that lead to the stabilized form are presented in Sec. 3. Numerical simulations are presented in Sec. 4, and conclusions are drawn in Sec. 5.

2 A Mixed Form of Elasticity

Let $\Omega \subset \mathbb{R}^{n_{sd}}$ be an open bounded region with piecewise smooth boundary Γ. The number of space dimensions, n_{sd}, is equal to 2 or 3. The unit outward normal vector to Γ is denoted by $\boldsymbol{n} = (n_1, n_2, \ldots n_{\mathrm{sd}})$. The mixed displacement-pressure form of elasticity that is valid for compressible and nearly incompressible behavior [1] is given as follows:

$$\nabla \cdot \boldsymbol{\sigma} + \boldsymbol{b} = \boldsymbol{0} \quad \text{in } \Omega \tag{1}$$

[1]Author to whom correspondence should be addressed.
[2]Formerly Graduate Research Assistant.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, August 9, 2004; final revision, January 11, 2005. Associate Editor: T. E. Tezduyar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

$$\mathbf{1}^T\boldsymbol{\varepsilon} - \frac{p}{K} = 0 \quad \text{in } \Omega \tag{2}$$

$$\boldsymbol{u} = \boldsymbol{g} \quad \text{on } \Gamma_g \tag{3}$$

$$\boldsymbol{\sigma}\cdot\boldsymbol{n} = \boldsymbol{t} \quad \text{on } \Gamma_h \tag{4}$$

where $\boldsymbol{u}(\boldsymbol{x})$ is the displacement vector field and $p(\boldsymbol{x})$ is the scalar pressure field. K is the bulk modulus defined as $K=E/3(1-2\nu)$. Equation (1) represents equilibrium in body Ω, Eq. (2) is the volumetric constitutive equation in Ω, Eq. (3) is the Dirichlet boundary condition on Γ_g, and Eq. (4) is the Neumann boundary condition on Γ_h.

2.1 Standard Variational Form. The standard variational form of the boundary value problem is stated in terms of the following spaces for trial solutions and weighting functions.

$$\mathcal{S} = \{\boldsymbol{u} | \boldsymbol{u} \in H^1(\Omega), \boldsymbol{u} = \boldsymbol{g} \text{ on } \Gamma_g\} \tag{5}$$

$$\mathcal{V} = \{\boldsymbol{w} | \boldsymbol{w} \in H^1(\Omega), \boldsymbol{w} = \boldsymbol{0} \text{ on } \Gamma_g\} \tag{6}$$

$$\mathcal{P} = \{p | p \in L_2(\Omega)\} \tag{7}$$

Because there are no explicit boundary conditions on pressures, $\mathcal{P}$ suffices as both a trial solution space and a weighting function space. The weighted residual form for the equilibrium and the volumetric constitutive equation is as follows:

$$\int_\Omega \nabla \boldsymbol{w} : \boldsymbol{\sigma}\, d\Omega = \int_\Omega \boldsymbol{w}\cdot\boldsymbol{b}\, d\Omega + \int_\Gamma \boldsymbol{w}\cdot\boldsymbol{t}\, d\Gamma \tag{8}$$

$$\int_\Omega q\left(\mathbf{1}^T\boldsymbol{\varepsilon} - \frac{p}{K}\right) d\Omega = 0 \tag{9}$$

where $\boldsymbol{w}$ and q are weighting functions corresponding to displacement $\boldsymbol{u}$ and pressure p, respectively. We split the stress tensor into two parts: Deviatoric stress $\boldsymbol{s}$ and pressure p

$$\begin{aligned}\boldsymbol{\sigma} &= \boldsymbol{s} + p\mathbf{1}\\ &= 2\mu\boldsymbol{I}_{\text{dev}}\boldsymbol{\varepsilon} + p\mathbf{1}\\ &= \boldsymbol{D}^{uu}\boldsymbol{\varepsilon} + \boldsymbol{D}^{up}p\end{aligned} \tag{10}$$

where μ is the shear modulus, $\mathbf{1}$ is the second order unit tensor, $\boldsymbol{I}$ is the rank four unit tensor, and $\boldsymbol{I}_{\text{dev}}=\boldsymbol{I}-\frac{1}{3}\,\mathbf{1}\otimes\mathbf{1}$ is the deviatoric projection of $\boldsymbol{I}$. $\boldsymbol{D}^{uu}=\boldsymbol{D}\left(\boldsymbol{I}-\frac{1}{3}\,\mathbf{1}\otimes\mathbf{1}\right)$ is the deviatoric projection of elastic matrix $\boldsymbol{D}$, and $\boldsymbol{D}^{\text{up}}=\mathbf{1}=[1\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0]^T$.

Remark 1: The standard form (8) and (9) is required to satisfy the celebrated Babushka-Brezzi *inf-sup* condition [20] to yield stable and convergent elements in the incompressible limit.

3 The Variational Multiscale Method

3.1 Multiscale Decomposition. In this section we apply the variational multiscale method, first proposed by Hughes [25], to the development of a stabilized formulation for a mixed form of elasticity that is applicable in the incompressible limit. We consider the bounded domain Ω discretized into nonoverlapping regions Ω^e (element domains) with boundaries Γ^e, $e=1,2,\ldots n_{umel}$, such that

$$\Omega = \overline{\bigcup_{e=1}^{n_{umel}} \Omega^e} \tag{11}$$

We denote the union of element interiors and element boundaries by Ω' and Γ', respectively.

$$\Omega' = \bigcup_{e=1}^{n_{umel}} (\text{int})\Omega^e \quad (\text{elem. interiors}) \tag{12}$$

$$\Gamma' = \bigcup_{e=1}^{n_{umel}} \Gamma^e \quad (\text{elem. boundaries}) \tag{13}$$

We assume an overlapping sum decomposition of the displacement field into coarse scales or resolvable scales and fine scales or the subgrid scales.

$$\boldsymbol{u}(\boldsymbol{x}) = \underbrace{\overline{\boldsymbol{u}}(\boldsymbol{x})}_{\text{coarse scale}} + \underbrace{\boldsymbol{u}'(\boldsymbol{x})}_{\text{fine scale}} \tag{14}$$

Remark 2: For the case of elasticity wherein the underlying continuum formulation does not possess any scales, the coarse and fine scales should be viewed as the computational scales in the solution to the Galerkin form of (8) and (9), i.e., $\overline{\boldsymbol{u}}(\boldsymbol{x})$ represents part of the solution that is resolved by a given grid, and $\boldsymbol{u}'(\boldsymbol{x})$ represents the error in the solution.

Likewise, we assume an overlapping sum decomposition of the weighting function into coarse and fine scale components indicated as $\overline{\boldsymbol{w}}$ and $\boldsymbol{w}'$, respectively.

$$\boldsymbol{w}(\boldsymbol{x}) = \underbrace{\overline{\boldsymbol{w}}(\boldsymbol{x})}_{\text{coarse scale}} + \underbrace{\boldsymbol{w}'(\boldsymbol{x})}_{\text{fine scale}} \tag{15}$$

We further make an assumption that the subgrid scales, although nonzero within the elements, vanish identically over the element boundaries.

$$\boldsymbol{u}' = \boldsymbol{w}' = \boldsymbol{0} \quad \text{on } \Gamma' \tag{16}$$

Employing (14) in the definition of the symmetric strain tensor, we can decompose the strain field into two components.

$$\begin{aligned}\boldsymbol{\varepsilon} &= \tfrac{1}{2}(\nabla\boldsymbol{u} + \nabla^T\boldsymbol{u})\\ &= \tfrac{1}{2}(\nabla\overline{\boldsymbol{u}} + \nabla^T\overline{\boldsymbol{u}}) + \tfrac{1}{2}(\nabla u' + \nabla^T u')\\ &= \overline{\boldsymbol{\varepsilon}} + \boldsymbol{\varepsilon}'\end{aligned} \tag{17}$$

where $\overline{\boldsymbol{\varepsilon}}$ is the coarse scale strain field while $\boldsymbol{\varepsilon}'$ is the fine scale strain field. Because of (16), $\boldsymbol{\varepsilon}'$ is defined locally within an element and is discontinuous across element boundaries.

We define the corresponding discrete spaces of functions $\mathcal{S}^h$, $\mathcal{P}^h$, and $\mathcal{V}^h$, and write the Galerkin form as;

$$\int_\Omega \nabla \boldsymbol{w}^h : \boldsymbol{\sigma}^h d\Omega = \int_\Omega \boldsymbol{w}^h\cdot\boldsymbol{b}\, d\Omega + \int_\Gamma \boldsymbol{w}^h\cdot\boldsymbol{t}\, d\Gamma \tag{18}$$

$$\int_\Omega q^h\left(\mathbf{1}^T\boldsymbol{\varepsilon}^h - \frac{p^h}{K}\right) d\Omega = 0 \tag{19}$$

The decomposed form of the displacement field $\boldsymbol{u}$ given in (14) can be represented via interpolation functions as

$$\boldsymbol{u}^h(\boldsymbol{x}) = \sum_{a=1}^{n_{el}} \overline{N}_a\overline{\boldsymbol{u}}_a + \sum_{a=1}^{n_b} N'_a\boldsymbol{u}'_a \tag{20}$$

where n_{el} is the number of displacement nodes in the element, and n_b is the number of interpolation functions for the fine scales in the element. $\overline{\boldsymbol{N}}$ represents coarse scale shape function that can be associated with the standard Lagrange interpolation functions, while $\boldsymbol{N}'$ represents fine scale shape functions and in general they can be represented by any functions that satisfy (16). Without loss of generality we assume that in the present problem $\boldsymbol{N}'$ are represented by the so-called bubble functions.

Remark 3: Shape functions $\overline{\boldsymbol{N}}$ and $\boldsymbol{N}'$ should be linearly independent so as to ensure that $\overline{\boldsymbol{u}}(\boldsymbol{x})\cap\boldsymbol{u}'(\boldsymbol{x})=\boldsymbol{0}$.

The pressure field $p^h(\boldsymbol{x})$ is represented as

$$p^h(\boldsymbol{x}) = \sum_{a=1}^{n_p} N_a^p p_a \tag{21}$$

where n_p is the number of pressure nodes, $\boldsymbol{N}^p$ represents the shape functions for the pressure field, and $\boldsymbol{p}$ represents the nodal values of the field.

3.2 Decomposition of the Galerkin Form. Substituting (15) into (18) and exploiting the linearity of the weighting function slot we can split Eq. (18) in a coarse scale and a fine scale problem. Since we have assumed fine scale pressure $p'=0$, therefore, $q'=0$. Accordingly, Eq. (19) only contributes to the coarse scale problem. Since the trial solutions and weighting functions are implied to be functions of h, therefore, to keep the notation simple, explicit dependence on h is suppressed.

Coarse scale sub-problem:

$$\int_\Omega \nabla \overline{\boldsymbol{w}} : \boldsymbol{\sigma}\, d\Omega = \int_\Omega \overline{\boldsymbol{w}} \cdot \boldsymbol{b}\, d\Omega + \int_\Gamma \overline{\boldsymbol{w}} \cdot \boldsymbol{t}\, d\Gamma \tag{22}$$

$$\int_\Omega q\left(\mathbf{1}^T\boldsymbol{\varepsilon} - \frac{p}{K}\right) d\Omega = 0 \tag{23}$$

Fine scale sub-problem:

$$\int_\Omega \nabla \boldsymbol{w}' : \boldsymbol{\sigma}\, d\Omega = \int_\Omega \boldsymbol{w}' \cdot \boldsymbol{b}\, d\Omega + \int_\Gamma \boldsymbol{w}' \cdot \boldsymbol{t}\, d\Gamma \tag{24}$$

Our objective at this point is to solve the fine scale problem, defined over the sum of element interiors, to obtain the fine scale displacement field $\boldsymbol{u}'$. This fine scale field can then be substituted in the coarse scale problem (22) and (23), thereby eliminating the fine scales, yet retaining their effect.

3.2.1 Solution of the Fine-Scale Sub-Problem. Substituting Eq. (10) into Eq. (24), and noting that $\boldsymbol{w}'|_{\Gamma'}=0$ one obtains

$$\int_{\Omega'} \nabla \boldsymbol{w}' : (\boldsymbol{D}^{uu}\boldsymbol{\varepsilon} + \boldsymbol{D}^{up}p) d\Omega = \int_{\Omega'} \boldsymbol{w}' \cdot \boldsymbol{b}\, d\Omega \tag{25}$$

Inserting Eq. (17) into Eq. (25), and then employing (20) and (21) we get

$$\int_{\Omega'} \nabla \boldsymbol{w}' : \boldsymbol{D}^{uu} : \nabla \overline{\boldsymbol{N}}\, d\Omega\, \overline{\boldsymbol{u}} + \int_{\Omega'} \nabla \boldsymbol{w}' : \boldsymbol{D}^{uu} : \nabla \boldsymbol{N}'\, d\Omega\, \boldsymbol{u}' + \int_{\Omega'} \nabla \boldsymbol{w}' : \boldsymbol{D}^{up} \cdot \boldsymbol{N}^p d\Omega\, \boldsymbol{p} = \int_{\Omega'} \boldsymbol{w}' \cdot \boldsymbol{b}\, d\Omega \tag{26}$$

The weighting functions $\boldsymbol{w}'$ corresponding to fine scales are also represented via bubble functions $\boldsymbol{N}'$

$$\boldsymbol{w}' = (\boldsymbol{N}'\,\boldsymbol{\delta d}')^T = (\boldsymbol{\delta d}')^T (\boldsymbol{N}')^T \tag{27}$$

Substituting Eq. (27) into Eq. (26) and by employing arbitrariness of $\boldsymbol{w}'$ we can write the variational problem in its matrix form as follows:

$$\boldsymbol{K}_1 \boldsymbol{u}' + \boldsymbol{K}_2 \overline{\boldsymbol{u}} + \boldsymbol{K}_3 \boldsymbol{p} = \boldsymbol{R}_1 \tag{28}$$

where

$$\boldsymbol{K}_1 = \int_{\Omega'} \nabla^T \boldsymbol{N}' \boldsymbol{D}^{uu} \nabla \boldsymbol{N}'\, d\Omega \tag{29}$$

$$\boldsymbol{K}_2 = \int_{\Omega'} \nabla^T \boldsymbol{N}' \boldsymbol{D}^{uu} \nabla \overline{\boldsymbol{N}}\, d\Omega \tag{30}$$

$$\boldsymbol{K}_3 = \int_{\Omega'} \nabla^T \boldsymbol{N}' \boldsymbol{D}^{up} \boldsymbol{N}^p d\Omega \tag{31}$$

$$\boldsymbol{R}_1 = \int_{\Omega'} (\boldsymbol{N}')^T \cdot \boldsymbol{b}\, d\Omega \tag{32}$$

The fine scale displacement coefficients $\boldsymbol{u}'$ can now be obtained from (28) as follows.

$$\boldsymbol{u}' = \boldsymbol{K}_1^{-1}[\boldsymbol{R}_1 - \boldsymbol{K}_2\overline{\boldsymbol{u}} - \boldsymbol{K}_3\boldsymbol{p}] \tag{33}$$

3.2.2 Solution of the Coarse-Scale Sub-Problem. Substituting Eq. (10) into Eq. (22), one obtains

$$\int_\Omega \nabla \overline{\boldsymbol{w}} : (\boldsymbol{D}^{uu}\boldsymbol{\varepsilon} + \boldsymbol{D}^{up}p) d\Omega = \int_\Omega \overline{\boldsymbol{w}} \cdot \boldsymbol{b}\, d\Omega + \int_\Gamma \overline{\boldsymbol{w}} \cdot \boldsymbol{t}\, d\Gamma \tag{34}$$

Inserting Eq. (17) into Eq. (34) and then employing (20) and (21) we get

$$\int_\Omega \nabla \overline{\boldsymbol{w}} : \boldsymbol{D}^{uu} : \nabla \overline{\boldsymbol{N}}\, d\Omega\, \overline{\boldsymbol{u}} + \int_\Omega \nabla \overline{\boldsymbol{w}} : \boldsymbol{D}^{uu} : \nabla \boldsymbol{N}'\, d\Omega\, \boldsymbol{u}' + \int_\Omega \nabla \overline{\boldsymbol{w}} : \boldsymbol{D}^{up} \cdot \boldsymbol{N}^p d\Omega\, \boldsymbol{p} = \int_\Omega \overline{\boldsymbol{w}} \cdot \boldsymbol{b}\, d\Omega + \int_\Gamma \overline{\boldsymbol{w}} \cdot \boldsymbol{t}\, d\Gamma \tag{35}$$

The weighting functions $\overline{\boldsymbol{w}}$ corresponding to the coarse scale displacements are represented via the standard interpolation functions $\overline{\boldsymbol{N}}$.

$$\overline{\boldsymbol{w}} = (\overline{\boldsymbol{N}}\,\overline{\boldsymbol{\delta d}})^T = (\overline{\boldsymbol{\delta d}})^T (\overline{\boldsymbol{N}})^T \tag{36}$$

Substituting Eq. (36) into Eq. (35), and employing arbitrariness of $\overline{\boldsymbol{w}}$, we can write the matrix form as follows:

$$\boldsymbol{K}_4\overline{\boldsymbol{u}} + \boldsymbol{K}_5\boldsymbol{u}' + \boldsymbol{K}_6\boldsymbol{p} = \boldsymbol{R}_2 \tag{37}$$

Substituting the fine scale coefficients $\boldsymbol{u}'$ from Eq. (33) leads to the following matrix form

$$\boldsymbol{K}_4\overline{\boldsymbol{u}} + \boldsymbol{K}_5\boldsymbol{K}_1^{-1}[\boldsymbol{R}_1 - \boldsymbol{K}_2\overline{\boldsymbol{u}} - \boldsymbol{K}_3\boldsymbol{p}] + \boldsymbol{K}_6\boldsymbol{p} = \boldsymbol{R}_2 \tag{38}$$

which can be further simplified as

$$(\boldsymbol{K}_4 - \boldsymbol{K}_5\boldsymbol{K}_1^{-1}\boldsymbol{K}_2)\overline{\boldsymbol{u}} + (\boldsymbol{K}_6 - \boldsymbol{K}_5\boldsymbol{K}_1^{-1}\boldsymbol{K}_3)\boldsymbol{p} = \boldsymbol{R}_2 - \boldsymbol{K}_5\boldsymbol{K}_1^{-1}\boldsymbol{R}_1 \tag{39}$$

where

$$\boldsymbol{K}_4 = \int_\Omega \nabla^T \overline{\boldsymbol{N}} \boldsymbol{D}^{uu} \nabla \overline{\boldsymbol{N}}\, d\Omega \tag{40}$$

$$\boldsymbol{K}_5 = \int_\Omega \nabla^T \overline{\boldsymbol{N}} \boldsymbol{D}^{uu} \nabla \boldsymbol{N}'\, d\Omega \tag{41}$$

$$\boldsymbol{K}_6 = \int_\Omega \nabla^T \overline{\boldsymbol{N}} \boldsymbol{D}^{up} \boldsymbol{N}^p d\Omega \tag{42}$$

$$\boldsymbol{R}_2 = \int_\Omega \overline{\boldsymbol{N}}^T \cdot \boldsymbol{b}\, d\Omega + \int_\Gamma \overline{\boldsymbol{N}}^T \cdot \boldsymbol{t}\, d\Gamma \tag{43}$$

Remark 4: It is important to note that (39) is completely expressed in terms of the coarse/resolvable scales of the problem.

3.2.3 The Volumetric Constitutive Equation. Substituting Eq. (10) in the volumetric constitutive Eq. (23) and employing (20) and (21) we get

$$\int_\Omega \boldsymbol{q}\mathbf{1}^T \nabla \overline{\boldsymbol{N}}\, d\Omega\, \bar{\mathbf{u}} + \int_\Omega \boldsymbol{q}\mathbf{1}^T \nabla \boldsymbol{N}'\, d\Omega\, \boldsymbol{u}' - \int_\Omega \boldsymbol{q}\frac{1}{K}\boldsymbol{N}^p d\Omega\, \boldsymbol{p} = 0 \tag{44}$$

The weighting function $\boldsymbol{q}$ corresponding to pressure $\boldsymbol{p}$ can be represented as

$$\boldsymbol{q} = \boldsymbol{N}^p \boldsymbol{\delta p} \tag{45}$$

Substituting Eq. (45) into Eq. (44) and employing arbitrariness of $\boldsymbol{q}$ leads to the following matrix form of equations.

$$\boldsymbol{K}_7 \overline{\boldsymbol{u}} + \boldsymbol{K}_8 \boldsymbol{u}' + \boldsymbol{K}_9 \boldsymbol{p} = \mathbf{0} \tag{46}$$

Substituting the fine scale coefficients $\boldsymbol{u}'$ from Eq. (33) into (46) yields the following.

$$\boldsymbol{K}_7 \overline{\boldsymbol{u}} + \boldsymbol{K}_8 \boldsymbol{K}_1^{-1}[\boldsymbol{R}_1 - \boldsymbol{K}_2 \overline{\boldsymbol{u}} - \boldsymbol{K}_3 \boldsymbol{p}] + \boldsymbol{K}_9 \boldsymbol{p} = \mathbf{0} \tag{47}$$

Simplifying this expression we get

$$(\boldsymbol{K}_7 - \boldsymbol{K}_8 \boldsymbol{K}_1^{-1} \boldsymbol{K}_2)\overline{\boldsymbol{u}} + (\boldsymbol{K}_9 - \boldsymbol{K}_8 \boldsymbol{K}_1^{-1} \boldsymbol{K}_3)\boldsymbol{p} = -\boldsymbol{K}_8 \boldsymbol{K}_1^{-1} \boldsymbol{R}_1 \tag{48}$$

where

$$\boldsymbol{K}_7 = \int_\Omega \boldsymbol{N}^{pT} \boldsymbol{D}^{pu} \nabla \overline{\boldsymbol{N}}\, d\Omega \tag{49}$$

$$\boldsymbol{K}_8 = \int_\Omega \boldsymbol{N}^{pT} \boldsymbol{D}^{pu} \nabla \boldsymbol{N}'\, d\Omega \tag{50}$$

$$\boldsymbol{K}_9 = -\int_\Omega \boldsymbol{N}^{pT} D^{pp} \boldsymbol{N}^p d\Omega \tag{51}$$

$$\boldsymbol{D}^{pu} = \mathbf{1}^T = [1\ 1\ 1\ 0\ 0\ 0] \tag{52}$$

$$D^{pp} = \frac{1}{K} \tag{53}$$

3.3 Matrix Form of the HVM Formulation. Combining Eqs. (39) and (48), the mixed displacement-pressure formulation can be written in the matrix form as follows:

$$\begin{bmatrix} \boldsymbol{K}_4 - \boldsymbol{K}_5 \boldsymbol{K}_1^{-1} \boldsymbol{K}_2 & \boldsymbol{K}_6 - \boldsymbol{K}_5 \boldsymbol{K}_1^{-1} \boldsymbol{K}_3 \\ \boldsymbol{K}_7 - \boldsymbol{K}_8 \boldsymbol{K}_1^{-1} \boldsymbol{K}_2 & \boldsymbol{K}_9 - \boldsymbol{K}_8 \boldsymbol{K}_1^{-1} \boldsymbol{K}_3 \end{bmatrix} \begin{Bmatrix} \overline{\boldsymbol{u}} \\ \boldsymbol{p} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{R}_2 - \boldsymbol{K}_5 \boldsymbol{K}_1^{-1} \boldsymbol{R}_1 \\ -\boldsymbol{K}_8 \boldsymbol{K}_1^{-1} \boldsymbol{R}_1 \end{Bmatrix} \tag{54}$$

where $\overline{\boldsymbol{u}}$ represents the unknown displacement degrees of freedom and $\boldsymbol{p}$ represents the unknown pressure degrees of freedom. Now that we have derived the stabilized/multiscale formulation, we can make some simplifications in (54) by noting the following: $\boldsymbol{K}_6 = \boldsymbol{K}_7^T$; $\boldsymbol{K}_8 = \boldsymbol{K}_3^T$; $\boldsymbol{K}_2 = \boldsymbol{K}_5^T$ and $\boldsymbol{K}_1^{-1} = \boldsymbol{K}_1^{-T}$. Accordingly, Eq. (54) can be written as

$$\begin{bmatrix} \boldsymbol{K}_4 - \boldsymbol{K}_2^T \boldsymbol{K}_1^{-1} \boldsymbol{K}_2 & \boldsymbol{K}_6 - \boldsymbol{K}_2^T \boldsymbol{K}_1^{-1} \boldsymbol{K}_3 \\ \boldsymbol{K}_6^T - \boldsymbol{K}_3^T \boldsymbol{K}_1^{-1} \boldsymbol{K}_2 & \boldsymbol{K}_9 - \boldsymbol{K}_3^T \boldsymbol{K}_1^{-1} \boldsymbol{K}_3 \end{bmatrix} \begin{Bmatrix} \overline{\boldsymbol{u}} \\ \overline{\boldsymbol{p}} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{R}_2 - \boldsymbol{K}_2^T \boldsymbol{K}_1^{-1} \boldsymbol{R}_1 \\ -\boldsymbol{K}_3^T \boldsymbol{K}_1^{-1} \boldsymbol{R}_1 \end{Bmatrix} \tag{55}$$

Remark 5: Equation (55) is the stabilized finite element matrix form for the mixed displacement-pressure form of elasticity.

Remark 6: The left-hand side matrix in (55) is a symmetric matrix which is expressed entirely in terms of the coarse/resolvable computational scales of the problem. Fine scales have been substituted for by the additional terms in the matrix. These terms stabilize the formulation while consistently representing the fine computational scales in the problem.

Remark 7: An analogy with the stability parameters employed in the stabilized methods for Stokes flow can be made by observing that $\boldsymbol{K}_1^{-1}$ is playing the role of the so-called stability parameter in (55). It is important to note that in the present method, an explicit definition of $\boldsymbol{K}_1^{-1}$ appears and is given in (29).

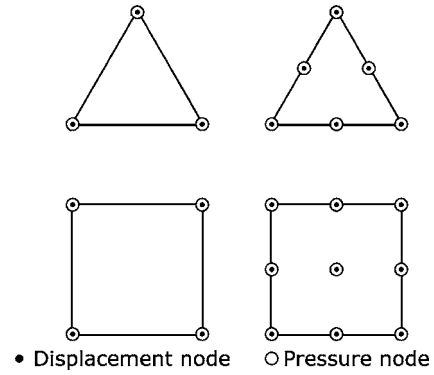


Fig. 1 A family of 2D linear and quadratic elements

Remark 8: We can make a direct comparison of Eq. (55) with the standard mixed form for elasticity that can be obtained by dropping the stabilization terms. Without the subscale effect, Eq. (55) will reduce to matrix form of Eqs. (8) and (9) as follows:

$$\begin{bmatrix} \boldsymbol{K}_4 & \boldsymbol{K}_6 \\ \boldsymbol{K}_6^T & \boldsymbol{K}_9 \end{bmatrix} \begin{Bmatrix} \overline{\boldsymbol{u}} \\ \boldsymbol{p} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{R}_2 \\ \mathbf{0} \end{Bmatrix}$$

As mentioned earlier, this form is required to satisfy the *inf-sup* condition [20] to yield stable and convergent elements in the incompressible limit.

4 Numerical Simulations

The stabilized formulation (55) has been used for several test problems presented in this section. Figure 1 shows the family of two-dimensional (2D) elements employed in the numerical studies. Dots correspond to the displacement nodes and circles correspond to the pressure nodes. Q4B and Q9B are 4- and 9-node quadrilaterals with one bubble function, and T3B and T6B are the 3- and 6-node triangles with one bubble function. In each case, appropriate quadrature rules were used to fully integrate the stiffness matrices.

Figure 2 shows the location of the bubble functions employed for the quadrilaterals and triangles, respectively. We have employed the simplest polynomial bubbles, which in the element natural coordinate frame are expressed as follows:

$$N(\xi, \eta)_{\text{quads.}} = (1 - \xi^2)(1 - \eta^2) \tag{56}$$

$$N(\xi, \eta)_{\text{triangles}} = 27rst = 27\xi\eta(1 - \xi - \eta) \tag{57}$$

The following section presents various patch tests and beam bending problems which serve as standard benchmark problems in the solid mechanics literature. The superior performance of the stabilized elements under severe geometrical distortions is also presented. Also presented are numerical rate of convergence studies for these elements, which confirm optimal convergence rates in the norms considered.

4.1 Plane Stress Patch Tests. The first set of numerical simulations consist of plane stress patch tests [26]. The uniform mesh

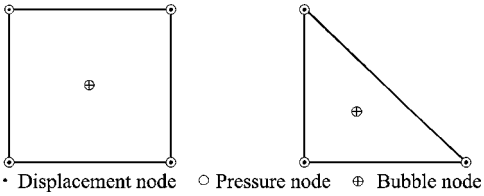


Fig. 2 Bubble functions employed for quadrilaterals and triangles

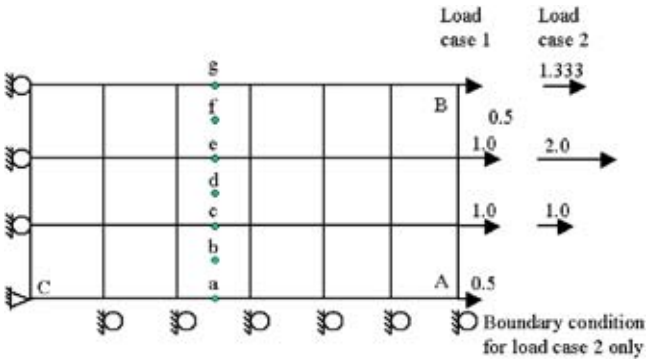


Fig. 3 Patch test with regular elements

configuration and the skewed mesh configuration used in the tests are shown in Figs. 3 and 4, respectively. Meshes for triangular elements were generated by splitting the quadrilaterals.

The elastic coefficients used in both cases are $E=1$ and $\nu=0.25$. The first test case in Fig. 3 is an axial stretch with nodal loads equivalent to a pure axial normal stress of unit intensity applied along edge AB. The exact solution is $u_1=x_1$ and $u_2=-x_2/4$. (The origin for the two meshes is taken to be at point C.) The second test case is a linearly varying normal stress of magnitude x_2 on edge AB. Due to antisymmetry of the problem, half of the mesh is modelled, with antisymmetry boundary conditions imposed on the nodes along edge AC. The bending solution to this problem is $u_1=x_1x_2$ and $u_2=-\frac{1}{2}((x_1)^2-\nu(x_2)^2)$.

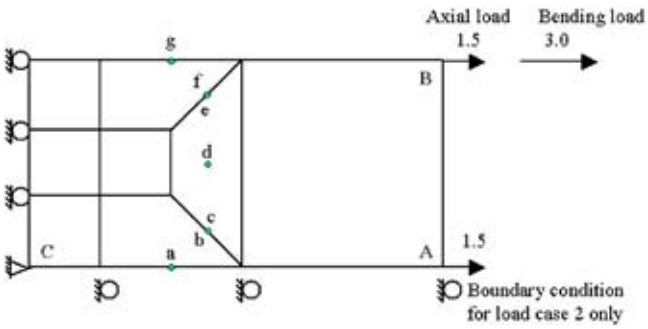


Fig. 4 Patch test with distorted elements

Table 1 Normalized displacements at point A in Figs. 3 and 4

	Uniform mesh		Skewed mesh	
Element Type	u_1(A)-axial	u_2(A)-bending	u_1(A)-axial	u_2(A)-bending
Q4	1.00	0.987	1.00	0.944
QM6	1.00	1.00	1.00	0.978
Q4B	1.00	0.996	1.00	0.972
T3B	1.00	0.958	1.00	0.879

Table 1 shows normalized displacements evaluated at node A for the axial and bending deflections for the two mesh configurations. For comparison purposes we have presented the response of Q4 which is the standard 4-node quadrilateral and QM6 which is the Taylor-Wilson [19] incompatible modes quadrilateral element. Accordingly, we present the performance of the linear triangle T3B and bilinear quadrilateral Q4B. The axial stretch in both the configurations is exactly satisfied. The bending solution to an applied moment in the form of a couple is a quadratic polynomial. In the uniform mesh configuration, only element QM6 with incompatible modes captures the solution exactly. This is because the incompatible modes contain quadratic terms in the coordinates of the parent domain ξ and η which coincide with the global coordinates for the uniform mesh and can, therefore, fully represent the exact solution. However, the incompatible modes are not complete second-order polynomials, and therefore, in the skewed mesh configuration, are unable to fully capture the exact displacement.

The axial stress at points a–g is exact for the axial stress case on both the mesh configurations. Therefore these results are not presented here. For the bending load case, the stresses obtained for Q4B at the points a–g are presented in Tables 2 and 3.

Remark 8: By studying patch tests we can detect the rigid body modes, any false zero energy modes and invariance of the element under change in global orientation. Preliminary tests, not presented herein, indicate that our elements do not possess false zero energy modes and are invariant under change in global orientation.

4.2 Sensitivity to Mesh Distortion. This is a standard test for evaluating the sensitivity of the elements to mesh distortion and serves as a benchmark problem in element evaluation. Such testing has been an important ingredient in element development, investigating completeness of the interpolation polynomial and checking the constant stress states of the element. A cantilever beam, modeled by two quadrilateral elements, is subjected to a bending moment in the form of a couple (see Fig. 5). The edge separating the two elements is then gradually rotated about its center, a distance of $\pm a$ on the top and the bottom surfaces, to distort the mesh. The degree of geometric distortion of these elements is represented by the dimension a. The same test is performed on triangles by bisecting the quadrilaterals. The elements used in the test are 6-node triangles and 9-node quadrilaterals (see Fig. 6). Linear triangles and bilinear quadrilaterals are not presented because of their inability to capture bending behavior

Table 2 Normalized axial stresses for the bending load case on the uniform mesh (except for point a, where the exact solution is zero)

Element	a	b	c	d	e	f	g
Q4	–0.033	0.988	0.988	0.987	0.986	0.987	0.998
QM6	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Q4B	–0.012	1.010	1.000	0.996	0.994	0.993	0.992

Table 3 Normalized axial stresses for the bending load case on the skewed mesh (except for point a, where the exact solution is zero)

Element	a	b	c	d	e	f	g
Q4	–0.008	0.798	1.204	0.957	0.908	0.986	0.946
QM6	–0.007	0.910	0.908	0.986	1.001	0.999	1.003
Q4B	–0.008	0.721	1.201	0.975	0.927	1.016	0.944

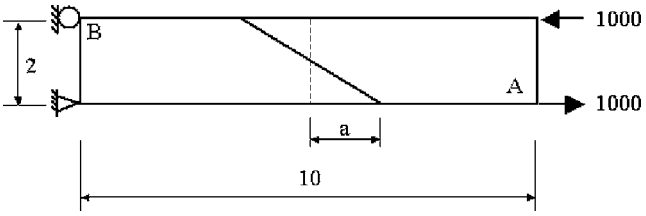


Fig. 5 Sensitivity to mesh distortion. 9-node quadrilaterals.

(which is a quadratic polynomial) even in the undeformed configuration. The normalized displacement at point A and the normalized stress at point B, respectively, are presented in Table 4. The exact solution is a quadratic polynomial, and hence quadratic elements show no deterioration with mesh distortion as long as the edges are kept straight (Table 4).

4.3 Rate of Convergence Study. This section presents the mathematical rate of convergence study for displacements in the $L_2(\Omega)$ norm and the energy norm, and for pressure in the $L_2(\Omega)$ norm. The exact solution depends on Poisson's ratio; the value 0.4999 is employed in the calculations to model the nearly incompressible behavior of the material. In the present study, plane strain conditions are assumed in force. The configuration considered here is shown in Fig. 7. A cantilever beam of length-to-depth ratio equal to five is subjected to a parabolically varying end load. Boundary conditions are set in accordance with an exact elasticity solution. This is a standard problem that is employed to assess the performance of plane-stress–strain elements subjected to dominant in-plane bending behavior. The meshes shown consist of 16 quadrilateral and 32 triangular elements. Finer meshes are constructed by bisection. In the case of the quadratic elements, a coarser mesh is also employed with one layer of elements through the depth.

The exact solution to an applied shear force is a third-order polynomial. Numerical examples in the following section show that with successive mesh refinements, the finite element solution converges to the exact solution at nearly the optimal rate of convergence for the norms considered. For the linear elements, i.e., 3-node triangle and the 4-node quadrilateral, the theoretical rate of convergence for the displacement field in the $L_2(\Omega)$ norm and the energy norm is 2 and 1, respectively, while optimal rate for the pressure field in $L_2(\Omega)$ norm is 2. Figures 8–10 present the numerical rates of convergence of linear elements. Figure 8 shows convergence in the $L_2(\Omega)$ norm of the displacement field and we

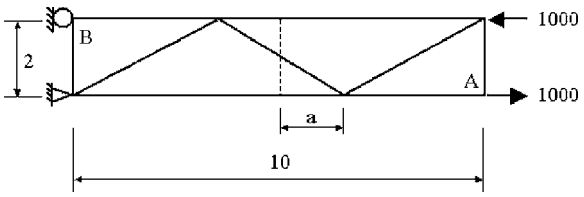


Fig. 6 Sensitivity to mesh distortion. 6-node triangles.

Table 4 Normalized displacements at node A and normalized stress at node B

Distortion	Normalized displacements at node A		Normalized stress at node B	
a	T6B	Q9B	T6B	Q9B
0	1.0	1.0	1.0	1.0
1	1.0	1.0	1.0	1.0
2	1.0	1.0	1.0	1.0
3	1.0	1.0	1.0	1.0
4	1.0	1.0	1.0	1.0

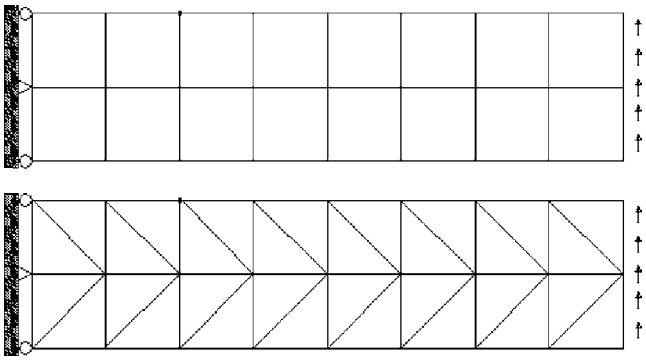


Fig. 7 Cantilever beam with edge shear

see nearly optimal rate of convergence. Figure 9 presents convergence in the energy norm and optimal rates are attained. The convergence in the $L_2(\Omega)$ norm of the pressure field is presented in Fig. 10 which is again nearly optimal.

For the quadratic elements, i.e., 6-node triangle and 9-node quadrilateral, the theoretical convergence rates for the displacement field in $L_2(\Omega)$ and energy norms are 3 and 2, respectively, while optimal rate for the pressure field in $L_2(\Omega)$ norm is 3. The corresponding numerical rates are presented in Figs. 11–13. Ex-

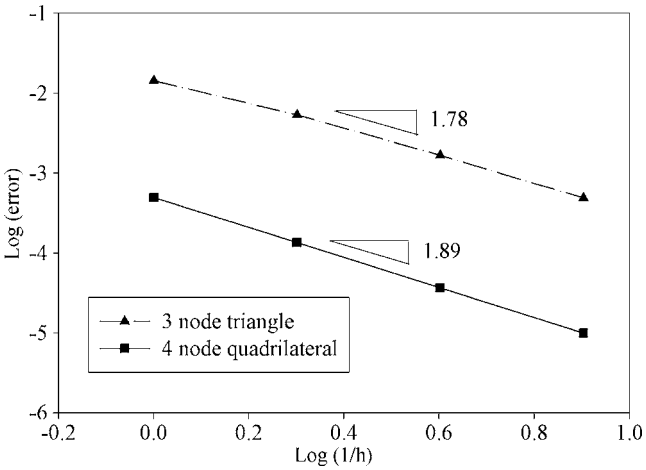


Fig. 8 Convergence rates for the L_2 norm of the displacement field (linear elements)

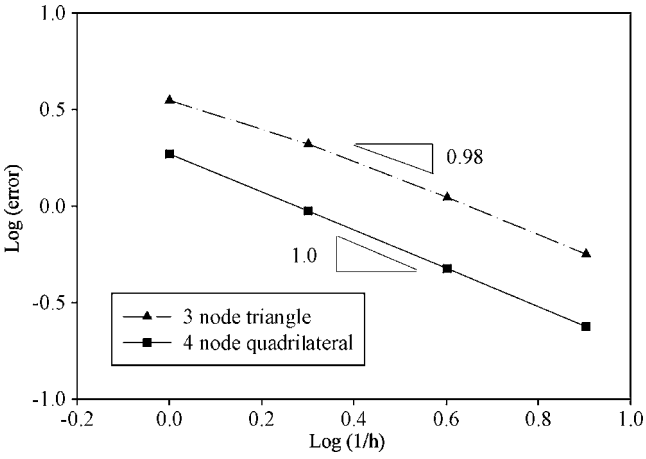


Fig. 9 Convergence rates for the energy norm (linear elements)

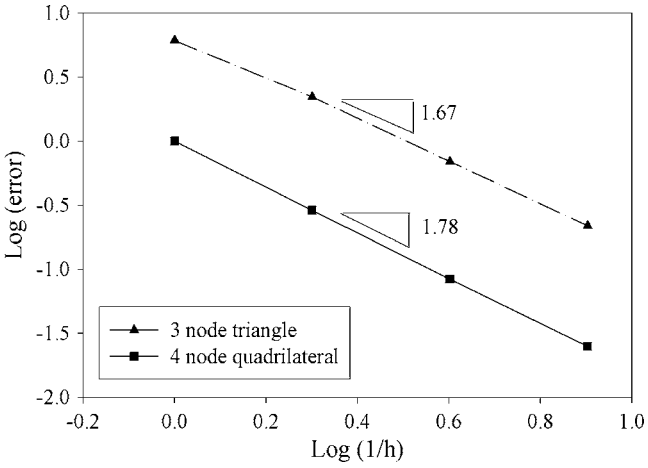


Fig. 10 Convergence rates for the L_2 norm of the pressure field (linear elements)

cept for the $L_2(\Omega)$ norm of the pressure field for 6-node triangle, which is sub-optimal, we get theoretical convergence rates of all the other fields in the quadratic elements.

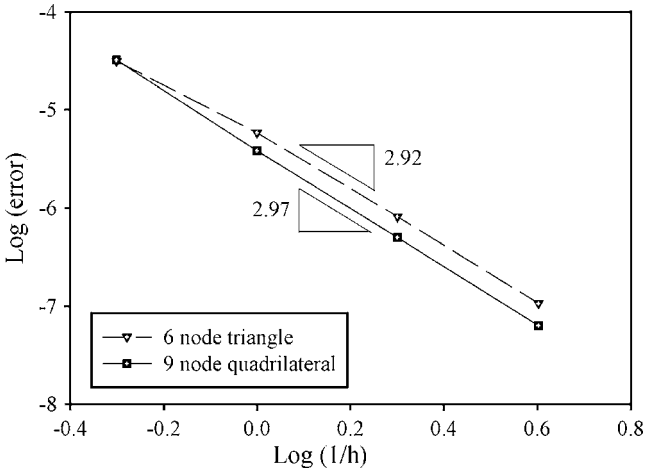


Fig. 11 Convergence rates for the L_2 norm of the displacement field (quadratic elements)

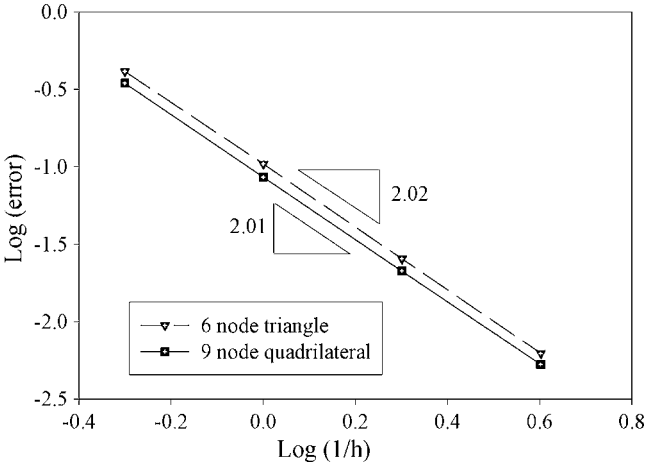


Fig. 12 Convergence rates for the energy norm (quadratic elements)

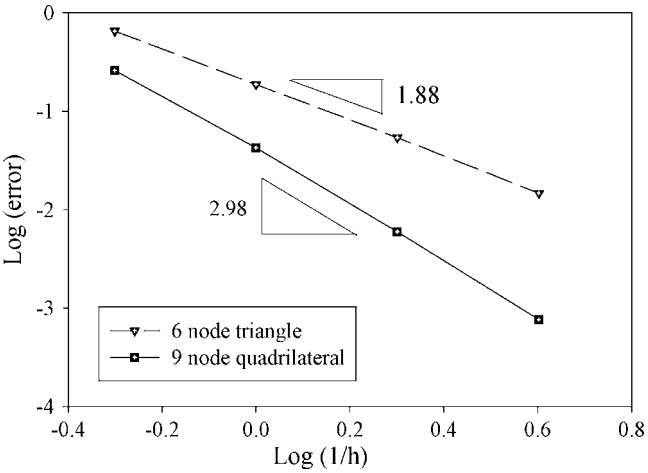


Fig. 13 Convergence rates for the L_2 norm of the pressure field (quadratic elements)

4.4 Accuracy Study. This test presents the engineering convergence properties for the family of elements. A cantilever beam is loaded via edge shear, and boundary conditions are set in accordance with the theory of elasticity (see Fig. 7). In these simulations Poisson's ratio is 0.4999 to simulate nearly incompressible behavior, and plane strain conditions are assumed in force. Figure 14 shows the pressure contours for the 3-node triangle T3B, while Fig. 15 presents the contours for the 4-node quadrilateral Q4B. Figure 16 shows a composite mesh made of triangles and quadrilaterals in the same computational domain with the superposed pressure contours. This figure is intended to show the advantage of using the stabilized methods wherein one can arbitrarily com-

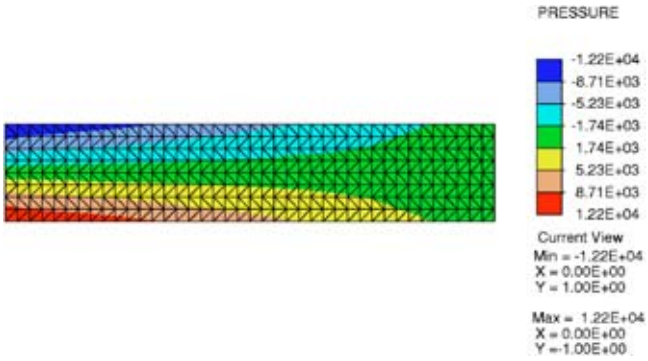


Fig. 14 Contours of the pressure field for the 3-node element mesh

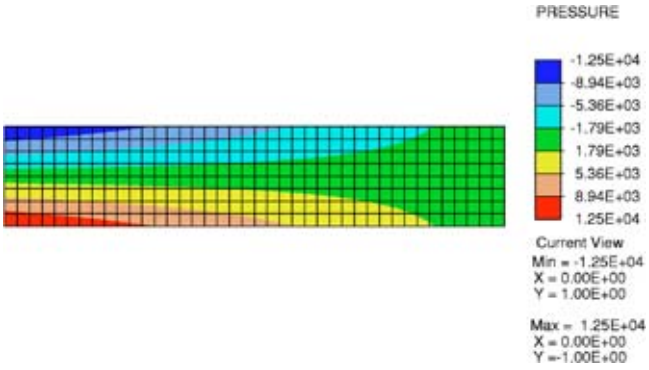


Fig. 15 Contours of the pressure field for the 4-node element mesh

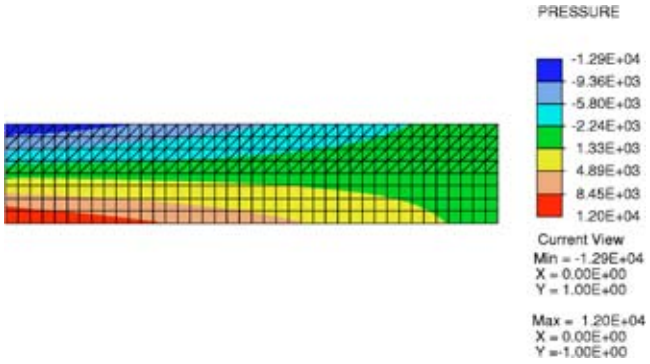


Fig. 16 Contours of the pressure field for the composite mesh

bine various element types in the same computational domain; a feature which can be very appealing from a practical view point of problem solving.

Figures 17 and 18 present the normalized tip deflection convergence and normalized stress convergence for the linear triangle T3B and the bilinear quadrilateral Q4B, respectively. Likewise, Figs. 19 and 20 present the normalized tip deflection convergence and normalized stress convergence for the quadratic elements T6B and Q9B, respectively. These plots present uniform convergence of displacements and stresses, and also show substantial increase in the coarse-mesh accuracy attained by increasing the order of

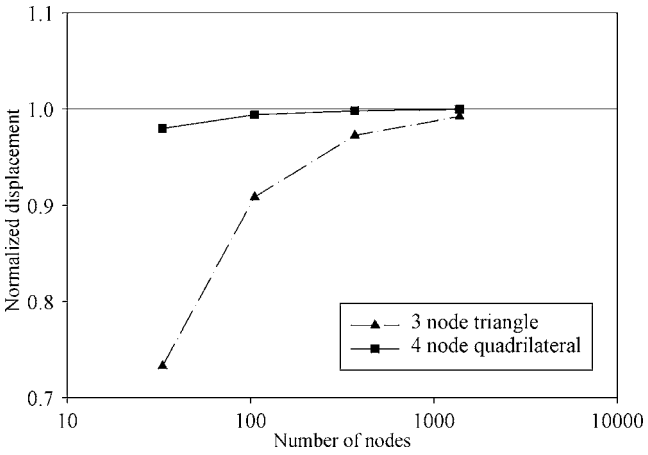


Fig. 17 Tip deflection convergence for 3-node and 4-node elements

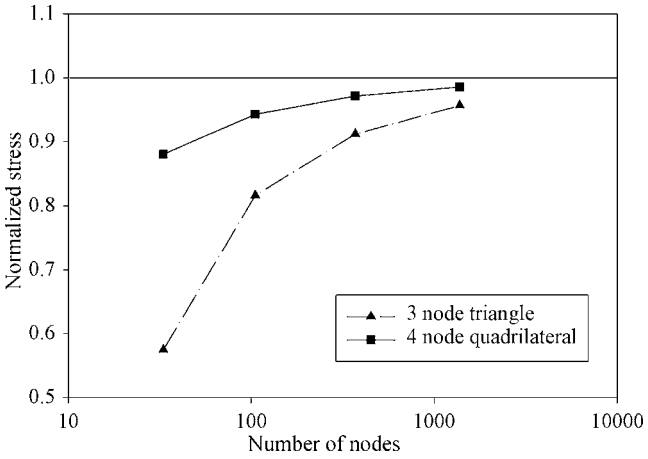


Fig. 18 Stress convergence for 3-node and 4-node elements

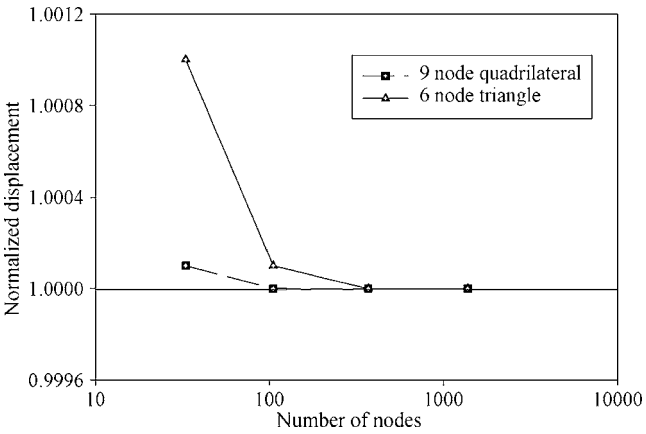


Fig. 19 Tip deflection convergence for 6-node and 9-node elements

interpolation. As expected, the linear triangle and bilinear quadrilateral are comparatively stiff, while the accuracy of all the quadratic elements is excellent.

4.5 Cook's Membrane. This problem, first proposed in Ref. [16] as a test case for general quadrilateral elements, shows the bending performance of the elements under excessive mesh distortion. The configuration is a tapered panel with one edge fixed and the opposite edge acted upon by a distributed shear load (see Fig. 21). There is no known analytic solution for this problem but the results for a 32×32 mesh are used for comparison purposes. Plane strain conditions are assumed enforced. The Poisson's ratio is 0.4999 as is used in Simo and Armero [18] and Kasper and

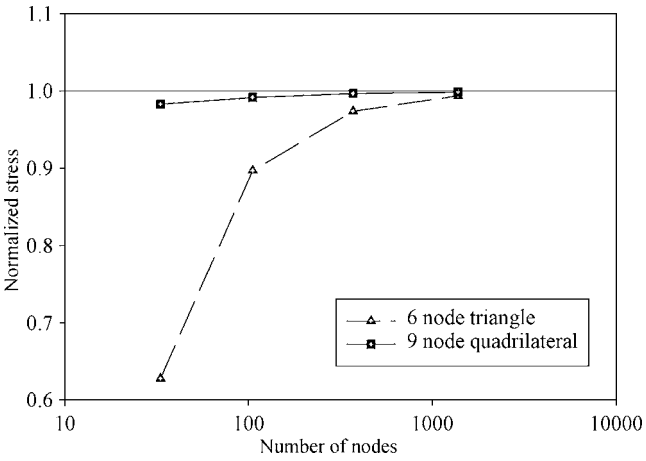


Fig. 20 Stress convergence for 6-node and 9-node elements

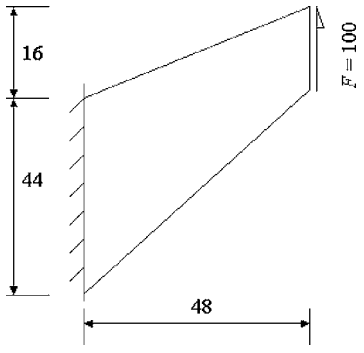


Fig. 21 Cook's membrane

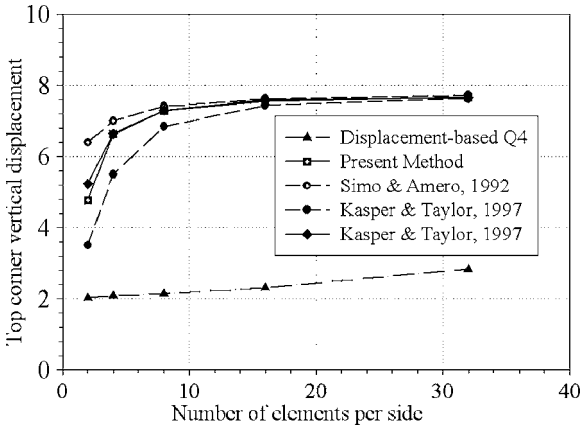


Fig. 22 Tip deflection convergence

Taylor [27]. The vertical deflection at the top corner of the tip is presented in Fig. 22. The standard displacement based 4-node quadrilateral element shows a poor response all across. One of the proposed stabilized-mixed elements, the equal-order 4-node element is compared with the enhanced strain element of Simo and Armero [18] and the mixed-enhanced strain element of Kasper and Taylor [27]. Figure 23 presents the response at the tip as a function of the Poisson's ratio. The present element works all the way to the incompressible limit, as does an element based on the Galerkin/Least-squares stabilization. However, pure displacement based standard element shows severe locking in the incompressible limit.

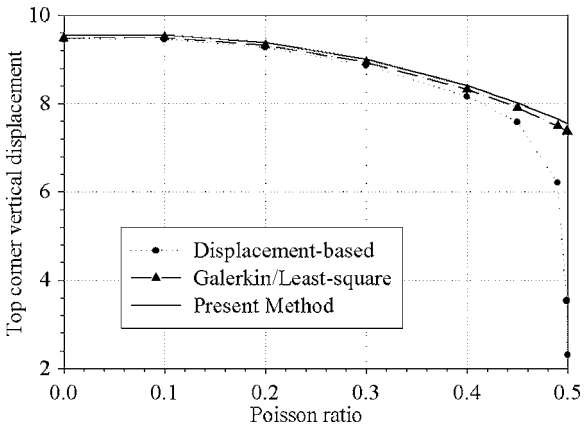


Fig. 23 Tip deflection convergence as a function of Poisson's ratio

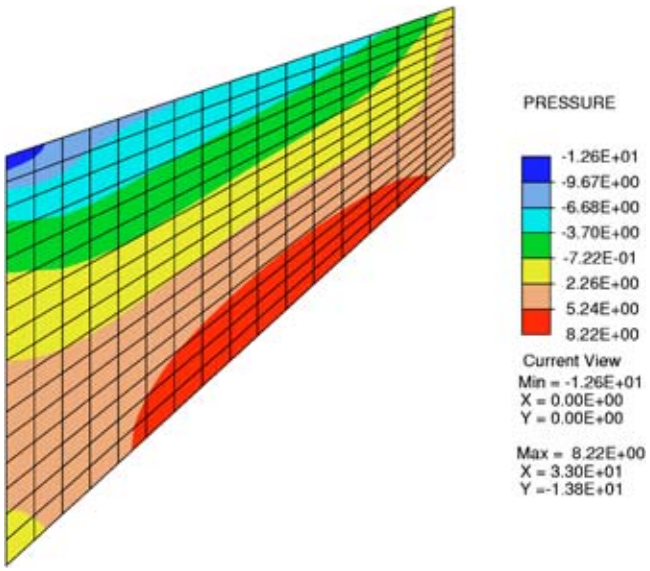


Fig. 24 Contours of the pressure field for the 4-node element mesh

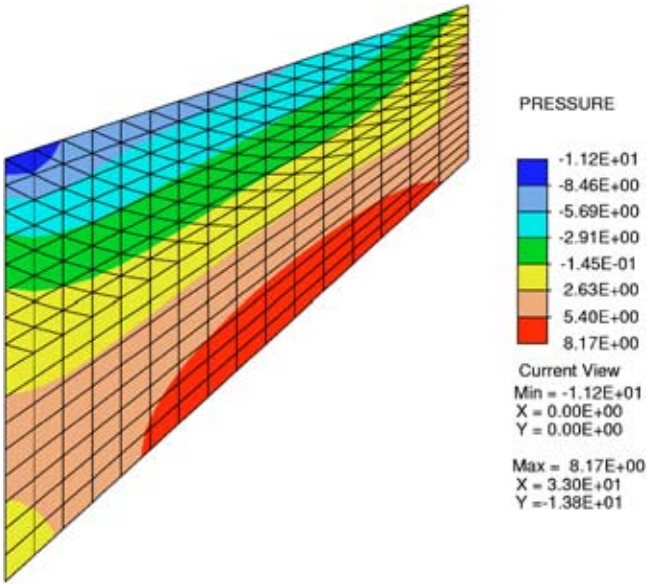


Fig. 25 Contours of the pressure field for the composite mesh

Figure 24 shows the pressure contours for the 4-node quadrilateral Q4B. Figure 25 shows a composite mesh made of triangles and quadrilaterals in the same computational domain with the superposed pressure contours. Once again we obtain a smooth pressure profile for the composite mesh.

5 Conclusions

We have presented an application of the variational multiscale method [25] for developing stabilized finite element formulations for compressible and nearly incompressible elasticity. The novelty of the present method is that the definition of the so-called stability parameter appears naturally in the derivation. The proposed method is based on sound variational foundations, thus providing a basis for a priori error analysis of the system. The resulting finite element formulation allows arbitrary combinations of interpolation functions for the displacement and pressure fields, and thus yields a family of stable and convergent elements. Various benchmark problems have been solved to show that the developed elements do not possess false zero energy modes and are invariant under change in global orientation. Numerical tests of plane stress and plane strain elements have been presented to show the superior accuracy of the elements. Rate of convergence studies have been carried out that show optimal convergence rates in the norms considered and corroborate the theoretical convergence rates for the displacement and the stress fields.

Acknowledgment

This work was supported by NSF Grant No. NSF-CMS-9813386 and ONR Grant No. N00014-02-1-0143. This support is gratefully acknowledged.

References

[1] Herrmann, L. R., 1965, "Elasticity Equations for Nearly Incompressible Materials by a Variational Theorem," AIAA J. **3**, pp. 1896–1900.
[2] Simo, J. C., Taylor, R. L., and Pister, K. S., 1985, "Variational and Projection

Methods for the Volume Constraint in Finite Deformation Elasto-Plasticity," Comput. Methods Appl. Mech. Eng. **51**, pp. 177–208.
[3] Arnold, D. N., Brezzi, F., and Douglas, Jr., J., 1984, "PEERS: A New Mixed Finite Element for Plane Elasticity," Jpn. J. Ind. Appl. Math., Part 1 **1**, pp. 347–367.
[4] Cruchaga, M. A., and Onate, E., 1997, "A Finite Element Formulation for Incompressible Flow Problems Using a Generalized Streamline Operator," Comput. Methods Appl. Mech. Eng. **143**, pp. 49–67.
[5] Kouhia, R., and Stenberg, R., 1995, "A Linear Nonconforming Finite Element Method for Nearly Incompressible Elasticity and Stokes Flow," Comput. Methods Appl. Mech. Eng. **124**, pp. 195–212.
[6] Malkus, D. S., and Hughes, T. J. R., 1987, "Mixed Finite Element Methods-Reduced and Selective Integration Techniques: A Unification of Concepts," Comput. Methods Appl. Mech. Eng. **15**(1), pp. 63–81.
[7] Stenberg, R., 1987, "On Some Three-Dimensional Finite Elements for Incompressible Media," Comput. Methods Appl. Mech. Eng. **63**, pp. 261–269.
[8] Belytschko. T., and Branchrach, W. E., 1986, "Efficient Implementation of Quadrilaterals With High Coarse Mesh Accuracy," Comput. Methods Appl. Mech. Eng. **54**, pp. 279–301.
[9] Belytschko, T., and Bindeman, L. P., 1991, "Assumed Strain Stabilization of the 4-Node Quadrilateral With 1-Point Quadrature for Nonlinear Problems," Comput. Methods Appl. Mech. Eng. **88**, pp. 311–340.
[10] Hughes, T. J. R., 1980, "Generalization of Selective Integration Procedures to Anisotropic and Nonlinear Media," Int. J. Numer. Methods Eng. **15**, pp. 1413–1418.
[11] Hughes, T. J. R., and Malkus, D. S., 1983, "A General Penalty/Mixed Equivalence Theorem for Anisotropic, Incompressible Finite Elements," *Hybrid and Mixed Finite Element Methods*, S. N. Atluri, R. H. Gallagher, and O. C. Zienkiewicz, eds., John Wiley, London, pp. 487–496.
[12] Simo, J. C., and Hughes, T. J. R., 1986, "On the Variational Foundations of Assumed Strain Methods," J. Appl. Mech. **53**, pp. 51–54.
[13] Simo, J. C., and Rifai, M. S., 1990, "A Class of Mixed Assumed Strain Methods and the Method of Incompatible Modes," Int. J. Numer. Methods Eng. **29**, pp. 1595–1636.
[14] Brezzi, F., and Bathe, K. J., 1990, "A Discourse on the Stability Conditions for Mixed Finite Element Formulations," Comput. Methods Appl. Mech. Eng. **82**(1-3), pp. 27–57.
[15] Chapelle, D., and Bathe, K. J., 1993, "The Inf-Sup Test," Comput. Struct. **47**(4-5), pp. 537–545.
[16] Cook, R. D., 1974, "Improved Two-Dimensional Finite Element," J. Struct. Div. ASCE **100**, pp. 1851–1863.
[17] Klaas, O., Maniatty, A. M., and Shephard, M. S., 1999, "A Stabilized Mixed Petrov-Galerkin Finite Element Method for Finite Elasticity. Formulation for Linear Displacement and Pressure Interpolation," Comput. Methods Appl. Mech. Eng. **180**, pp. 65–79.
[18] Simo, J. C., and Armero, F., 1992, "Geometrically Nonlinear Enhanced Strain Mixed Methods and the Method of Incompatible Modes," Int. J. Numer. Methods Eng. **33**, pp. 1413–1449.
[19] Taylor, R. L., Beresford, P. J., and Wilson, E. L., 1976, "A Nonconforming Element for Stress Analysis," Int. J. Numer. Methods Eng. **10**, pp. 1211–1219.
[20] Brezzi, F., and Fortin, M., 1991, *Mixed and Hybrid Finite Element Methods*, Vol. 15, Springer Series in Computational Mathematics, Springer-Verlag, NY, 1991.
[21] Hughes, T. J. R., 1987, *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*, Prentice-Hall, Englewoods Cliffs, NJ (Dover Edition, 2000), Chap. 4.
[22] Hughes, T. J. R., and Franca, L. P., 1987, "A New Finite Element Formulation for Computational Fluid Dynamics: VII. The Stokes Problem With Various Well-Posed Boundary Conditions: Symmetric Formulations That Converge for All Velocity/Pressure Spaces," Comput. Methods Appl. Mech. Eng. **65**, pp. 85–96.
[23] Commend, S., Truty, A., and Zimmermann, T., 2004, "Stabilized Finite Elements Applied to Elastoplasticity: I. Mixed Displacement-Pressure Formulation," Comput. Methods Appl. Mech. Eng. **193**, pp. 3559–3586.
[24] Garikipati, K., and Hughes, T. J. R., 2000, "A Variational Multiscale Approach to Strain Localization-Formulation For Multidimensional Problems," Comput. Methods Appl. Mech. Eng. **188**, pp. 39–60.
[25] Hughes, T. J. R., 1995, "Multiscale Phenomena: Green's Functions, the Dirichlet-to-Neumann Formulation, Subgrid Scale Models, Bubbles and the Origins of Stabilized Methods," Comput. Methods Appl. Mech. Eng. **127**, pp. 387–401.
[26] Hughes, T. J. R., Masud, A, and Harari, I., 1995, "Numerical Assessment of Some Membrane Elements With Drilling Degrees of Freedom," Comput. Struct. **55**(2), pp. 297–314.
[27] Kasper, E. P., and Taylor, R. L., 2000, "A Mixed-Enhanced Strain Method. Part I: Geometrically Linear Problems," Comput. Struct. **75**, pp. 237–250.

Numerical Analysis of Nanotube Based NEMS Devices — Part II: Role of Finite Kinematics, Stretching and Charge Concentrations

Changhong Ke

Horacio D. Espinosa[1]
e-mail: espinosa@northwestern.edu

Nicola Pugno[2]

Department of Mechanical Engineering,
Northwestern University,
Evanston, IL 60208-3111

In this paper a nonlinear analysis of nanotube based nano-electromechanical systems is reported. Assuming continuum mechanics, the complete nonlinear equation of the elastic line of the nanotube is derived and then numerically solved. In particular, we study singly and doubly clamped nanotubes under electrostatic actuation. The analysis emphasizes the importance of nonlinear kinematics effects in the prediction of the pull-in voltage of the device, a key design parameter. Moreover, the nonlinear behavior associated with finite kinematics (i.e., large deformations), neglected in previous studies, as well as charge concentrations at the tip of singly clamped nanotubes, are investigated in detail. We show that nonlinear kinematics results in an important increase in the pull-in voltage of doubly clamped nanotube devices, but that it is negligible in the case of singly clamped devices. Likewise, we demonstrate that charge concentration at the tip of singly clamped devices results in a significant reduction in pull-in voltage. By comparing numerical results to analytical predictions, closed form formulas are verified. These formulas provide a guide on the effect of the various geometrical variables and insight into the design of novel devices. [DOI: 10.1115/1.1985435]

1 Introduction

Nano-electromechanical systems (NEMS) are attracting significant attention because of their properties to enable superior electronic computing and sensing. By exploiting nanoscale effects, NEMS present interesting and unique characteristics. For instance, NEMS based devices can have an extremely high fundamental mechanical oscillation frequency [1–4], while preserving a robust mechanical response [5]. Several NEMS applications have been proposed, such as mass sensors [6], rf resonators [6], field effect transistors [7] and electrometers [8]. Carbon nanotubes (CNTs) have long been considered ideal building blocks for NEMS devices due to their superior electromechanical properties. CNT-based NEMS reported in the literature include nanotweezers [9,10], nonvolatile random access memory devices [11], nanorelays [12], rotational actuators [13] and recently proposed feedback-controlled nanocantilever NEMS devices [14]. All these reported devices can be simply modeled as CNT cantilevers or fixed-fixed CNTs hanging over an infinite conductive substrate. In order to design a functional NEMS device, its electromechanical characteristic should be well quantified in advance. During the past years, a lot of progress has been achieved in regard to the modeling of CNT-based nano-devices [15]. Generally, sufficiently large diameter multiwalled carbon nanotubes (MWNTs), i.e., with diameters of ~20 nm and higher, can be modeled to a good approximation as homogeneous cylindrical beams and perfect conductors, meaning that quantum effects and finite scale charge distribution are negligible at this dimension [16]. Three main types of forces have to be considered in the modeling of the electromechanical characteristic of CNT-based NEMS devices: the elastic forces, the electrostatic forces, and the van der Waals forces arising from the atomic interactions. For the elastic restoring forces, the classical continuum mechanics theory is applicable to CNT devices as demonstrated by molecular dynamics simulation [15]. The electrostatic forces are typically computed by using a capacitance model [17], so that a precise modeling of the capacitance of CNTs is a key issue in their description. We discuss this point, emphasizing the role of charge concentration at the tip of cantilever nanotubes, based on classical electrostatics. For the van der Waals forces, a continuum model based on Lennard-Jones potential theory was employed in the literature [15]. The effect of the van der Waals force on the performance of the CNT devices could be significant in the case of small gaps between nanotube and substrate or for sufficiently long nanotubes [14,15]. Another important but typically omitted effect in the modeling of nano-devices is finite kinematics, which accounts for large displacements. Note that for doubly clamped nanotubes, a dramatic increase in the elastic energy stored in the nanotube is expected as a consequence of the stretching imposed by the rope-like behavior. In this paper, we investigate the electromechanical characteristics of singly and doubly clamped CNT-based NEMS, as illustrated in Fig. 1: a biased MWNT cylinder of length L, placed above an infinite ground plane, at a height H. The inner radius and outer radius of MWNT are R_{int} and R_{ext}, respectively. The applied voltage between nanotube and substrate is V.

The paper is organized as follows. First, an analysis of the charge distribution arising from the electrostatic field is presented. The nonlinear elastic line equation is then derived. This equation is integrated numerically and compared to analytical predictions

[1]To whom correspondence should be addressed.
[2]On leave from the Department of Structural Engineering, Politecnico di Torino, Italy.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, August 23, 2004; final revision; January 26, 2005. Associate Editor. R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

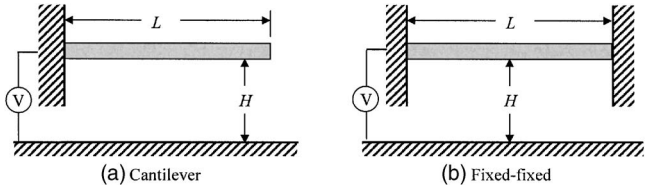


Fig. 1 Schematic of nanotube based NEMS devices

derived elsewhere. We close the paper with conclusions concerning the conditions under which finite kinematics and end charge concentration cannot be neglected for two boundary conditions, singly and doubly clamped devices.

2 Analysis of Charge Distribution

During the past years, significant progress has been made in regard to the computation of the charge distribution along finite-length nanotubes. For conductive nanotubes, essentially classical distribution of charge density with a significant charge concentration at the tube end has been observed [18–20]. Recently a model based on three-dimensional electrostatic calculations has been proposed in [21] to describe the charge distribution along finite length MWNTs cylinder, in particular, the concentrated charges at the tube ends.

Figure 2 shows the charge distribution along the length of a freestanding nanotube of length L, subjected to a bias voltage of 1 V. The contour plot shows the charge density (side view), while the curve shows the charge per unit length along the nanotube. The calculation is performed using the CFD-ACE+ software (a commercial code from CFD Research Corporation based on finite and boundary element methods). The calculations and model reported in [21] are valid as long as the conductive nanotube radius, R_{ext}, is larger than $\sim$10 nm, and the length of the tube, L, is much larger than H and R_{ext}. In these cases quantum effects and size-limit effects in the charge distribution can be considered negligible [21].

The capacitance per unit length along the cantilever nanotube, under moderate deflections, is approximated as [21]

$$C[r(x)] = C_d[r(x)]\{1 + 0.85[(H + R_{ext})^2 R_{ext}]^{1/3}\delta(x - x_{tip})\} = C_d[r(x)]\{1 + f_c\}, \quad (1)$$

where the first term in the bracket accounts for the uniform charge along the side surface of the tube and the second term, f_c, accounts for the concentrated charge at the end of the tube; $x = x_{tip} \neq L$, as a result of the finite kinematics; $\delta(x)$ is the Dirac distribution function. $C_d[r(x)]$ is the distributed capacitance along the side surface per unit length for an infinitely long tube, which is given by [17]

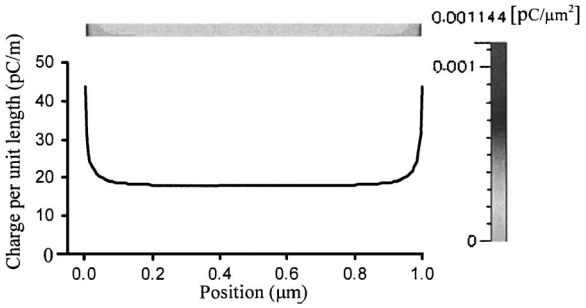


Fig. 2 Charge distribution for a biased nanotube. The device parameters are R_{ext}=9 nm, H=100 nm and L=1 μm.

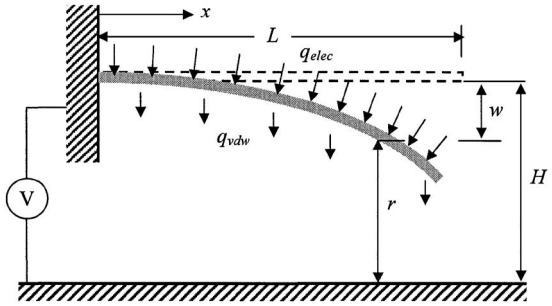


Fig. 3 Schematic of the finite kinematics configuration of a cantilever nanotube device subjected to electrostatic forces and van der Waals forces

$$C_d[r(x)] = \frac{2\pi\varepsilon_0}{a\cosh\left(1 + \frac{r(x)}{R_{ext}}\right)} \quad (2)$$

where r is the distance between the lower fiber of the nanotube and the substrate, and ε_0 is the permittivity of vacuum ($\varepsilon_0 = 8.854 \times 10^{-12}$ C^{2}N^{-1} m^{-2}). Thus, the electrostatic force per unit length of the nanotube is given by differentiation of the energy as follows:

$$q_{elec} = \frac{1}{2}V^2\frac{dC}{dr} = \frac{1}{2}V^2\left(\frac{dC_d}{dr}\right)\{1 + f_c\} = \frac{\pi\varepsilon_0 V^2}{\sqrt{r(r + 2R_{ext})}a\cosh^2\left(1 + \frac{r}{R_{ext}}\right)}(1 + f_c) \quad (3)$$

In the above equation $r(x) = H - w(x)$, with w being the deflection and V the bias voltage.

3 Nonlinear Elastic Line Equations

3.1 Singly Clamped Nanotube. The deflection of a cantilver nanotube under electrostatic force and van der Waals force is shown in Fig. 3. The electrostatic force per unit area remains perpendicular to the outer surface of the nanotube under finite kinematics as imposed by the electrical field. The electrostatic force per unit length of the cylinder is also perpendicular to the cylinder axis. Here we ignore the force applied to the end surface of the cantilever.

Accordingly, if we just consider the bending of the cantilever, the governing equation of the elastic line under finite kinematics is [22]

$$EI\frac{d^2}{dx^2}\left(\frac{\frac{d^2w}{dx^2}}{\left(1 + \left(\frac{dw}{dx}\right)^2\right)^{3/2}}\right) = (q_{vdw} + q_{elec})\sqrt{1 + \left(\frac{dw}{dx}\right)^2}, \quad (4)$$

where E is the Young modulus, $I = \pi(R_{ext}^4 - R_{int}^4)/4$ is the moment of the inertia of the nanotube; q_{vdw} is the van der Waals force (per unit length) between the nanotube and the substrate and can be evaluated by the method reported in [15], assuming the substrate consists of 30 graphite layers. Equation (4) represents the elastic line equations for a nanotube under finite kinematics. As a consequence of the large flexibility of the nanotube, it remains in the elastic regime. Equation (4) clearly represents a more accurate description of the elastic behavior of nanotubes, than the more common equation assuming small displacements, i.e., $dw/dx \ll 1$

$$EI\frac{d^4w}{dx^4} = q_{elec} + q_{vdw} \quad (5)$$

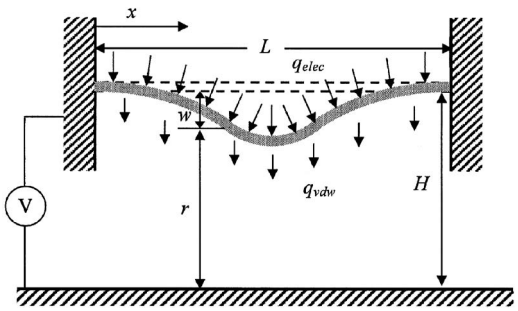


Fig. 4 Schematic of the finite kinematics configuration of a fixed-fixed nanotube device subjected to electrostatic forces and van der Waals forces

3.2 Doubly Clamped Nanotube. For a doubly clamped nanotube, stretching becomes significant as a consequence of the rope-like behavior of a fixed-fixed nanotube subjected to finite kinematics shown in Fig. 4. A tension T in the nanotube has to be introduced, so that the elastic line equation becomes [22]

$$EI\frac{d^2}{dx^2}\left(\frac{\frac{d^2w}{dx^2}}{\left(1+\left(\frac{dw}{dx}\right)^2\right)^{3/2}}\right)-T\left(\frac{\frac{d^2w}{dx^2}}{\left(1+\left(\frac{dw}{dx}\right)^2\right)^{3/2}}\right)=(q_{\mathrm{vdw}}+q_{\mathrm{elec}})\sqrt{1+\left(\frac{dw}{dx}\right)^2} \tag{6}$$

The tension T is related to the axial strain ε, namely [23],

$$T=EA\varepsilon\approx\frac{EA}{2L}\int_0^L\left(\frac{dw}{dx}\right)^2dx \tag{7}$$

where A is the cross-sectional area of the nanotube. Combining Eqs. (6) and (7), we can obtain the governing elastic line equation for the equilibrium position as

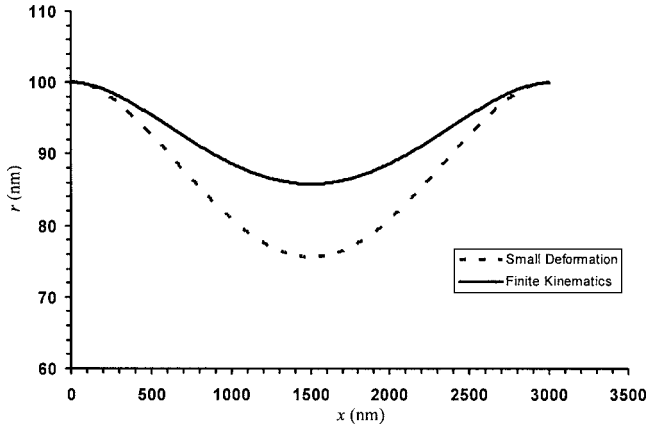


Fig. 5 Elastic line for fixed-fixed nanotube at V=5 V. The solid line is for finite kinematics, the dotted line assumes small deformations.

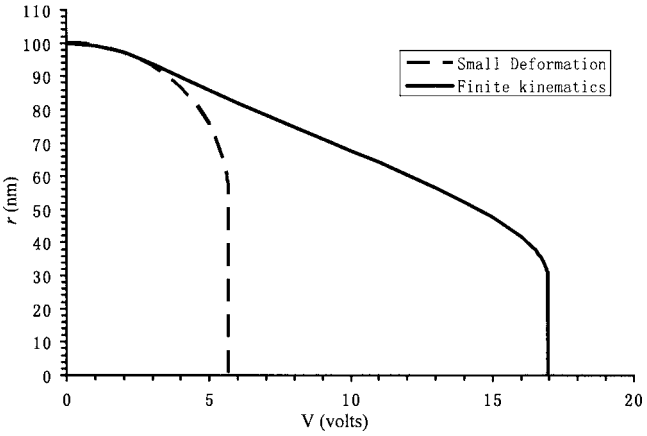


Fig. 6 Electromechanical characteristic (central displacement-voltage curve) for fixed-fixed nanotube device. The dashed line is for small deformation model (pure bending), the solid line is for finite kinematics model (bending plus stretching).

$$EI\frac{d^2}{dx^2}\left(\frac{\frac{d^2w}{dx^2}}{\left(1+\left(\frac{dw}{dx}\right)^2\right)^{3/2}}\right)-\frac{EA}{2L}\int_0^L\left(\frac{dw}{dx}\right)^2dx\left(\frac{\frac{d^2w}{dx^2}}{\left(1+\left(\frac{dw}{dx}\right)^2\right)^{3/2}}\right)=(q_{\mathrm{vdw}}+q_{\mathrm{elec}})\sqrt{1+\left(\frac{dw}{dx}\right)^2} \tag{8}$$

Here the electrostatic force per unit length q_{elec} is given by Eq. (3) setting f_c=0, since the tip charge concentration for this boundary condition does not take place.

If $dw/dx\ll 1$, the classical equation for small displacements, Eq. (5), which neglects the stretching of the nanotube, is again recovered. In addition, for moderately finite kinematics: $(dw/dx)^2\ll 1$, so that Eq. (8) becomes

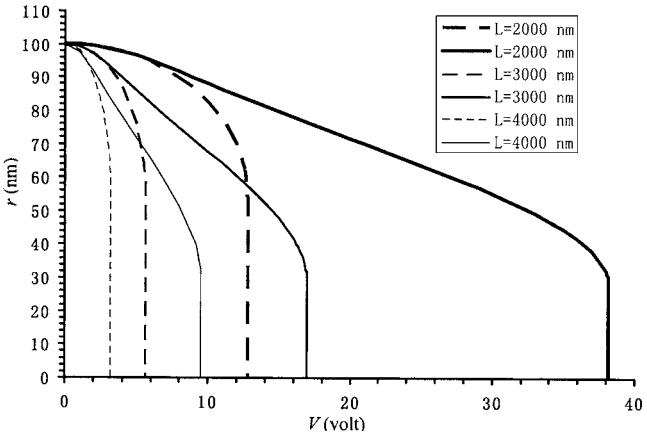


Fig. 7 Electromechanical characteristics (central displacement-voltage curve) for fixed-fixed nanotube devices with different lengths L. The dashed lines are for small deformation model (pure bending), the solid lines are for finite kinematics model (bending plus stretching).

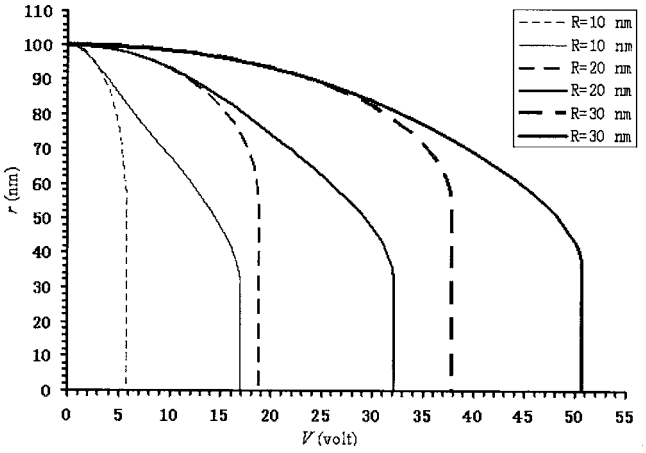


Fig. 8 Electromechanical characteristics (central displacement-voltage curve) for fixed-fixed nanotube devices with different R_{ext} and H=100 nm. The dashed lines are for small deformation model (pure bending), the solid lines are for finite kinematics model (bending plus stretching).

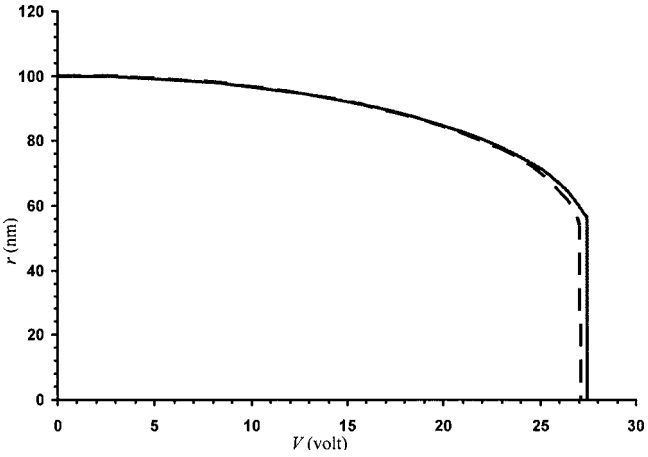


Fig. 10 The effect of finite kinematics on the characteristics of the cantilever nanotube based device (tip displacement versus voltage). The solid lines show the result accounting for finite kinematics, while the dashed line shows the result if finite kinematics is neglected. Both cases consider the concentrated charge at the end of the cantilever nanotube.

$$EI\frac{d^4w}{dx^4} - \frac{EA}{2L}\int_0^L \left(\frac{dw}{dx}\right)^2 dx \frac{d^2w}{dx^2} = q_{\text{elec}} + q_{\text{vdw}} \tag{9}$$

4 Nonlinear Numerical and Theoretical Pull-In Voltage Predictions

Solving numerically the previous nonlinear equations for singly, and doubly clamped nanotube NEMS devices by direct integration (Eqs. (4) and (5)) and finite difference method (Eqs. (5) and (9)), respectively, the *pull-in* voltage corresponding to the nanotube collapsing onto the ground substrate can be predicted. This parameter is key in an optimal device design, corresponding to the transition between open and close states in applications such as nano-switches, nano-tweezers, etc.

The computed results are reported in Figs. 5–9 for doubly clamped and in Figs. 10 and 11 for singly clamped nanotube devices. Unless otherwise specified we have considered R_{ext} =10 nm, R_{int}=0, E=1 TPa, H=100 nm, and for fixed-fixed nanotube L=3000 nm, whereas for cantilever nanotube L =500 nm. In Fig. 5, the elastic lines of the nanotube under a difference in the electrostatic potential of 5 V are reported. The solid line corresponds to the finite kinematics case while the dash line corresponds to the small deformations case. The role of stiffening due to the rope-like behavior is quite remarkable. In Fig. 6 the central deflection of the nanotube as a function of the applied voltage is reported for both cases, i.e., with and without stretching. The two vertical lines correspond to the reaching of the pull-in voltages. In Figs. 7 and 8 similar results considering different lengths and radii are reported for both cases. Again, the role of the stretching has been found not negligible. The effect of H on the pull-in voltage is illustrated in Fig. 9. It is interesting to note that when H changes from 100 to 200 nm the pull-in voltage more than quadruples.

For the cantilevered nanotube, the displacement of the tip as a function of the applied voltage is reported in Fig. 10. In this figure, the effect of the finite kinematics is shown. As expected, the role of the finite kinematics becomes less significant than for the doubly clamped boundary condition. The pull-in voltages (instabilities) correspond to the vertical lines. Both numerical solutions reported in Fig. 10 consider the charge concentration at the tip of

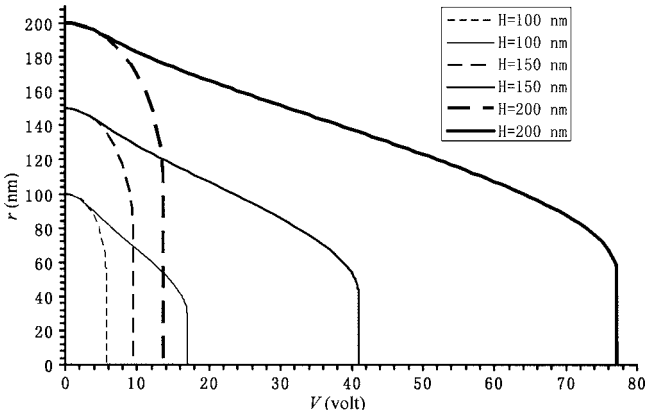


Fig. 9 Electromechanical characteristics (central displacement-voltage curve) for fixed-fixed nanotube devices with different H. The dashed lines are for small deformation model (pure bending), the solid lines are for finite kinematics model (bending plus stretching).

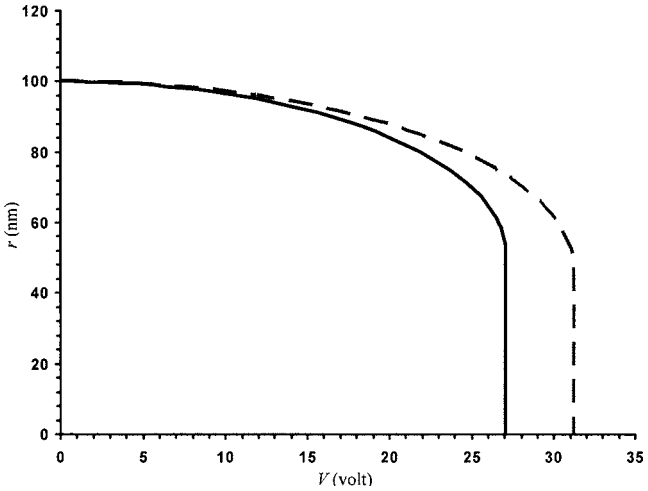


Fig. 11 The effect of the charge concentration on the characteristics of the cantilever nanotube based device (tip displacement versus voltage). The solid line shows the deflection curve with the concentrated charge. The dashed line shows the deflection curve without the concentrated charge. Both curves are from small deflection model.

Table 1 Comparison between pull-in voltages evaluated numerically and theoretically ([24,25]) by Eqs. (10) for doubly (*D*) and singly (*C*) clamped nanotube devices, respectively, $E=1$ TPa, $R_{int}=0$. For cantilever nanotube device the symbol (w) denotes that the effect of charge concentration has been included.

Case	BC	H [nm]	L [nm]	$R=R_{ext}$ [nm]	V_{PI} [V] (theo. linear)	V_{PI} [V] (num. linear)	V_{PI} [V] (theo. non-linear)	V_{PI} [V] (num. non-linear)
1	*D*	100	4000	10	3.20	3.18	9.06	9.54
2	*D*	100	3000	10	5.69	5.66	16.14	16.95
3	*D*	100	2000	10	12.81	12.73	36.31	38.14
4	*D*	150	3000	10	9.45	9.43	38.93	40.92
5	*D*	200	3000	10	13.53	13.52	73.50	77.09
6	*D*	100	3000	20	19.21	18.74	31.57	32.16
7	*D*	100	3000	30	38.57	37.72	51.96	50.63
8	*S*	100	500	10	27.28w)	27.05(w)	27.52(w)	27.41 (w)
9	*S*	100	500	10	27.28(w)	27.05(w)	30.87	31.66

the cantilever nanotube. Figure 11 shows the error in the pull-in voltage, if the charge concentration is ignored. It is inferred that the error can be appreciable.

Recently, analytically derived formulas to compute the pull-in voltage, corresponding to the approximated solutions of the previous nonlinear equations, have been obtained equating to zero the first two derivatives (related to equilibrium and instability) of the free energy of the system [24,25]. Because devices of interest have gaps in the range of 0.1–1 μm, achievable with currently available manufacturing techniques, the effect of van der Waals force is negligible before pull-in happens [15]. Thus, we consider cases in which $q_{vdw}\approx 0$ in the analytical analysis [24] and in the comparison between numerical and analytical predictions. Accordingly, the pull-in voltages for singly (*S*) clamped NEMS devices can be computed as

$$V_{\underset{PI}{S}} \approx k_S\sqrt{1+K_S^{FK}}\frac{H}{L^2}\ln\left(\frac{2H}{R_{ext}}\right)\sqrt{\frac{EI}{\varepsilon_0}}, \tag{10a}$$

$$k_S \approx 0.85, \quad K_S^{FK} \approx \frac{8H^2}{9L^2} \tag{10b}$$

Subscripts *S* refer to single clamped boundary conditions. Superscript *FK* refers to finite kinematics. Moreover, taking into account the additional energy concentrated at the tip of the cantilever nanotube and following the method described in [24], one finds the additional corrective term for the charge concentration at the tip, according to Eq. (1), as

$$V_{\underset{PI}{S}}^{TIP} = \frac{V_{\underset{PI}{S}}}{\sqrt{1+K^{TIP}}}, \quad K^{TIP} \approx \frac{2.55\left[R_{ext}(H+R_{ext})^2\right]^{1/3}}{L} \tag{10c}$$

For doubly (*D*) clamped NEMS devices, the pull-in voltage can be expressed as

$$V_{\underset{PI}{D}} = k_D\sqrt{1+k_D^{FK}}\frac{H+R}{L^2}\ln\left(\frac{2(H+R)}{R}\right)\sqrt{\frac{EI}{\varepsilon_0}} \tag{11a}$$

$$k_D = \sqrt{\frac{1024}{5\pi S'(c_{PI})}\left(\frac{c_{PI}}{H+R}\right)}, \; k_D^{FK} = \frac{128}{3003}\left(\frac{c_{PI}}{\rho}\right)^2 \tag{11b}$$

$$\rho^2 = \frac{I}{A} = \frac{R_{ext}^2+R_{int}^2}{4}$$

$$S(c) = \sum_{i=1}^{\infty}\left(\frac{1}{\left[\ln\left(\frac{2(H+R)}{R}\right)\right]^i}\sum_{j=1}^{\infty}a_{ij}\left(\frac{c}{(H+R)}\right)^j\right) \tag{11c}$$

Subscripts *D* refer to double clamped boundary conditions; c_{PI} is the central deflection of the nanotube at the pull-in, and the $\{a_{ij}\}$ in Eq. (11*c*) are known constants [25].

From the Figs. 6–10 numerically predicted pull-in voltages can be obtained. We compare these results with the theoretical predictions resulting from Eqs. (10) and (11). Note that the comparison does not involve a best fit parameter. The results are reported in Table 1. Columns six and seven in Table 1 compare analytical and numerical pull-in voltage predictions under the assumption of small deformations. The agreement is good (with a maximum discrepancy of 5%). Columns eight and nine in Table 1 compare analytical and numerical pull-in voltage predictions under the assumption of finite kinematics.

5 Conclusions

In this paper a nonlinear analysis for singly and doubly clamped nanotube based nano-electromechanical system (NEMS) devices has been reported. Assuming Continuum Mechanics, the complete nonlinear equation of the elastic line of the nanotube is first derived and then numerically solved for the two considered boundary conditions. The analysis emphasizes the important role of the nonlinear effects in the prediction of the pull-in voltage, a key design parameter corresponding to the switching between the on/off states of the device. Moreover, the nonlinear analysis, neglected in previous studies, shows that finite kinematics resulting in stretching, significantly influences the pull-in voltage of doubly clamped devices. In the case of singly clamped nanotube devices, the finite kinematics effect is negligible but the effect of charge concentration is quite significant. The numerical results agree with the theoretical predictions, Eqs. (10) and (11), for the case of singly and doubly clamped nanotubes. A correction is required to account for tip charge concentration, as described by Eq. (10*c*). In summary, Eqs. (10) and (11) can be used with confidence in the design of novel NEMS. Moreover, these formulas can be employed to gain insight into the effect of device geometry and architecture.

Acknowledgment

A special thanks is due to Prof. N. Moldovan for many key comments during the research and manuscript writing phases. The authors also acknowledge the support from the FAA through Award No. DTFA03-01-C-00031 and the NSF through Award No. CMS-0120866. We would like to express our appreciation to Dr.

J. Newcomb and Dr. J. Larsen-Base for supporting this work. Work was also supported in part by the Nanoscale Science and Engineering Initiative of the National Science Foundation under NSF Award No. EEC-0118025.

References

[1] Yang, Y. T., Ekinci, K. L., Huang, X. M. H., Schiavone, L. M., Roukes, M. L., Zorman, C. A., and Mehregany, M., 2001, "Monocrystalline Silicon Carbide Nanoelectromechanical Systems," Appl. Phys. Lett., **78**, pp. 162–164.

[2] Cleland, A. N., and Roukes, M. L., 1996, "Fabrication of High Frequency Nanometer Scale Mechanical Resonators from Bulk Si Crystals," Appl. Phys. Lett., **69**, pp. 2653–2655.

[3] Erbe, A., Blick, R. H., Tilke, A., Kriele, A., and Kotthaus, P., 1998, "A Mechanical Flexible Tunneling Contact Operating at Radio Frequency," Appl. Phys. Lett., **73**, pp. 3751–3753.

[4] Huang, X. M. H., Zorman, C. A., Mehregany, M., and Roukes, M. L., 2003, "Nanodevice Motion at Microwave Frequencies," Nature (London), **421**, p. 496.

[5] Roukes, M. L., 2000, "Nanoelectromechanical System," Technical Digest of the 2000 Solid-State Sensor and Actuator Workshop.

[6] Abadal, G., Davis, Z. J., Helbo, B., Borrise, X., Ruiz, R., Boisen, A., Campabadal, F., Esteve, J., Figueras, E., Perez-Murano, F., and Barniol, N., 2001, "Electromechanical Model of a Resonating Nano-Cantilever-Based Sensor for High-Resolution and High-Sensitivity Mass Detection," Nanotechnology, **12**, pp. 100–104.

[7] Martel, R., Schmidt, T., Shea, H. R., Hertel, T., and Avouris, Ph., 1998, "Single- and Multi-Wall Carbon Nanotube Field-Effect Transistors," Appl. Phys. Lett., **73**, pp. 2447–2449.

[8] Cleland, A. N., and Roukes, M. L., 1998, "A Nanometer-Scale Mechanical Electrometer," Nature (London), **392**, pp. 160–162.

[9] Akita, S., Nakayama, Y., Mizooka, S., Takano, Y., Okawa, T., Miyatake, Y., Yamanaka, S., Tsuji, M., and Nosaka, T., 2001, "Nanotweezers Consisting of Carbon Nanotubes Operating in an Atomic Force Microscope," Appl. Phys. Lett., **79**, pp. 1691–1693.

[10] Kim, P., and Lieber, C. M., 1999, "Nanotube Nanotweezers," Science, **126**, pp. 2148–2150.

[11] Rueckes, T., Kim, K., Joslevich, E., Tseng, G. Y., Cheung, C., and Lieber, C. M., 2000, "Carbon Nanotube-Based Nonvolatile Random Access Memory for Molecular Computing," Science, **289**, pp. 94–97.

[12] Kinaret, J., Nord, T., and Viefers, S., 2003, "A Carbon-Nanotube-Based Nanorelay," Appl. Phys. Lett. **82**, pp. 1287–1289.

[13] Fennlmore, A. M., Yuzvlnsky, T. D., Han, W. Q., Fuhrer, M. S., Cummings, J., and Zettl, A., 2003, "Rotational Actuator Based on Carbon Nanotubes," Nature (London), **424**, pp. 408–410.

[14] Ke, C.-H., and Espinosa, H. D., 2004, "Feedback Controlled Nanocantilever Device," Appl. Phys. Lett. **85**, pp. 681–683.

[15] Dequesnes, M., Rotkin, S. V., and Aluru, N. R. 2002, "Calculation of Pull-in Voltage for Carbon-Nanotube-Based Nanoelectromechanical Switches," Nanotechnology, **13**, pp. 120–131.

[16] Krcmar, M., Saslow, W. M., and Zangwill, A., 2003, "Electrostatic of Conducting Nanocylinder," J. Appl. Phys., **93**, pp. 3495–3500.

[17] Hayt, W., and Buck, J., 2001, *Engineering Electromagnetics*, 6th ed., McGraw–Hill, New York.

[18] Keblinski, P., Nayak, S. K., Zapol, P., and Ajayan, P. M., 2002, "Charge Distribution and Stability of Charged Carbon Nanotube," Phys. Rev. Lett., **89**, p. 255503.

[19] Bulashevich, K. A., and Rotkin, S. V., 2002, "Nanotube Devices: A Microscopic Model," JETP Lett., **175**, pp. 205–209.

[20] Rotkin, S. V., Shrivastava, V., Bulashevich, K. A., and Aluru, N. R., 2002, "Atomic Capacitance of a Nanotube Electrostatic Device," Int. J. Nanosci., **1**, pp. 337–346.

[21] Ke, C.-H., and Espinosa, H. D., 2005, "Numerical Analysis of Nanotube Based NEMS Devices—Part I: Electrostatic Charge Distribution on Multiwalled Nanotubes," ASME J. Appl. Mech., **72**, pp. 721–725.

[22] Fertis, D. G., 1999, *Nonlinear Mechanics*, 2nd ed. CRC, Boca Raton, FL.

[23] Sathyamoorthy, M., 1998, *Nonlinear Analysis of Structures*, CRC, Boca Raton, FL.

[24] Ke, C-H., Pugno, N., Peng, B., and Espinosa, H. D., 2005, "Experiments and Modelling of Carbon-Nanotube Based NEMS Devices," J. Mech. Phys. Solids, **53**, pp. 1314–1333.

[25] Pugno, N., Ke, C.-H., and Espinosa, H. D., 2005, ″Analysis of Doubly-Clamped Nanotube Devices Under Large Displacements,″ ASME J. Appl. Mech., **72**, pp. 445–449.

Flaw Tolerance in a Thin Strip Under Tension

Huajian Gao

Shaohua Chen

Max Planck Institute for Metals Research,
Heisenbergstrasse 3,
D-70569 Stuttgart, Germany

Recent studies on hard and tough biological materials have led to a concept called flaw tolerance which is defined as a state of material in which pre-existing cracks do not propagate even as the material is stretched to failure near its limiting strength. In this process, the material around the crack fails not by crack propagation, but by uniform rupture at the limiting strength. At the failure point, the classical singular stress field is replaced by a uniform stress distribution with no stress concentration near the crack tip. This concept provides an important analogy between the known phenomena and concepts in fracture mechanics, such as notch insensitivity, fracture size effects and large scale yielding or bridging, and new studies on failure mechanisms in nanostructures and biological systems. In this paper, we discuss the essential concept for the model problem of an interior center crack and two symmetric edge cracks in a thin strip under tension. A simple analysis based on the Griffith model and the Dugdale-Barenblatt model is used to show that flaw tolerance is achieved when the dimensionless number $\Lambda_{\rm ft}=\Gamma E/(S^2H)$ *is on the order of 1, where* Γ *is the fracture energy, E is the Young's modulus, S is the strength, and H is the characteristic size of the material. The concept of flaw tolerance emphasizes the capability of a material to tolerate cracklike flaws of all sizes.*
[DOI: 10.1115/1.1988348]

1 Introduction

Hard and tough biological materials such as bone, teeth and wood must be able to survive cracklike flaws of many size scales in order to successfully perform their designated mechanical and biological functions. The self-sensing, self-adapting, and self-repairing capabilities of bone require not only a dynamical network of blood vessels for supply of nutrient, but also constant removal and replacement of old and damaged materials with fresh and healthy materials. The fact that all these processes should occur while an animal is conducting its normal activities indicates that biological materials must be designed to tolerate cracklike flaws of many size scales. We have found it useful to adopt flaw tolerance as a basic principle in understanding mechanics of biological systems [1–7]. On the one hand, the concept of flaw tolerance can be related to the known phenomena and concepts such as notch insensitivity, fracture size effects and large scale yielding or bridging in fracture mechanics [8–24]; on the other hand, it can be related to the theory of evolution which states that survivability (in this case against mechanical flaws) is a key to propagation of animal species. In the state of flaw tolerance, pre-existing cracklike flaws do not propagate and do not participate in the failure process. This view has been a central theme in our recent studies on mechanics and mechanical properties of protein-mineral nanocomposite structures of bone and bonelike materials [1–5] as well as mechanics of hierarchical adhesion systems of gecko [6,7]. In these biological systems, it has been shown that, as the characteristic size of the critical structural link is reduced to below a critical size, a class of elastic solutions emerge with the interesting feature of uniform stress distribution even in the vicinity of a crack. The idea of flaw tolerance has been used to explain the nanometer sizes of mineral crystals in bone [1–5] and of the adhesive nanoprotrusions of gecko [6,7]. In a flaw tolerant biological system, failure occurs not by propagation of a pre-existing crack, but by uniform rupture at the limiting strength of the material.

For brittle materials such as various biominerals, the critical length scale associated with flaw tolerance was found to be proportional to $\gamma E/\sigma_{\rm th}^2$ [1–7], where E is the Young's modulus, $\sigma_{\rm th}$ the theoretical strength of the material, and γ the fracture surface energy. Similar length scales with various physical interpretations have appeared in the classical Dugdale-Barenblatt model [8,9] and large scale yielding or bridging models applied to mechanics of earthquake rupture [10], notch insensitivity [11,12], fracture size effects in concrete [13–17] and in sea ice [18], and fiber bridging in composites [19–21]. Such length scale also appeared in dynamic failure phenomena such as the minimum fragment sizes in comminuting of glassy materials [22,23] and in dynamic fragmentation [24]. The concept of flaw tolerance can thus be closely related to various established phenomena and concepts in fracture mechanics. In the present paper, we discuss flaw tolerance and its related concepts for the simple model problem of a thin strip under tension. The Griffith model and the Dugdale-Barenblatt model are used to show that there exists a critical strip width below which cracks of all sizes are tolerated. This critical size is identified as a unique material parameter independent of the flaw size. Below the critical size, the strip fails not by crack growth but by uniform rupture of the material. It is interesting to compare this concept to that of large scale yielding or bridging; in the latter case, one compares the size of a crack with that of the yield or bridging zone near the crack tip; if these sizes are comparable, the fracture strength is predicted to approach the yield or limiting strength of the material. In the concept of flaw tolerance, since it is not the size of a flaw but the size of a material which is linked to a critical length scale, we can introduce the following dimensionless number:

$$\Lambda_{\rm ft}=\Gamma E/(S^2H) \tag{1}$$

where Γ is the fracture energy, E is the Young's modulus, S is the limiting strength of the material, and H is the characteristic size of the material or structure. We illustrate in the following that, when the flaw tolerance number $\Lambda_{\rm ft}$ is on the order of 1, the material has the intrinsic capability to tolerate cracklike flaws of all sizes. Alternatively, we can also define a critical length

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, September 9, 2004; final revision, January 15, 2005. Associate Editor: R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

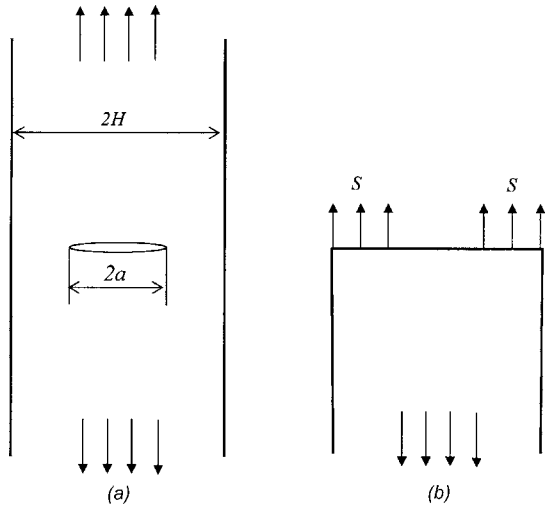


Fig. 1 Flaw tolerance in a center cracked strip under tension. (*a*) The crack configuration and (*b*) the state of flaw tolerance. The stress around the crack uniformly reaches the theoretical strength of the material with no stress concentration near the crack tip.

$$\ell_{\text{ft}} = \frac{\Gamma E}{S^2}. \tag{2}$$

The state of flaw tolerance is achieved when the characteristic dimension of a structure falls below this critical length.

2 The Basic Definition

Consider a strip of width $2H$ containing an interior center crack of size $2a$ [Fig. 1(*a*)] or two symmetric edge cracks (The case of a single edge crack is complicated by a net bending moment on the material ahead of the crack and is not treated here.) [Fig. 2(*a*)] each of length a. In the absence of cracklike flaws, the material is assumed to fail by uniform rupture at the limiting strength S. Our attention will be focused on whether there exists a solution in which the normal traction outside the crack region is uniform and equal to the limiting strength S, as depicted for the case of the center crack in Fig. 1(*b*) and for the double edge crack in Fig. 2(*b*). The existence of such solutions would mean that the crack will not propagate even as the surrounding material uniformly reaches the limiting strength. We introduce the following definition:

Definition: The state of flaw tolerance in a material is defined as such that a pre-existing crack, no matter its size and orientation, does not propagate even as the applied stress approaches the limiting strength of the material. The material ahead of the crack fails not by crack propagation, but by uniform rupture at the limiting strength of the material.

This view has been adopted in our recent studies on mechanics and mechanical properties of bonelike biological materials [1–5] and hierarchical adhesion structures of the gecko [6,7]. Here we illustrate the basic concept for the simple strip geometries shown in Figs. 1 and 2. Considering a unit thickness in the out-of-plane direction, if the thin strip has achieved flaw tolerance, the failure load should be

$$F_{\text{ft}} = 2(H-a)S \tag{3}$$

and the strength of the flaw tolerant strip is simply

$$S_{\text{ft}} = \frac{H-a}{H}S = (1-\phi)S \tag{4}$$

where $\phi = a/H$ measures the relative size of the crack and can be generally interpreted as an effective area density of crack over the cross-section of the strip. The above equations are analogous to continuum damage mechanics [25] where ϕ corresponds to the damage parameter. It has been recently established (see e.g., [16]) that simple damage mechanics descriptions like Eq. (4) is directly applicable only for structure sizes less than a critical length scale similar to Eq. (2) and, because of this, continuum damage mechanics needs to be combined with nonlocal modeling in order to capture damage localization and size effects for sizes exceeding the size of the fracture process zone or the nonlocal averaging volume [16].

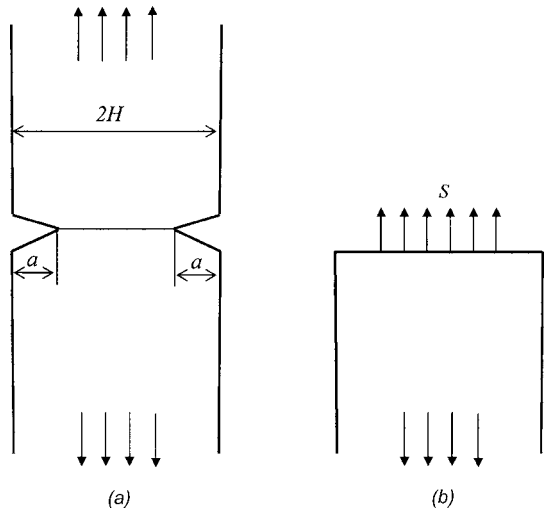


Fig. 2 Flaw tolerance of a double edge cracked strip under tension. (*a*) The double edge crack configuration and (*b*) the corresponding flaw tolerance solution.

3 Flaw Tolerant Solution Based on the Griffith Model

In linear elastic fracture mechanics, the stress intensity factor for the cracked strips shown in Figs. 1 and 2 can be expressed as [26]

$$K_I = \sigma\sqrt{\pi a}F(\phi) \tag{5}$$

where $\phi = a/H$ and the function $F(\phi)$ can be approximated from a periodic crack solution [26] as

$$F(\phi) = \sqrt{\frac{2}{\pi\phi}\tan\frac{\pi\phi}{2}} \tag{6}$$

or by the following empirical formulas [26]:

$$F(\phi) = (1 - 0.025\phi^2 + 0.06\phi^4)\sqrt{\sec\frac{\pi\phi}{2}} \quad \text{center crack} \tag{7}$$

$$F(\phi) = \left(1 + 0.122\cos^4\frac{\pi\phi}{2}\right)\sqrt{\frac{2}{\pi\phi}\tan\frac{\pi\phi}{2}} \quad \text{double edge crack} \tag{8}$$

These empirical formulas are accurate within 0.5% error for all values of ϕ. In the Griffith-Irwin fracture mechanics, the condition for a crack to propagate is

$$G = \frac{(1-\nu^2)}{E}K_I^2 \geq \Gamma. \tag{9}$$

On the other hand, the condition for existence of a flaw tolerant solution is that a pre-existing crack, regardless of its size, does not propagate. Therefore, in the Griffith model, the flaw tolerance condition corresponds to

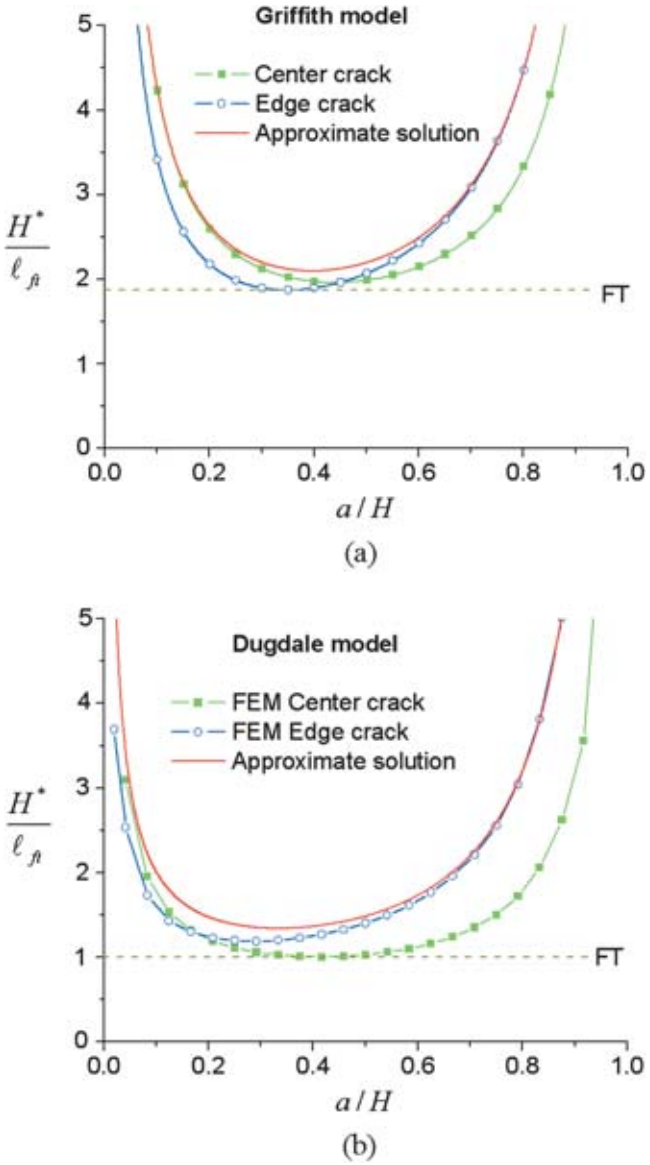


Fig. 3 The normalized critical size for a strip of width 2*H* to tolerate a crack of length 2*a*. The minimum value of such curves corresponds to the critical size for the strip to tolerate cracks of all sizes. Calculations based on (*a*) the Griffith model and (*b*) the Dugdale model.

$$G(0 \leqslant \phi < 1) = \left.\frac{(1-\nu^2)}{E} K_I^2\right|_{0 \leqslant \phi < 1} < \Gamma. \tag{10}$$

Since a flaw tolerant strip has the optimal strength $S_{ft}=(1-\phi)S$, inserting $\sigma=S_{ft}$ into Eq. (5) and then into Eq. (10) yields

$$H < H_{cr} = \min_{0 \leqslant \phi < 1} H^*(\phi), \quad H^*(\phi) = \frac{1}{\pi(1-\nu^2)\phi(1-\phi)^2 F^2(\phi)} \ell_{ft} \tag{11}$$

where $H^*(\phi)$ is the critical size for the strip to tolerate a crack of given size $a=\phi H$ and H_{cr} is the critical size to tolerate cracks of all sizes. Taking $\nu=0.3$, the results of $H^*(\phi)/\ell_{ft}$ based on the Griffith model are plotted in Fig. 3(*a*) for both center and edge cracks. Comparing both types of cracks, $H^*(\phi)/\ell_{ft}$ has the minimum value of about 1.87. This minimum is indicated by a horizontal dashed line labeled "FT" (flaw tolerance). Any strip with width below this line is predicted to tolerate center and edge cracks of all sizes. Therefore, the critical size for flaw tolerance based on the Griffith model is estimated to be

$$H_{cr} = 1.87\ell_{ft} = 1.87\frac{\Gamma E}{S^2} \tag{12}$$

Therefore, the flaw tolerance condition based on the Griffith model is thus

$$H \leqslant 1.87\ell_{ft} \text{ or } \Lambda_{ft} \geqslant 0.53 \tag{13}$$

A similar problem has been solved previously by Carpinteri [17] with respect to a tension collapse that precedes brittle crack propagation in a strip for any relative crack length. Carpinteri [17] defined a "brittleness number" as $s=K_{IC}/\sigma_u\sqrt{h}$ where K_{IC} denotes the fracture toughness, σ_u the strength, and h the size of the strip. Carpinteri found that if this "brittleness number" is larger than 0.54, tension collapse occurs before brittle crack propagation for any crack length. The flaw tolerance number Λ_{ft} defined in Eq. (1) is twice the square of Carpinteri's brittleness number; $s=0.54$ corresponds to $\Lambda_{ft}=0.58$.

4 Flaw Tolerant Solution Based on the Dugdale Model

The Griffith model is a highly idealized model of brittle fracture which neglected the details of atomic interaction laws. At the next level of sophistication, the concept of flaw tolerance can be illustrated by the Dugdale-Barenblatt model [8,9] with the following cohesive law

$$\sigma(\delta) = \begin{cases} S & \delta \leqslant \delta_0 \\ 0 & \delta > \delta_0 \end{cases} \tag{14}$$

where $\sigma=\sigma(\delta)$ is the normal traction, δ_0 is the effective range of cohesive interaction, and δ the opening displacement in the plane of the crack [Fig. 4(*a*)]. In the Dugdale model [8], the condition of flaw tolerance is

$$\delta_{tip}(0 \leqslant \phi < 1) \leqslant \delta_0 \tag{15}$$

where δ_{tip} denotes the crack-tip opening displacement. The basic length scale for flaw tolerance can also be expressed in terms of δ_0 as

$$\ell_{ft} = \frac{\delta_0 E}{S} \tag{16}$$

where we have used the relation $\Gamma=S\delta_0$ in the Dugdale model.

If the flaw tolerance solution exists, there will be a uniform distribution of normal stress S outside the crack region, regardless of the crack size [Fig. 4(*a*)]. This problem is equivalent to a semi-infinite strip subject to a uniform stress equal to S on the part of the top surface outside the crack region, as shown in Fig. 4(*b*). This stress causes a nonuniform upward normal displacement along the top surface. The crack-tip opening displacement can be calculated from the linear elastic problem of Fig. 4(*b*) as

$$\delta_{tip} = -2[u_y(x=a, y=0) - u_y(x=H, y=0)] = \frac{4(1-\nu^2)SH}{\pi E} f(\phi) \tag{17}$$

where $f(\phi)$ ($\phi=a/H$) is a dimensionless function which cannot be determined in closed form. An approximation by the periodic crack solution shown in Fig. 5 is derived in the Appendix as

$$f(\phi) = \int_{\phi}^{1} \ln \frac{\sin \pi(\xi+\phi)/2}{\sin \pi(\xi-\phi)/2} d\xi. \tag{18}$$

The flaw tolerance condition (15) can then be written in the form

$$H \leqslant H_{cr} = \min_{0 \leqslant \phi < 1} H^*(\phi), \quad H^*(\phi) = \frac{\pi}{4(1-\nu^2) f(\phi)} \ell_{ft}. \tag{19}$$

Assuming $\nu=0.3$, the function $H^*(\phi)/\ell_{ft}$ based on the Dugdale

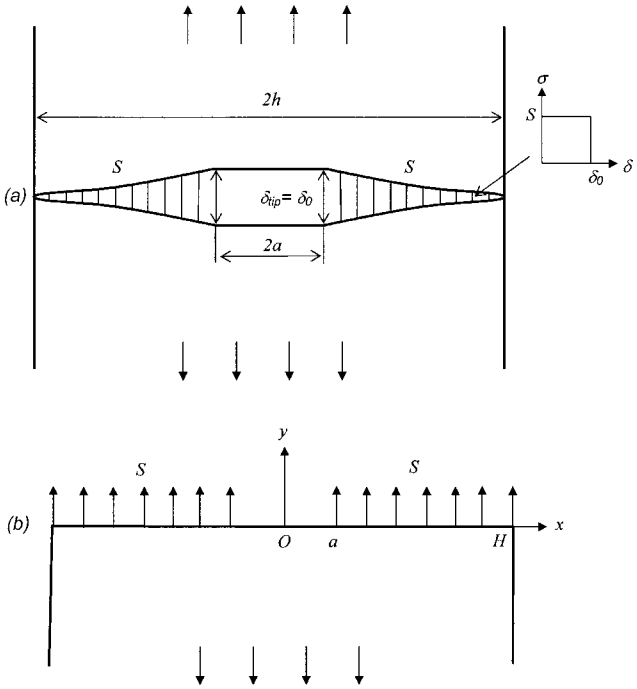


Fig. 4 The flaw tolerance solution based on the Dugdale model of the center cracked strip. (*a*) A Dugdale interaction law is assumed in the plane of the crack. The condition of flaw tolerance is equivalent to requiring δ_{tip} not to exceed δ_0 for any crack size *a*. In this case, the opening displacement in the plane of the crack outside the crack region should lie within the range of cohesive interaction δ_0. (*b*) The reduced elastic problem of a semi-infinite strip subjected to a uniform stress *S* over the top surface outside the crack region.

model is plotted in Fig. 3(*b*) using the approximate solution for $f(\phi)$ in Eq. (18) and numerical calculations from a finite element analysis.

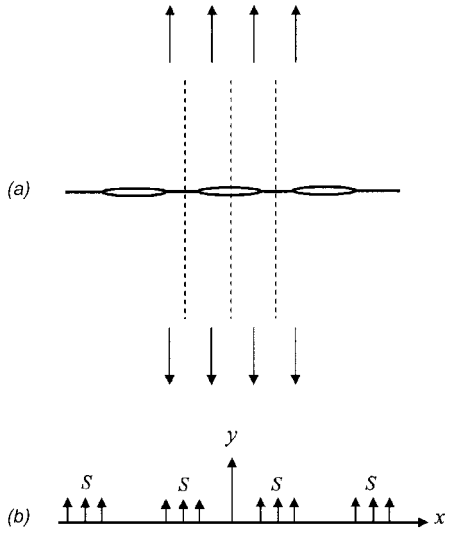


Fig. 5 A periodic crack problem used to derive an approximate solution for the strip cracks in Figs. 1 and 2. Symmetry conditions along the dotted lines suggest that the periodic crack problem can be used as an approximate solution to the strip crack problems. In the mode III case, the period crack solution becomes the exact solutions for the strip cracks. (*a*) The period crack configuration and (*b*) the flaw tolerant solution.

Figure 3(*b*) indicates that, for both types of cracks, $H^*(\phi)/\ell_{ft}$ has a minimum value around $0.32\pi \approx 1$ for a center crack with crack size close to one half of the strip width. Strips thinner than this critical value are predicted to tolerate center and edge cracks of all sizes. Therefore, the Dugdale-model predicts that the critical size for the strip to achieve flaw tolerance is

$$H_{cr} = \ell_{ft} = \frac{\Gamma E}{S^2} = \frac{\delta_0 E}{S}. \tag{20}$$

The flaw tolerance condition can thus be expressed as

$$H \leqslant \ell_{ft} \text{ or } \Lambda_{ft} \geqslant 1 \tag{21}$$

Note that the prediction based on the Dugdale model is within a factor of 2 of that based on the Griffith model. Equation (21) is very convenient to use.

While the Dugdale model represents an improvement upon the Griffith model, the cohesive law in Eq. (14) is still unrealistic and should be regarded as a highly idealized simplification of cohesive interactions. Despite this idealization, the Dugdale model has been successfully used to describe molecular forces in adhesive contact of elastic spheres [27]. More rigorous descriptions of atomic interactions can be achieved by atomistic modeling without additional gain in concept.

5 Discussion

5.1 Nanoscale Flaw Tolerance in Brittle Materials. We have used the Griffith model and the Dugdale-Barenblatt model of a cracked thin strip as a benchmark problem to illustrate the concept of flaw tolerance applied to mechanics of biological systems [1–7]. This concept is closely related to the known phenomena and concepts such as notch insensitivity, fracture size effects, and large scale yielding or bridging in fracture mechanics. The focus of flaw tolerance is on the capability of materials (such as biominerals) to tolerate cracklike flaws below a critical size. In the state of flaw tolerance, the material ahead of a crack fails not by crack propagation, but by uniform rupture at the limiting strength of the material. Obviously, for a cracked material, the optimal strength corresponds to a uniform distribution of failure stress spread over all remaining cohesive bonds. This optimal state can be reached via size reduction. We have illustrated via both Griffith and Dugdale models that there exists a critical width for a strip below which center and edge cracks of all sizes are tolerated. This critical size is a material constant independent of the flaw size.

For brittle materials, the fracture energy is twice the surface energy ($\Gamma = 2\gamma$) and the limiting strength is the theoretical strength ($S = \sigma_{th}$). A previous estimate of critical size of flaw tolerance H_{cr} based on a platelet containing a thumbnail crack with depth equal to half of the plate thickness is [1]

$$H_{cr} = \pi \frac{\gamma E}{\sigma_{th}^2} = \frac{\pi \ell_{ft}}{2}. \tag{22}$$

This result, which is similar to Eqs. (12) and (20), has led to the postulate that the nanometer size of mineral crystals in bone is selected to ensure optimum fracture strength and maximum tolerance of flaws (for robustness) [1]. For biominerals, taking $\gamma = 1$ J/m^2, $E = 100$ GPa, and $\sigma_{th} = E/30$ leads to an estimate of ℓ_{ft} around 18 nm and $H_{cr} = \pi \ell_{ft}/2$ around 30 nm. The Young's modulus of various brittle materials can vary, depending on the atomic structure and the purity of the material, in quite a wide range between a few gigapascals to around a thousand gigapascals. For example, the Young's modulus of $CaCO_3$ can be as low as 50 GPa and that of diamond can reach as high as 1200 GPa. Typical estimates of the theoretical strength can range between 1%–10% of the Young's modulus. If we take $\gamma = 1$ J/m^2, $E = 50$–1200 GPa, and $\sigma_{th} = (1\% - 10\%)E$, the characteristic length ℓ_{ft} of flaw tolerance can be estimated to vary in the range

$$\ell_{\text{ft}} \approx 2\text{Å} - 400\ \text{nm}. \tag{23}$$

Therefore, the critical length of flaw tolerance can vary from near atomic scale for materials like diamond to a few hundred nanometers for biominerals. Recent studies on mechanical strength of single wall carbon nanotubes and SiC and beta-Si_3N_4 nanowhiskers (see [28] and references therein) seem to show that nanodefects in these nearly perfect nanostructures can strongly affect their strength. This may be understood from the point of view that the characteristic length of flaw tolerance for these materials is on the atomic scale with practically vanishing range of flaw tolerance so that the resulting nanostructures remain sensitive to the presence of nanodefects. For biological materials, the inorganic components in bone, teeth and sea shells have significantly lower Young's modulus and hence can become insensitive to cracklike flaws at relatively larger size scales (a few to a few hundred nanometers). In other words, materials with nearly perfect atomic structures and high Young's modulus may remain sensitive to cracklike flaws even down to atomic scales while materials with lower Young's modulus and less perfect lattice structures or less pure chemical constituents are likely to have saturated their limiting strength at tens or hundreds of nanometers size scales.

5.2 Flaw Tolerance in Hierarchical Materials. The concept of flaw tolerance suggests that sufficiently small bodies tolerate cracklike flaws. Although it is difficult to devise a rigorous mathematical proof for arbitrary crack and specimen geometries, a simple argument based on dimensional analysis would show that the basic concept is in fact general. Consider a small scale material containing a variety of cracklike flaws. For cracks with sizes smaller than the overall dimension of the material, the dominant length is the crack size a. On the other hand, for those flaws with size approaching the material dimension H, the dominant length becomes H-a. Therefore, the "largest flaw in the material" will be around half the material dimension, a fact which can also be noted from Fig. 3. Application of Griffith fracture criterion would show that the fracture strength is

$$\sigma^f \propto \sqrt{\Gamma E/H}. \tag{24}$$

Comparing this prediction to the limiting strength S immediately shows that the largest flaw in the material will be tolerated as soon as the material is reduced to a critical size on the order of $\ell_{\text{ft}} = \Gamma E/S^2$ defined in Eq. (2). The growth of all flaws will be suppressed once the largest flaw is tolerated. This viewpoint was implicitly adopted in [1] when estimating the critical length of flaw tolerance for mineral platelets in bone.

The concept of flaw tolerance implies that the conventional engineering concept of stress concentration near flaws becomes invalid when the characteristic dimension of a material or a structure is reduced to below a critical length. When this occurs, the concept of flaws can assume a totally different meaning from that used in conventional engineering. This has been shown to have special meaning for the inorganic components in bone [1–5] and the adhesive protrusions in gecko [6,7]. How can the concept of flaw tolerance be used to design materials with flaw tolerance property at macroscopic scales? In biological materials or systems, we believe that the size limit of flaw tolerance is increased via hierarchical structure design; for example, bone has seven levels of structural hierarchies [29]. Another fact known to civil engineers is that a mesh of steel bars can be embedded in concrete to push the size effects to larger structures [16]. In metals, the plastic yield strength itself exhibit strong size-dependence (e.g., [30]) near micrometer length scales. In cases involving fiber-reinforced or other forms of composites, the limiting strength of the composite will depend on how the length scale of interest compares with the characteristic size of individual fibers or other reinforcing elements. It will be of particular interest to develop nanoengineered hierarchical materials with flaw tolerance property at all hierarchical levels, all the way up to macroscopic length scales. The essence of flaw tolerance in hierarchical materials is to render the dimensionless number $\Lambda_{\text{ft}} = \Gamma E/(S^2 H)$ independent of H via hierarchical design. This may be achieved by one or more of the following strategies as H increases: (a) increase Γ; (b) increase E; (c) decrease S. As the size of metallic materials increase, inelastic deformation mechanisms such as dislocations and creep tend to become more active while yield strength tends to decrease, somewhat corresponding to strategies (a) and (c). In certain cases, these strategies can be realized by designing with nonhomogeneous or graded materials. Biological materials may provide more hints on how to design flaw tolerant hierarchical materials for macroscopic applications. The study of how biological systems survive cracklike flaws should be of interest not only to the mechanics community but also to a variety of other disciplines including materials science, biology, biological materials, evolution, nanotechnology, and biotechnology.

Acknowledgment

Support of this work has been provided by the Max Planck Society, the National Science Foundation of China and the Chang Jiang Scholar Program through the Ministry of Education of China. Helpful discussions with Professor Jan D. Achenbach of Northwestern University are gratefully acknowledged.

Appendix: An Approximate Solution for the Dugdale Model of Flaw Tolerance in a Cracked Strip

The nondimensional function $f(\phi)$ in Eq. (17) cannot be determined in closed form. An approximate solution can be obtained based on the periodic crack problem depicted in Fig. 5(a). According to the theory of elasticity [31], the normal displacement along the surface of a plane-strain half space subjected to an arbitrary distribution of normal traction $p(\xi)$ can be expressed as

$$\frac{\partial u_y(x)}{\partial x} = \frac{2(1-\nu^2)}{\pi E}\int_{-\infty}^{\infty} \frac{p(\xi)}{\xi - x} d\xi. \tag{A1}$$

When the traction distribution is periodic with period $2H$, Eq. (A1) can be simplified as

$$\frac{\partial u_y(x)}{\partial x} = -\frac{(1-\nu^2)}{EH}\int_0^{2H} p(\xi)\text{ctan}\,\frac{\pi(\xi - x)}{2H} d\xi. \tag{A2}$$

For the piecewise uniform distribution of normal traction shown in Fig. 5(b), Eq. (A2) can be integrated as

$$\frac{\partial u_y(x)}{\partial x} = -\frac{2(1-\nu^2)S}{\pi E}\ln\left|\frac{\sin(\pi(x+a)/2H)}{\sin(\pi(x-a)/2H)}\right|. \tag{A3}$$

In the Dugdale model of the periodic crack, the crack-tip opening displacement can be calculated from Eq. (A3) as

$$\delta_{\text{tip}} = 2\int_H^a \frac{\partial u_y(x)}{\partial x} dx = \frac{4(1-\nu^2)SH}{\pi E} f\left(\frac{a}{H}\right) \tag{A4}$$

where

$$f(\phi) = \int_\phi^1 \ln\frac{\sin\,\pi(\xi+\phi)/2}{\sin\,\pi(\xi-\phi)/2} d\xi. \tag{A5}$$

Symmetry considerations will show that this result can be used as an approximate solution to the strip crack problems shown in Figs. 1 and 2. If the plane strain modulus $(1-\nu^2)/E$ is replaced by $(1+\nu)/E$, the solution of Eqs. (A4) and (A5) also applies to the periodic crack under mode III antiplane loading. In the mode III case, Eq. (A4) with $(1-\nu^2)/E$ replaced by $(1+\nu)/E$ is also the exact solution for the strip crack problems depicted in Figs. 1 and 2.

References

[1] Gao, H., Ji, B., Jäger, I. L., Arzt, E., and Fratzl, P., 2003, "Materials Become

Insensitive to Flaws at Nanoscale: Lessons from Nature," Proc. Natl. Acad. Sci. U.S.A., **100**, pp. 5597–5600.
[2] Gao, H., and Ji, B., 2003, "Modeling Fracture in Nanomaterials via a Virtual Internal Bond Method," Eng. Fract. Mech., **70**, pp. 1777–1791.
[3] Gao, H., Ji, B., Buehler, M. J., and Yao, H., 2004, "Flaw Tolerant Bulk and Surface Nanostructures of Biological Systems," Mech. Chem. Biosyst., **1**, pp. 37–52.
[4] Ji, B., and Gao, H., 2004, "Mechanical Properties of Nanostructure of Biological Materials," J. Mech. Phys. Solids, **52**(9), pp. 1963–1990.
[5] Ji, B., and Gao, H., 2004, "A Study of Fracture Mechanisms in Biological Nanocomposites via the Virtual Internal Bond Model," Mater. Sci. Eng., A, **366**, pp. 96–103.
[6] Gao, H., and Yao, H., 2004, "Shape Insensitive Optimal Adhesion of Nanoscale Fibrillar Structures," Proc. Natl. Acad. Sci. U.S.A., **101**, pp. 7851–7856.
[7] Gao, H., Wang, X., Yao, H., Gorb, S., and Arzt, E., 2005, "Mechanics of Hierarchical Adhesion Structure of Gecko," Mech. Mater., **37**, pp. 275–285.
[8] Dugdale, D. S., 1960, "Yielding of Steel Sheets Containing Slits," J. Appl. Mech., **8**, pp. 100–104.
[9] Barenblatt, G. I., 1985, "The Mathematical Theory of Equilibrium Cracks in Brittle Fracture," *Adv. Appl. Mech.*, VII, Academic, New York, pp. 55–129.
[10] Rice, J. R., 1980, "The Mechanics of Earthquake Rupture," International School of Physics "E. Fermi," Course 78, 1979: Italian Physical Society/North Holland, Amsterdam.
[11] Bilby, B. A., Cottrell, A. H., and Swinden, K. H., 1963, "The Spread of Plastic Yield from a Notch," Proc. R. Soc. London, Ser. A, **272**, pp. 304–314.
[12] Suo, Z., Ho, S., and Gong, X., 1993, "Notch Ductile-to-Brittle Transition due to Localized Inelastic Band," J. Eng. Mater. Technol., **115**(3), pp. 319–326.
[13] 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.
[14] Bazant, Z. P., 1976, "Instability, Ductility and Size Effect in Strain-Softening Concrete," J. Eng. Mech., **102**, pp. 331–344.
[15] Bazant, Z. P., and Cedolin, L., 1983, "Finite Element Modeling of Crack Band Propagation," J. Struct. Eng., **109**, pp. 69–92.
[16] Bazant, Z. P., and Planas, J., 1998, *Fracture and Size Effect in Concrete and Other Quasibrittle Materials*, CRC Press, Boca Raton, FL.
[17] Carpinteri, A., 1997, *Structural Mechanics: A Unified Approach*, Chapman and Hall, London.
[18] Mulmule, S. V., and Dempsey, J. P., 2000, "LEFM Size Requirement for the Fracture Testing of Sea Ice," Int. J. Fract., **102**, pp. 85–98.
[19] Bao, G., and Suo, Z., 1992, "Remarks on Crack-Bridging Concepts," Appl. Mech. Rev., **45**, pp. 355–366.
[20] Cox, B. N., and Marshall, D. B., 1994, "Concepts for Bridged Cracks in Fracture and Fatigue," Acta Metall. Mater., **42**, pp. 341–363.
[21] Massabo, R., and Cox, B. N., 1999, "Concepts for Bridged Mode II Delamination Cracks," J. Mech. Phys. Solids, **47**, pp. 1265–1300.
[22] Kendall, K., 1978, "Complexities of Compression Failure," Proc. R. Soc. London, Ser. A, **361**, pp. 245–263.
[23] Karihaloo, B. L., 1979, "A Note on Complexities of Compression Failure," Proc. R. Soc. London, Ser. A, **368**, pp. 483–493.
[24] Drugan, W. J., 2001, "Dynamic Fragmentation of Brittle Materials: Analytical Mechanics-Based Models," J. Mech. Phys. Solids, **49**, pp. 1181–1208.
[25] Lemaitre, J., 1998, *A Course on Damage Mechanics*, 2nd ed., Springer Verlag, Berlin.
[26] Tada, J., Paris, P. C., and Irwin, G. R., 1973, *The Stress Analysis of Cracks Handbook*, 2nd ed., 1985, Del Research Corporation, St. Louis.
[27] Maugis, D., 1992, "Adhesion of Spheres: JKR-DMT Transition Using a Dugdale Model," J. Colloid Interface Sci., **150**, pp. 243–269.
[28] Pugno, N. M., and Ruoff, R. S., 2004, "Quantized Fracture Mechanics," Philos. Mag., **84**, pp. 2829–2845.
[29] Weiner, S., and Wagner, H. D., 1998, "The Material Bone: Structure-Mechanical Function Relations," Annu. Rev. Mater. Sci., **28**, pp. 271–298.
[30] Nix, W. D., and Gao, H., 1998, "Indentation Size Effects in Crystalline Materials: A Law for Strain Gradient Plasticity," J. Mech. Phys. Solids, **46**, pp. 411–425.
[31] Timosheko, S. P., and Goodier, J. N., 1987, *Theory of Elasticity*, McGraw-Hill, New York.

Yeau-Ren Jeng[1]
Professor
Department of Mechanical Engineering,
National Chung Cheng University,
Chia-Yi, Taiwan
e-mail: imeyrj@ccu.edu.tw

Chung-Ming Tan
Department of Mechanical Engineering,
National Chung Cheng University,
Chia-Yi, Taiwan
and
Department of Mechanical Engineering,
Wufeng Institute of Technology,
Chia-Yi, Taiwan

Static Atomistic Simulations of Nanoindentation and Determination of Nanohardness

This paper develops a nonlinear finite element formulation to analyze nanoindentation using an atomistic approach, which is conducive to observing the deformation mechanisms associated with the nanoindentation cycle. The simulation results of the current modified finite element formulation indicate that the microscopic plastic deformations of the thin film are caused by instabilities of the crystalline structure, and that the commonly used procedure for estimating the contact area in nanoindentation testing is invalid when the indentation size falls in the nanometer regime. [DOI: 10.1115/1.1988349]

1 Introduction

Nanoindentation has evolved into a valuable means of determining the mechanical properties of thin films and surfaces in nanometer regimes [1–4]. In conventional hardness tests, "hardness" is defined as the maximum load divided by the area of the residual impression. The prevailing definition of "nanohardness" during nanoindentation is the maximum load divided by the contact area at that load, which is calculated from an analysis of the load-depth curve based on Hertzian continuum mechanics contact theory.

Computer simulation has attracted an increasing interest in recent years as a means of gaining valuable insights into the atomic processes associated with the nanoindentation cycle. Perez et al. [5] employed total-energy pseudopotential calculations to study the plastic behavior which occurs during nanoindentation, and concluded that these plastic deformations are activated by charge delocalizations. Zimmerman et al. [6] conducted atomistic simulations to clarify the surface step effects on nanoindentation. Their results indicated that the yielding load decreases when indentation takes place in the vicinity of a surface step. Fuente et al. [7] studied nanoindentation using scanning tunneling microscopy and atomistic simulations, which successfully reproduced the dislocation loops observed experimentally. Knap et al. [8] addressed the effects of the indenter radius size on the nanoindentation of Au (001), and demonstrated that the indenter force is an unreliable indicator of the onset of dislocation for indenter sizes in the experimental range.

In the main, these studies adopted a Molecular Dynamics (MD) approach to investigate the dislocation activities induced by the incipient plasticity of single perfect crystals during nanoindentation. However, MD simulation is very time-consuming with respect to static atomistic simulations since it involves time resolutions of a picosecond or less. Furthermore, nanoindentation experiments are performed under quasistatic conditions in order to avoid dynamic effects such as heating and creep and to provide a clearer view of the deformation mechanisms of the substrate. However, MD simulation of quasistatic conditions incurs significant computational costs.

Hence, this present study utilizes an alternative approach which resembles FEM methodology and is referred to hereafter as the modified FEM approach [9,10]. Traditionally, FEM is formulated basing on the constitutive law, i.e., the stress-strain relation. In contrast, the modified FEM is formulated from the basis of the interatomic potential. And it uses the concept of the energy minimization scheme [6,7] using second derivatives such as the Newton-Raphson method. Since atoms oscillate thermodynamically around their minimum-energy positions, the proposed approach simply calculates the incremental changes in the minimum-energy positions during the indentation process. In this way, the indentation process becomes quasistatic, thereby reducing the computational task. This study employs the modified FEM approach to simulate elastic-plastic deformations in a copper substrate during a complete nanoindentation cycle. Furthermore, the relative influences of indenter geometry and indentation depth upon the nanohardness are explored and discussed.

2 Methodology

2.1 Atomistic Model. As shown in Figs. 1(*a*) and 1(*b*), the simulated system configurations include a perfect three-dimensional crystalline slab of copper atoms with a (001) surface, and two indenters of different geometries The first indenter is in the form of a sphere whose outer surface atoms form many facets due to a crystalline structure. Meanwhile, the second indenter has a triangular pyramidal form. In the simulation, it is assumed that the hardness of the indenter's diamond tip far exceeds that of the thin copper film, and hence indenter deformation may be neglected during the indentation process. The simulation assumes boundary conditions in which the atoms located at the four sides and base of the simulated film are fully constrained. The interatomic potential of the copper substrate is modeled using the Sutton-Chen potential [11], which has the same functional form as an EAM potential, i.e.,

$$U = \sum_i U_i \tag{1}$$

$$U_i = \frac{1}{2}\sum_{j\neq i} \phi(r_{ij}) + F(\rho_i) = \varepsilon\left(\frac{1}{2}\sum_{j\neq i}\left(\frac{a}{r_{ij}}\right)^n - c\sqrt{\rho_i}\right) \tag{2}$$

where ρ_i is an electron densitylike term for atom i, which is defined as

$$\rho_i = \sum_{j\neq i} f(r_{ij}) = \sum_{j\neq i}\left(\frac{a}{r_{ij}}\right)^m \tag{3}$$

where r_{ij} is the distance between atoms i and j. The copper constants ε(eV), a(Å), c, m, and n have values of 1.238×10^{-2}, 3.6,

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division December 30, 2004; final revision January 19, 2005. Associate Editor: Z. Suo. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

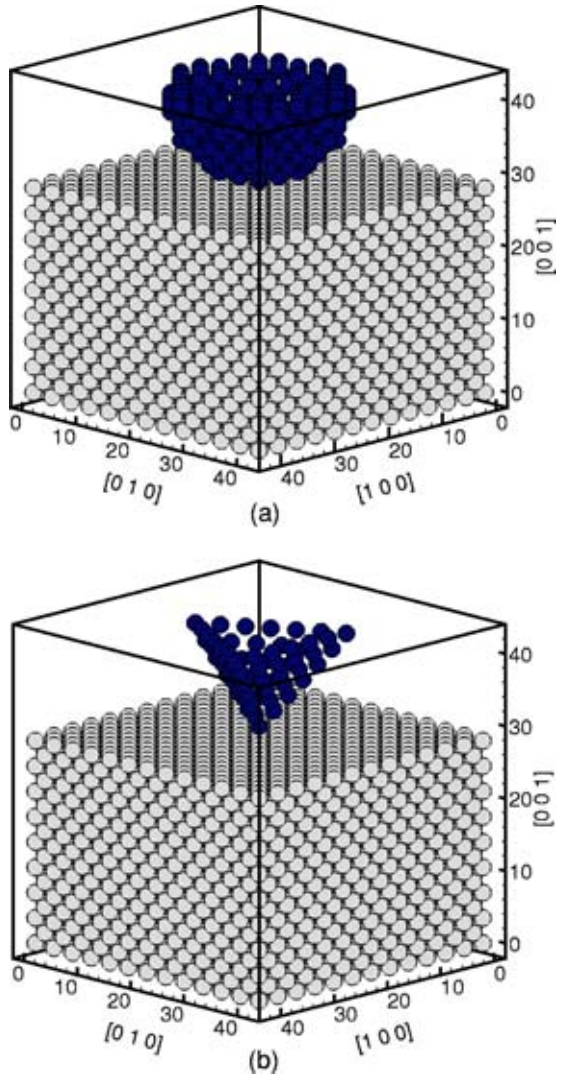


Fig. 1 Atomistic model used in nanoindentation simulations (Units: angstrom)

39.432, 6, and 9, respectively. The potential between the carbon and copper atoms is simulated using the Born-Mayer potential [12], which produces an impulsive force only. This potential has the following form:

$$\phi(r_{ij}) = A\exp[-2\alpha(r_{ij}-r_0)] \tag{4}$$

where r_{ij} is the distance between carbon atom i and copper atom j, and the carbon/copper constants A(eV), a(1/Å), and r_0(Å) have values of 0.3579, 0.9545, and 2.5, respectively.

2.2 Finite Element Formulation. In this procedure, atoms are regarded as nodes, and their potentials are considered to be elements. It is assumed that atom i is located at position (x_i, y_i, Z_i) with displacements u_i, v_i, and w_i in the x-, y-, and z-directions, respectively. By defining the nodal displacement vector for the i atom as $\{u\}_i$ and the corresponding external nodal force vector as $\{F\}_i=(f_i, g_i, h_i)^T$, the total potential energy for atom i can be expressed as

$$E_i = \frac{1}{2}\sum_{j\neq i}\phi(r_{ij}) + F\left(\sum_{j\neq i} f(r_{ij})\right) - \{u\}_i^T\{F\}_i \tag{5}$$

where the atomic distance r_{ij} is given by

$$r_{ij} = \{(x_i+u_i-x_j)^2 + (y_i+v_i-y_j)^2 + (z_i+w_i-z_j)^2\}^{1/2} \tag{6}$$

The differential of the atomic distance with respect to $\{u\}_i$ can be expressed as

$$dr_{ij} = \frac{1}{r_{ij}}[x_i+u_i-x_j, y_i+v_i-y_j, z_i+w_i-z_j]d\{u\}_i = [B]d\{u\}_i \tag{7}$$

The principle of minimum work enforces the minimization of the total energy of the whole system ($E_{\text{total}}=\Sigma_i E_i$) with respect to $\{u\}_i$ such that

$$\frac{\partial E_{\text{total}}}{\partial\{u\}_i} = \sum_{j\neq i}\left[\frac{\partial F}{\partial\rho_i}\frac{\partial f}{\partial r_{ij}} + \frac{\partial F}{\partial\rho_j}\frac{\partial f}{\partial r_{ij}} + \frac{\partial\phi}{\partial r_{ij}}\right] - \{F\}_i = \{0\}. \tag{8}$$

Equation (8) expresses the equilibrium equation at atom i, which represents the equilibrium of the forces acting on atoms i. The unbalance force, $\{\xi\}_i$, can then be defined as

$$\{\xi\}_i = \sum_{j\neq i}\left[\frac{\partial F}{\partial\rho_i}\frac{\partial f}{\partial r_{ij}} + \frac{\partial F}{\partial\rho_j}\frac{\partial f}{\partial r_{ij}} + \frac{\partial\phi}{\partial r_{ij}}\right] - \{F\}_i \tag{9}$$

In order to solve this nonlinear equilibrium equation in an efficient iterative way, it is necessary to differentiate $\{\xi\}_i$ with respect to $\{u\}_i$, i.e.,

$$d\{\xi\}_i = \left\{\frac{\partial^2 F}{\partial\rho_i^2}\left(\sum_{j\neq i}\frac{\partial F}{\partial\rho_i}\frac{[B]}{r_{ij}}\right)\left(\sum_{j\neq i}\frac{\partial F}{\partial\rho_j}\frac{[B]^T}{r_{ij}}\right) + \sum_{j\neq i}\left\{\left(\frac{\partial F}{\partial\rho_i}+\frac{\partial F}{\partial\rho_j}\right)\times\left[\left(\frac{\partial^2 f}{\partial r_{ij}^2} - \frac{1}{r_{ij}}\frac{\partial f}{\partial r_{ij}}\right)[B][B]^T + \frac{1}{r_{ij}}\frac{\partial f}{\partial r_{ij}}[I]\right]\right\} + \sum_{j\neq i}\left[\left(\frac{\partial^2\phi}{\partial r_{ij}^2} - \frac{1}{r_{ij}}\frac{\partial\phi}{\partial r_{ij}}\right)[B][B]^T + \frac{1}{r_{ij}}\frac{\partial\phi}{\partial r_{ij}}[I]\right]\right\}d\{u\}_i = [K_T]_i d\{u\}_i \tag{10}$$

Subsequently, the conventional finite element formulation assembly procedure can be employed to assemble Eq. (8) in order to obtain the total system equation, i.e.,

$$d\{\xi\} = [K_T]d\{u\} \tag{11}$$

Similarly, Eq. (7) can be assembled to obtain the equilibrium equation of the total system, i.e.,

$$\sum_i \xi_i = \{f\}_{\text{internal}} - \{F\}_{\text{external}} = \{0\} \tag{12}$$

In terms of the finite element formulation, Eq. (11) represents the tangent stiffness equation, while the terms $\{f\}_{\text{internal}}$ and $\{F\}_{external}$ in Eq. (12) denote the internal force vector and the external force vector, respectively.

The present simulation adopts the Newton-Raphson iterative technique to solve Eq. (12) via the following displacement control scheme. First, it is assumed that the external force vector $\{F\}$ retains a specified form during the iteration process, i.e.,

$$\{F\}^i = \{F\}^{i-1} + \lambda^i\{\hat{F}\}, \quad i=1,\ 2,\ \ldots \tag{13}$$

where $\{\hat{F}\}$ is the reference load vector.

If Eq. (13) is substituted into Eq. (11), the iterative tangent stiffness equation becomes

$$[K_T]d\{u\}^i = \lambda^i\{\hat{F}\} + d\{\xi\}^i, \quad i=1,\ 2,\ \ldots \tag{14}$$

The iterative displacement increment can be written in a similar form, i.e.,

$$d\{u\}^i = \lambda^i\{u\}_a^i + d\{u\}_b^i, \quad i=1,\ 2,\ \ldots \tag{15}$$

where

$$\begin{cases}[K_T]\{u\}_a^i = \{\hat{F}\}\\ [K_T]d\{u\}_b^i = d\{\xi\}^i\end{cases}, \quad i=1,\ 2,\ \ldots \tag{16}$$

The displacement control scheme is so called because the qth component of the incremental displacement vector is maintained as a constant during the iteration process, i.e.,

$$\lambda^i u^i_{aq} + du^i_{bq} = du^i_q$$

$$du^i_q = \begin{cases} du_q, & i = 1 \\ 0, & i > 1 \end{cases} \tag{17}$$

The modified FEM approach yields the complete nanoindentation equilibrium path. Once the equilibrium position of each atom has been determined, the stress tensor at the atomic site can be determined [13], i.e.,

$$\sigma^i_{km} = \frac{1}{V^i} \sum_{j \neq i}^{N} f^{ij}_k r^{ij}_m \tag{18}$$

where i refers to the atom in question, j refers to the neighboring atom, r^{ij}_m is the displacement vector from atom i to atom j, N is the number of nearest neighboring atoms, and V^i is the volume of the atom in question.

Slip vector analysis [6] is utilized to observe the dislocation activity. The slip vector is given by

$$s = \frac{1}{n_s} \sum_{j}^{n_s} (r^j_t - r^j_\theta) \tag{19}$$

where j is one of the nearest neighbors of the atom in question, n_s is the number of neighboring atoms which have slipped, and r^j_t and r^j_θ are the vector differences of atom j and the atom in question at indentation displacement steps t and θ, respectively.

3 Results and Discussion

Figure 2(*a*) presents the variation in load (i.e., force experienced by the indenter) with indentation depth for spherical indenter nanoindentation cycles performed to maximum indentation depths of 4 Å, 7 Å, and 10 Å. Meanwhile, Fig. 2(*b*) shows the equivalent load-depth curves for pyramidal indenter nanoindentation cycles. It is clear that each of the indentation cycles in Fig. 2 represents a hysteretic loop, which indicates that plastic deformations take place during the loading process. In the present study, the stability of a crystalline structure is correlated to, and can be monitored by, the positiveness of the modified FEM tangent stiffness matrix. The irreversible plastic deformations observed in the simulation can be attributed to changes of the crystalline structure caused by instabilities induced by high localized stress. This perspective of the plastic deformation mechanism bears witness in the work of Li et al., [14], but the precise prediction of the onset of incipient plasticity falls beyond the scope of the present study. Figure 2 reveals that compared to pyramidal indentation, spherical indentation leads to an enhanced yielding phenomenon, which resembles that observed in the uniaxial compression testing of ductile metals. Comparing the hysteresis loops of Figs. 2(*a*) and 2(*b*), it is clear that the enclosed areas are greater in the case of spherical indenter indentation. This indicates that more plastic work is dissipated in indentation cycles performed with this particular indenter geometry.

Figure 3 provides a schematic illustration of the elastic-plastic deformation mechanism observed during the present simulations. It can be seen that irreversible plastic deformation occurs as a result of slips in which the crystal lattice retains its shape and orientation. In addition to plastic deformation, the change in size and distortion of the crystal lattice also induces elastic deformation.

Figure 4 presents the atomic configurations of the copper crystals at the maximum indentation depth following unloading for the two indenters of different geometries. For clarity of observation, the atomic configurations shown in this figure correspond to the atomic arrangements of the single (1 0 0) atomic layer located immediately beneath the two indenter tips. Equation (19) can be

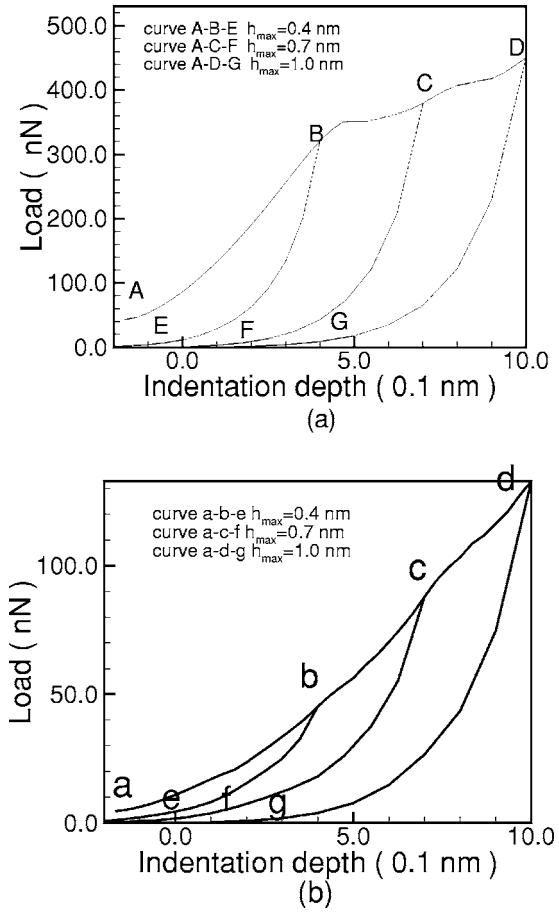


Fig. 2 Load-depth curves for maximum indentation depths of 4 Å, 7 Å, and 10 Å: (*a*) spherical indenter, and (*b*) pyramidal indenter

used to calculate the slip vector at each atomic site. The majority of the atoms do not slip. These atoms are shaded a dark blue color in the figure. The atoms with nonzero slip vectors are shaded with different colors. By comparing subplots (*a*) and (*b*), it can be seen that the elastic dilatation and distortion of the crystal lattice are recovered and that a few more atoms have slipped following unloading. Furthermore, comparing subplots (*c*) and (*d*) shows that the elastic deformation is recovered and that some atoms with small slip vectors are also recovered. Therefore, atoms with small slip vectors are induced by elastic distortion of the crystal lattice.

The high stresses induced as the indenter impresses the substrate cause structural instability of the crystal, which causes some

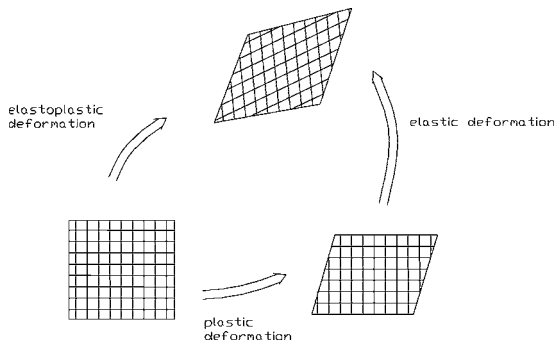


Fig. 3 Representation of elasto-plastic deformation mechanism observed in the simulations

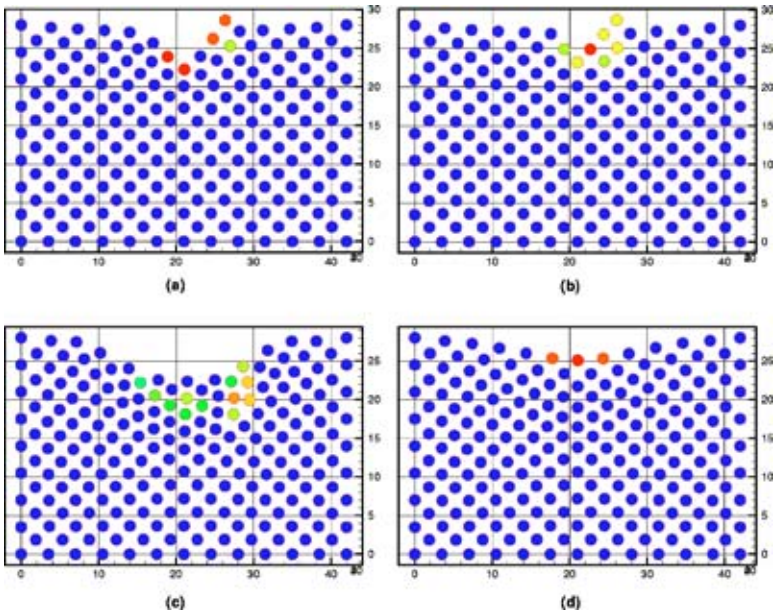


Fig. 4 Atomic configurations of copper crystal at maximum indentation depth following unloading for two indenters of different geometry. Note that (*a*) and (*b*) represent the cross sections through the indenter tip and parallel with the (1 0 0) plane for the pyramidal indenter, while (*c*) and (*d*) present the equivalent cases for the spherical indenter.

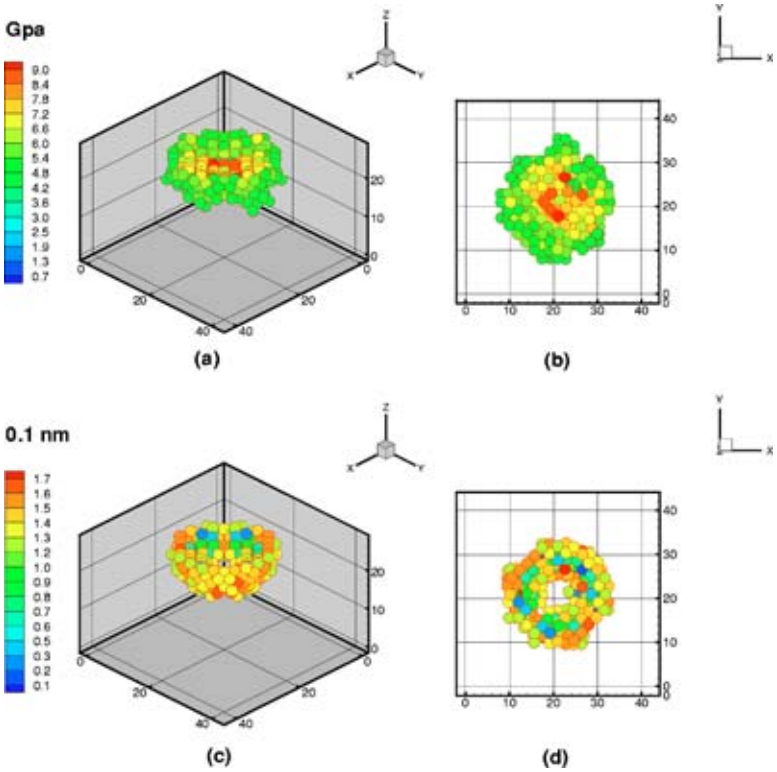


Fig. 5 von Mises shear stress and slip vector distributions for indentation to a depth of 4 Å using a spherical indenter. (*a*)–(*b*) Distributions of von Mises shear stress viewed in the [111] and [001] directions, respectively. (*c*)–(*d*) Distributions of the norm of the slip vector viewed in the [111] and [001] directions, respectively.

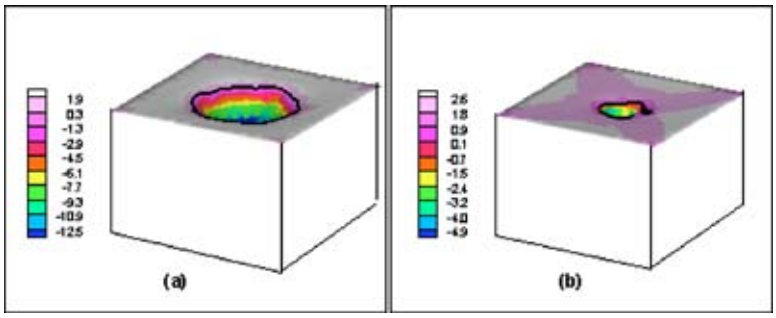


Fig. 6 Flooded contour diagrams of contact pressure distributions on the copper surface. (*a*) Spherical indenter; (*b*) pyramidal indenter. The two bold contour lines represent the boundaries of the contact areas where contact pressure vanishes. (Stress units: GPa.)

of the atoms to slip. Therefore, a correlation must exist between the stresses at the atomic sites and the atom slippage. Using Eq. (18), the stress tensor is calculated at each atomic site of the copper crystal for indentation to a depth of 4 Å using the spherical indenter. Figures 5(*a*) and 5(*b*) show the distributions of von Mises shear stress exceeding 4.5 GPa in the [111], [001] directions, respectively. Meanwhile, Figs. 5(*c*) and 5(*d*) present the distributions of the norm of the slip vector in the [111], [001] directions, respectively. A comparison of the subplots in the upper and lower rows of Fig. 5 indicates that the von Mises shear stress is a qualitatively good indicator of the plastic deformation induced in the simulated nanoindentation.

The principal goal of nanoindentation testing is to determine the elastic modulus and nanohardness of the specimen material from experimental measurements of indenter load and penetration depth. These readings provide an indirect measure of the contact area at full load, from which the mean contact pressure, and hence the hardness, may then be estimated. Therefore, the validity of the nanohardness and elastic modulus results depends largely upon the analytical procedure employed to process the raw data. The present simulations utilize the contact pressure to identify the true contact area at full load. Figures 6(*a*) and 6(*b*) present flooded contour diagrams of the contact pressure distributions on the copper surface for spherical and pyramidal indentations, respectively. The bold contour lines represent the boundaries of the contact areas at which the contact pressure vanishes. Once the true contact area has been determined, the nanohardness can be computed by dividing the load by the projected area of contact at that load. Table 1 presents the nanohardness magnitudes calculated using this method for spherical and pyramidal indentations with maximum indentation depths of 4 Å, 7 Å, and 10 Å.

In nanoindentation tests, analysis of the load-depth curve based on continuum mechanics contact theory gives an indirect measure of the contact area. The contact depth (h_c) is estimated from the total indentation depth (h_t) via

Table 1 Nanohardness values calculated using two methods for two different indenter geometries and three indentation depths.

Geometry of indenter	Indentation depth h_t (Å)	Nanohardness calculated using the true contact area (GPa)	Nanohardness calculated using Eq. (20) (GPa)
pyramidal	4	5.1	15.7
pyramidal	7	4.9	11.8
pyramidal	10	4.7	9.5
spherical	4	7.6	8.1
spherical	7	6.2	7.3
spherical	10	6.3	7.8

$$h_c = h_t - \varepsilon P_{\max}/S_{\max} \tag{20}$$

where $\varepsilon = 0.72$ for the pyramidal indenter, $\varepsilon = 0.75$ for the spherical indenter, and $S_{\max}$ is the stiffness, which is equal to the slope of the unloading curve (dp/dh) at the maximum load ($P_{\max}$).

The projected contact area is obtained by the projected contact area-to-contact depth relationship for the indenter geometry. The nanohardness is then determined by dividing the maximum load by the projected contact area. Table 1 presents the results when this procedure is applied to the simulated load-depth curves to determine the corresponding nanohardness values.

Table 1 shows that the nanohardness values calculated using Eq. (20) are far higher than those calculated using the true contact area identified by the contact pressure. Hence, Eq. (20) significantly underestimates the contact depth in all of the simulation cases. The results also reveal that the nanohardness varies with the indentation depth and the indenter geometry. Therefore, it can be concluded that the indentation size effect is significant in the nanoscale indentations simulated in this present study.

It can also be seen that the nanohardness obtained using the pyramidal indenter is higher than that obtained using the spherical indenter when the contact area is measured via the contact pressure. In contrast, if the contact area is obtained from Eq. (20), the pyramidal indenter indentation yields a higher nanohardness. Figure 2 has demonstrated that the spherical indenter indentation consumes greater plastic work. This observation is consistent with the result that the nanohardness calculated by direct measurement of the contact area is higher. Therefore, it can be concluded that the commonly used procedure for estimating the contact area in nanoindentation testing is invalid when the indentation size falls in the nanometer regime. The reason why Eq. (20) underestimates the real contact area of nanoscale indentation depth is that the assumption implied by Eq. (20) fails significantly. In other words, the deformation behavior at the incipient unloading is elastic-plastic instead of purely elastic behavior implied by Eq. (20).

4 Conclusion

In conclusion, this paper has presented a more efficient approach than conventional MD simulation for the investigation of the elastic-plastic deformations which occur during nanoscale indentations of a thin film. The proposed approach utilizes the minimum energy principle via nonlinear finite element formulation to reduce the indentation process to a quasistatic problem. The simulation results of the current modified finite element formulation indicate that the microscopic plastic deformations of the thin film are caused by instabilities of the crystalline structure. It has been shown that the analytical procedure commonly adopted in nanoindentation testing significantly overestimates the nanohardness. The simulation results have confirmed that both the indentation depth and the indenter geometry influence the nanohardness results.

Acknowledgment

The authors gratefully acknowledge the financial support provided to this study by the National Science Council of Taiwan, under Grant No. NSC 91-2218-E-274-001, and American AFOSR, under Contract No. F62562-03-P-0378.

References

[1] Lucas, B. N., Oliver, W. C., and Swindeman, J. E., 1998, "The Dynamics of Frequency-Specific, Depth-Sensing Indentation Test," Mater. Res. Soc. Symp. Proc., **522**, pp. 3–14.

[2] Hay, J. C., and Pharr, G. M., 1998, "Experiment Investigation of the Sneddon Solution and an Improved Solution for the Analysis of Nanoindentation Data," Mater. Res. Soc. Symp. Proc., **522**, pp. 39–44.

[3] Hay, J. L., O'Hern, M. E., and Oliver, W. C., 1998, "The Importance of Contact Radius for Substrate-Independent Property Measurement of Thin Film," Mater. Res. Soc. Symp. Proc., **522**, pp. 27–32.

[4] Lu, W., and Komvopoulos, K., 2001, "Nanotribological and Nanomechanical Properties of Ultrathin Amorphous Carbon Films Synthesized by Radio Frequency Sputtering," J. Tribol., **123**, pp. 641–650.

[5] Perez, R., Payne, M. C., and Simpson, A. D., 1995, "First Principles Simulations of Silicon Nanoindentation," Phys. Rev. Lett., **75**, pp. 4748–4751.

[6] Zimmerman, J. A., Kelchner, C. L., Klein, P. A., Hamilton, J. C., and Foiles, S. M., 2001, "Surface Step Effects on Nanoindentation," Phys. Rev. Lett., **87**, p. 165507.

[7] Fuente, O. R. de la, Zimmerman, J. A., Gonzalez, M. A., Figuera, J. de la, Hamilton, J. C., Pai, W. W., and Rojo, J. M., 2002, "Dislocation Emission around Nanoindentations on a (001) fcc Metal Surface Studied by Scanning Tunneling Microscopy and Atomistic Simulations," Phys. Rev. Lett., **88**, p. 036101.

[8] Knap, J., and Ortiz, M., 2003, "Effect of Indenter-Radius Size on Au(001) Nanoindentation," Phys. Rev. Lett., **90**, p. 226102.

[9] Jeng, Y. R., and Tan, C. M., 2002, "Computer Simulation of Tension Experiments of a Thin Film Using an Atomic Model," Phys. Rev. B, **65**, p. 174107.

[10] Jeng, Y. R., and Tan, C. M., 2004, "Theoretical Study of Dislocation Emission around a Nanoindentation Using a Static Atomistic Model," Phys. Rev. B, **69**, p. 104109.

[11] Rafii-Tabar, H., 2000, "Modelling the Nano-Scale Phenomena in Condensed Matter Physics via Computer-Based Numerical Simulations," Phys. Rep., **325**, pp. 239–310.

[12] Inamura, T., Suzuki, H., and Takezawa, N., 1991, "Cutting Experiments in a Computer Using Atomic Models of a Copper Crystal and a Diamond Tool," Int. J. Jpn. Soc. Precis. Eng., **25**, pp. 259–266.

[13] Alber, I., Bassani, J. L., Khantha, M., Vitek, V., and Wang, G. J., 1992, "Grain Boundaries as Heterorgeneous Systems: Atomic and Continuum Elastic Properties," Philos. Trans. R. Soc. London, Ser. A, **339**, pp. 555–586.

[14] Li, J., Vliet, K. J. V., Zhu, T., Yip, S., Suresh, S., 2002, "Atomistic Mechanisms Governing Elastic Limit and Incipient Plasticity in Crystals," Nature (London), **418**, pp. 307–310.

An Acoustic Model for Wave Propagation in a Weak Layer

Michael El-Raheb
ATK Mission Research,
23052 Alcade Drive,
Laguna Hills, CA 92653

An acoustic model is developed for transient wave propagation in a weak layer excited by prescribed pressure or prescribed acceleration at the boundary. The validity of the acoustic model is investigated for the two excitations. A comparison of transient response from the acoustic model and a 3D axisymmetric elastic model reveals that for prescribed acceleration the acoustic model fails to capture important features of the elastic model even as Poisson ratio ν approaches 1/2. However for prescribed pressure, the two models agree since shear stress is reduced. For prescribed acceleration adopting the modal approach, the mixed boundary-value problem on the excited boundary is converted to a pure traction problem utilizing the influence method. To validate the elaborate modal approach a finite difference model is also developed. [DOI: 10.1115/1.1988367]

1 Introduction

Laboratory simulation of blunt trauma in living tissue relies on measuring propagation of stress waves from low velocity impact in a weak viscoelastic material such as ordnance gelatin. It has acoustic impedance close to that of water yet living tissue dissipates energy from viscoelasticity and possesses shear rigidity controlling transverse propagation. It has been widely assumed that gelatin is similar to water because it has approximately the same density and bulk speed of sound. In a weak solidlike gelatin, effects of the free surface and lateral propagation of a forcing pulse are controlled by shear modulus G and the speed of shear waves, respectively. These types of propagation are independent of a loss mechanism like viscoelasticity. Loss produces an attenuation of the pulse over and above that from dispersion. It reduces the participation of high frequency modes by smoothing the average response and its gradients.

In a fluid like water, propagation is mostly volumetric, with shear related to dissipation that is proportional to velocity gradient and kinematic viscosity. At the free surface a different kind of wave develops controlled by gravity and depth of the fluid. It can be argued that although water and gelatin have very similar acoustic impedances, shear rigidity of gelatin may control how a stress wave propagates laterally and its character at and close to the free surface. If gelatin is like water then it can be treated as an acoustic fluid governed by the wave equation. In this work the wave equation is derived as a limiting case of the linear elastodynamic equations of a homogeneous solid. In fact when Poisson ratio assumes the value of 1/2, the elastic field converts to the acoustic field. One issue addressed in this work is the sensitivity of the solution to Poisson ratio close to 1/2.

To measure transmission of stress waves produced by low velocity impact on gelatin, a layer is bonded onto a metallic substrate instrumented by sensitive carbon gauges. Upon impact, stress waves propagate across the layer reaching the substrate with substantial reduction in intensity from dispersion and viscous losses. Measuring impact and transmitted pressures are needed to construct the material's constitutive model. Carefully controlled experiments with sufficient accuracy reproducing transient histories for correlation with computed results are very hard to execute. The problem lies in the weakness of the material. Gauges cannot be placed inside the material while gauges at the interface between material and metal substrate suffer from lack of cohesion adding uncertainty to measured data. This difficulty forces investigators to rely on sensitivity studies from analysis and general purpose discretization programs in order to understand phenomena. Moreover, literature in this field addresses quasistatic measurements of elongation omitting important dynamic effects such as strain-rate dependence in the microsecond regime. The simulation of these experiments led to the realization that approximating gelatin as a viscous fluid is valid only for unrealistic impact conditions when pressure over the footprint is uniform.

Acoustic wave propagation governed by the Helmholtz equation has been treated extensively in the literature. Solution techniques range from the analytical for simple geometries to numerical for problems with complicated geometry, medium inhomogeneity, and nonlinearity. Theil [1] treats the 1D viscoelastically damped wave equation analytically. Yserentant [2] shows how a consistent discretization of the acoustic equation can be recovered from the particle model of compressible fluids (see Ref. [3]). Sina and Khashayar [4] solve the 3D wave equation analytically for arbitrary nonhomogeneous media adopting the differential transfer matrix. Sujith et al. [5] present an exact solution to 1D transient waves in curvilinear coordinates adopting transformation of variables suggested by the WKB approximation. Hamdi et al. [6] present exact solitary wave solutions of the 1D wave propagation in nonlinear media with dispersion. Yang [7] solves numerically the wave equation with attenuation from linear friction utilizing grid modification to track wave fronts accurately. Narayan [8] solves the 3D transient acoustics in inhomogeneous media by finite difference and Schemann and Bornemann [9] apply the adaptive Rothe integrator. Bailly and Juve [10] present a numerical solution to the 2D acoustic propagation from transient sources using the dispersion-relation-preserving scheme in space and a fourth-order Runge-Kutta in time. Wagner et al. [11] and Gaul and Wenzel [12] use a hybrid boundary element method for frequency and transient acoustic response in bounded and unbounded regions. Mehdizadeh and Paraschivoiu [13] develop a spectral element method to solve the 3D Helmholtz equation retaining accuracy for large wave numbers. None of the references above addresses 3D transient propagation from impact analytically.

Acoustic wave propagation in a free disk is developed here adopting a modal analysis validated by a finite difference method. Transient response to prescribed pressure and prescribed acceleration at the boundary is analyzed. Since the primary goal of this work is to investigate the validity of the established belief that

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, December 21, 2004; final revision, February 7, 2005. Associate Editor: O. M. O'Reilly. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

tissue can be treated as a fluid, the acoustic equation is derived from the elastic equations of a solid in the limit when Poisson ratio and shear stresses vanish (Appendix).

Section 2 develops the acoustic model utilizing the modal approach for both prescribed pressure and prescribed acceleration. In the modal approach, the forcing function at the boundary is treated adopting the static-dynamic superposition method (see Berry and Naghdi [14]). The solution is expressed as a superposition of a static term satisfying the inhomogeneous boundary conditions, and a dynamic solution in terms of the eigenfunctions satisfying homogeneous boundary conditions.

Since the projectile's strength and acoustic impedance are much greater than those of tissue, the excitation transmitted over the boundary at the projectile-tissue interface can be approximated as a given time dependent prescribed motion in contrast to an unknown pressure excitation. However, this type of excitation would lead to a mixed boundary condition; i.e., pressure gradient prescribed over part of the boundary and zero pressure prescribed over the remaining part. This difficulty can be overcome by the influence method which superimposes response from a set of unit pressures with time-dependent weights prescribed on annular portions of the footprint. These weights are updated at each time step from the condition that combined acceleration at the center of each ring over the footprint equals the prescribed instantaneous acceleration. In this way, the forcing function is converted to pure traction with time-varying spatial dependence.

Section 3 develops the finite difference approach. Radial and axial dependence are discretized by central differences while time dependence is integrated by the Kutta-Runge method.

Section 4 compares acoustic histories from the two approaches validating the modal approach. Histories of the acoustic model are compared to those from a 3D axisymmetric elastic model demonstrating the inadequacy of the acoustic model when applied to a solid with Poisson ratio near $1/2$ and forced by applied acceleration. Sensitivity of the acoustic histories to type of excitation and to parameters of the prescribed acceleration profile is also presented. The effect of Poisson ratio ν on peak elastic stress is evaluated confirming that for prescribed acceleration mismatch of acoustic and elastic results is not caused by small deviations in Poisson ratio ν from $1/2$ in the elastic model. Finally, results from the two models are compared for prescribed uniform pressure revealing that the mismatch diminishes when shear stress is reduced.

2 Modal Analysis

In the analysis to follow, all variables are independent of circumferential angle due to the assumption of axisymmetry. This condition applies for a cylindrical projectile at normal incidence. Consider a traction-free disk with radius r_d and length h bonded to a rigid substrate. Appendix A derives the acoustic equation in the limit when shear stress vanishes in the linear elastodynamic equations of a solid. In the analysis to follow, r and z denote radial and axial coordinates. Acoustic propagation in the disk is governed by the acoustic equation

$$(\partial_{rr} + 1/r\partial_r + \partial_{zz})p_d - 1/c_b^2\partial_{tt}p_d = 0 \tag{1a}$$

with the following boundary conditions:

$$p(r_d,z;t) = 0 \tag{1b}$$

$$p(r,0;t) = [H(r) - H(r-r_p)]f(t) \quad \text{prescribed pressure}$$

$$\left.\begin{aligned} \partial_z p(r,0;t) &= -\rho\ddot{f}_w(t) \quad 0 \leqslant r \leqslant r_p \\ p(r,0;t) &= 0 \qquad\qquad r_p < r < r_d \end{aligned}\right\} \quad \text{prescribed acceleration} \tag{1c}$$

$$\partial_z p(r,h;t) = 0, \quad \text{fixed face, or alternatively}$$

$$p(r,h;t) = 0, \quad \text{free face} \tag{1d}$$

where $H(r)$ is the Heaviside function, r_p is footprint radius of the external excitation which is projectile radius, $f(t)$ is time dependence of prescribed pressure, and $\ddot{f}_w(t)$ is time dependence of prescribed acceleration. Express $p(r,z;t)$ as a superposition of two terms (see Berry and Naghdi [14])

$$p(r,z;t) = \begin{cases} p_s(r,z)f(t) + p_d(r,z;t), & \text{prescribed pressure} \\ -p_s(r,z)\rho\ddot{f}_w(t) + p_d(r,z;t), & \text{prescribed acceleration} \end{cases} \tag{2}$$

where $p_s(r,z)$ is the static solution of Eq. (1a) with inhomogeneous boundary conditions (1b)–(1d) assuming $f(t)=1$ or $\ddot{f}_w(t) = 1/\rho$, and $p_d(r,z;t)$ is a dynamic solution of Eq. (1a) satisfying the homogeneous boundary conditions (1b)–(1d) with $f(t)=0$ or $\ddot{f}_w(t)=0$.

The prescribed acceleration boundary condition in Eq. (1c) is mixed. In other words, part of the boundary has prescribed pressure gradient and the other part has prescribed pressure. This difficulty can be overcome by dividing the circle bounding the footprint into $n+1$ equidistant radial stations with increment Δr_p

$$0, r_1, r_2, \ldots, r_{n-1}, r_n, \quad r_k - r_{k-1} = \Delta r_p = \text{const}$$

where $r_n = r_p$. Assume a uniform pressure of unit intensity acting over each annular segment $r_{k-1} \rightarrow r_k$ that is termed source segment. Where subscript z denotes partial derivative with respect to z, evaluating the pressure gradient $P_{z,lk}(r,z;t)$ from the kth source segment at the center of the lth segment $r_{cl} = (r_l + r_{l-1})/2$ that is termed target point and following the expansion in (2) yields

$$P_{z,lk}(r_{cl},0;t) = -p_{zs,lk}(r_{cl},0)\rho\ddot{f}_w(t) + p_{zd,lk}(r_{cl},0;t) \tag{3}$$

where $p_{zs,lk}(r_{cl},0;t)$ and $p_{zd,lk}(r_{cl},0)$ are static and dynamic pressure gradients at the lth target point due to the kth source segment. Enforcing the condition of prescribed pressure gradient $p_{zf}(t)$ over the footprint at each time step yields a set of simultaneous equations in the weights $c_k(t)$

$$\sum_{k=1}^{n} P_{z,lk}(r_{cl},0;t)c_k(t) = p_{zf}(t), \quad 1 \leqslant l \leqslant n \tag{4}$$

The combined pressure from all annular source segments is the superposition of $P_{lk}(r,z;t)$ factored by time dependent weights $c_k(t)$

$$p(r,z;t) = \sum_{k=1}^{n} P_{lk}(r,z;t)c_k(t), \quad 1 \leqslant l \leqslant n$$

$$P_{lk}(r,z;t) = -p_{s,lk}(r,z)\rho\ddot{f}_w(t) + p_{d,lk}(r,z;t) \tag{5a}$$

Solutions of $p_{s,k}(r,z;t)$ and $p_{d,k}(r,z;t)$ for each unit source segment are outlined in what follows. The static solution for the kth source segment $p_s(r,z)$ takes the form

$$p_{s,k}(r,z) = \sum_{m=1}^{m_r} \psi_{sm,k}(z)J_0(k_{rm}r)$$

$$\psi_{sm,k}(z) = \alpha_{mk}\sinh(k_{rm}z) + \beta_{mk}\cosh(k_{rm}z) \tag{5b}$$

where $J_0(k_{rm}r)$ is the Bessel function of the first kind and zeroth order. Substituting (5b) in the boundary conditions (1b)–(1d) and enforcing orthogonality of $J_0(k_{rm}r)$ yields

$$J_0(k_{rm}r_d) = 0, \quad 1 \leqslant m \leqslant m_r \tag{6a}$$

$$\beta_{m,k} = \frac{2(r_k J_1(k_{rm}r_k) - r_{k-1}J_1(k_{rm}r_{k-1}))}{r_d^2 J_1^2(k_{rm}r_d)k_{rm}} \tag{6b}$$

$$\alpha_{m,k}=\begin{cases}-\beta_{m,k}\tanh(k_{rm}h), & \text{fixed face, or alternatively}\\ -\beta_{m,k}/\tanh(k_{rm}h), & \text{free face}\end{cases} \tag{6c}$$

Note that in (3), $p_{zs,lk}(r_{cl},0)=\partial_z p_{s,k}(r_{cl},0)$.

The dynamic solution $p_{d,k}(r,z;t)$ satisfies

$$(\partial_{rr}+1/r\partial_r+\partial_{zz})p_{d,k}-1/c_b^2\partial_{tt}p_{d,k}=0 \tag{7}$$

and the homogeneous boundary conditions in (1*b*)–(1*d*). Expand $p_d(r,z;t)$ in terms of its orthogonal eigenfunctions

$$p_{d,k}(r,z;t)=\sum_m\sum_n a_{mn,k}(t)\psi_{dn}(z)J_0(k_{rm}r) \tag{8}$$

Applying the homogeneous boundary conditions in (1*b*)–(1*d*) to $J_0(k_{rm}r)$ and $\psi_{dn}(z)$ produces

$$J_0(k_{rm}r_d)=0, \quad 1\leq m\leq m_r \tag{9a}$$

$$\psi_{dn}(z)=\begin{cases}\cos(k_{zn}z),\cos(k_{zn}h)=0\rightarrow k_{zn}h=\dfrac{1}{2}(2n-1)\pi, & \text{fixed face}\\ \sin(k_{zn}h),\sin(k_{zn}h)=0\rightarrow k_{zn}h=n\pi, & \text{free face}\end{cases}, \quad 1\leq n\leq n_z \tag{9b}$$

$$k_{zn}^2+k_{rm}^2=k_{mn}^2, \quad \omega_{mn}=c_b k_{mn} \tag{9c}$$

where ω_{mn} is the eigenfrequency corresponding to mode (m,n). Substituting (3) in (1*a*) with use made of (5*a*), (5*b*), (6), (8), and (9) and enforcing orthogonality of $\psi_{dn}(z)$ and $J_0(k_{rm}r)$ yields

$$\ddot{a}_{mn,k}(t)+\omega_{mn}^2 a_{mn,k}(t)=-N_{sd\ mn,k}\rho f_w^{IV}(t); \quad f_w^{IV}(t)=\partial^4 f_w(t)/\partial t^4$$

$$N_{sd\ mn,k}=\frac{2}{h}\int_0^h \psi_{sm,k}(z)\psi_{dn}(z)dz, \quad 1\leq m\leq m_r, \quad 1\leq n\leq n_z \tag{10}$$

In deriving Eq. (10) the term $\nabla_0^2(-p_s)\rho\ddot{f}_w(t)$, $(\nabla_0^2\equiv\partial_{rr}+1/r\partial_r)$ vanishes since static pressure $p_s(r,z)$ satisfies the equation $\nabla_0^2 p_s=0$. Acoustic displacements $(w,u)_k$ are determined from (A4)

$$\partial_z p_{d,k}=-\rho\partial_{tt}^2 w_k$$

$$\partial_r p_{d,k}=-\rho\partial_{tt}^2 u_k \tag{11}$$

The solution to (10) is expressed as a Duhamel integral

$$a_{mn,k}(t)=-\frac{\rho N_{sd\ mn,k}}{\omega_{mn}}\int_0^t \sin\omega_{mn}(t-\tau)f_w^{IV}(\tau)d\tau \tag{12}$$

Note that in (11) $\partial_z p_{d,k}(r_{cl},0;t)=p_{zd,lk}(r_{cl},0;t)$ as defined in (3). Once histories of $\partial_z p_{d,k}$ and $\partial_r p_{d,k}$ are determined from solving (10), histories of w_k and u_k are found by integrating (11) numerically.

3 Finite Difference

Consider a disk with traction-free boundaries satisfying the conditions

$$\partial_r p(0,z;t)=0 \tag{13a}$$

$$p(r_d,z;t)=0 \tag{13b}$$

$$\partial_z p(r,0;t)=0 \tag{13c}$$

$$p(r,h;t)=[H(r)-H(r-r_p)]f(t) \quad \text{prescribed pressure}$$

$$\left.\begin{aligned}\partial_z p(r,h;t)&=-\rho\ddot{f}_w(t), & 0\leq r\leq r_p\\ p(r,h;t)&=0, & r_p<r\leq r_d\end{aligned}\right\} \quad \text{prescribed acceleration} \tag{13d}$$

where ( ) denotes time derivative. Unlike the analysis in Sec. 2 where z has its origin at the excited boundary, in the finite difference scheme z has its origin at the nonexcited boundary. Condition (13*a*) is symmetry about the axis of revolution $r=0$, (13*b*) is traction-free boundary at $r=r_d$, (13*c*) is fixed boundary at $z=0$, and (13*d*) is prescribed acceleration for $0\leq r\leq r_p$ and traction-free boundary for $r_p\leq r\leq r_d$ at $z=h$. Form the rectangular grid

$$\begin{aligned}&i=1\rightarrow n_r, \quad d_r\leq r\leq r_d-d_r, \quad d_r=r_d/(n_r+1)\\ &j=1\rightarrow n_z, \quad d_z\leq z\leq h-d_z, \quad d_z=h/(n_z+1)\end{aligned} \tag{14}$$

In this grid, nodes do not include points on the boundaries. Expressing Eq. (1*a*) in central difference to first order yields the following relations depending on position:

(a) Internal points $d_r<r<r_d-d_r$, $d_z<z<h-d_z\Rightarrow 2\leq i\leq n_r-1$, $2\leq j\leq n_z-1$

$$\alpha_1 p_{i+1,j}+\alpha_2 p_{i-1,j}+\alpha_3 p_{i,j}+\alpha_4(p_{i,j+1}+p_{i,j-1})=1/c_b^2\ddot{p}_{i,j}$$

$$\alpha_1=\left(\frac{1}{d_r^2}+\frac{1}{2r_i d_r}\right), \quad \alpha_2=\left(\frac{1}{d_r^2}-\frac{1}{2r_i d_r}\right), \quad \alpha_3=-2\left(\frac{1}{d_r^2}+\frac{1}{d_z^2}\right), \quad \alpha_4=\frac{1}{d_z^2} \tag{15a}$$

(b) Corner point at $r=d_r$, $z=d_z\Rightarrow i=1$, $j=1$

$$\alpha_1 p_{i+1,j}+(\alpha_2+\alpha_3+\alpha_4)p_{i,j}+\alpha_4 p_{i,j+1}=1/c_b^2\ddot{p}_{i,j} \tag{15b}$$

(c) Points along axis $r=d_r$, $d_z<z<h-d_z\Rightarrow i=1$, $2\leq j\leq n_z-1$

$$\alpha_1 p_{i+1,j}+(\alpha_2+\alpha_3)p_{i,j}+\alpha_4(p_{i,j+1}+p_{i,j-1})=1/c_b^2\ddot{p}_{i,j} \tag{15c}$$

(d) Corner point at $r=d_r$, $z=h-d_z\Rightarrow i=1$, $j=n_z$

For prescribed pressure

$$\alpha_1 p_{i+1,j}+(\alpha_2+\alpha_3)p_{i,j}+\alpha_4 p_{i,j-1}-1/c_b^2\ddot{p}_{i,j}=\alpha_4 f(t) \tag{15d}$$

For prescribed acceleration

$$\alpha_1 p_{i+1,j}+(\alpha_2+\alpha_3+\alpha_4)p_{i,j}+\alpha_4 p_{i,j-1}-1/c^2\ddot{p}_{i,j}=-\rho\ddot{f}_w(t)/d_z$$

(e) Points along boundary $d_r<r<r_d-d_r$, $z=d_z\Rightarrow 2\leq i\leq n_r-1$, $j=1$

$$\alpha_1 p_{i+1,j} + \alpha_2 p_{i-1,j} + (\alpha_3 + \alpha_4) p_{i,j} + \alpha_4 p_{i,j+1} = 1/c_b^2 \ddot{p}_{i,j} \tag{15e}$$

(f) Points along boundary $d_r < r < r_d - d_r$, $z = h - d_z \Rightarrow 2 \leq i \leq n_r - 1$, $j = n_z$

For $0 \leq r \leq r_p$ and prescribed pressure

$$\alpha_1 p_{i+1,j} + \alpha_2 p_{i-1,j} + \alpha_3 p_{i,j} + \alpha_4 p_{i,j-1} - 1/c_b^2 \ddot{p}_{i,j} = \alpha_4 f(t) \tag{15f}$$

For $0 \leq r \leq r_p$ and prescribed acceleration

$$\alpha_1 p_{i+1,j} + \alpha_2 p_{i-1,j} + (\alpha_3 + \alpha_4) p_{i,j} + \alpha_4 p_{i,j-1} - 1/c_b^2 \ddot{p}_{i,j} = -\rho \ddot{f}_w(t)/d_z$$

For $r_p < r \leq r_d$

$$\alpha_1 p_{i+1,j} + \alpha_2 p_{i-1,j} + \alpha_3 p_{i,j} + \alpha_4 p_{i,j-1} - 1/c_b^2 \ddot{p}_{i,j} = 0$$

(g) Corner point at $r = r_d - d_r$, $z = d_z \Rightarrow i = n_r$, $j = 1$

$$\alpha_2 p_{i-1,j} + (\alpha_3 + \alpha_4) p_{i,j} + \alpha_4 p_{i,j+1} = 1/c^2 \ddot{p}_{i,j} \tag{15g}$$

(h) Points along boundary $r = r_d - d_r$, $d_z < z < h - d_z \Rightarrow i = n_r$, $2 \leq j \leq n_z - 1$

$$\alpha_2 p_{i-1,j} + \alpha_3 p_{i,j} + \alpha_4 (p_{i,j+1} + p_{i,j-1}) = 1/c_b^2 \ddot{p}_{i,j} \tag{15h}$$

(i) Corner point at $r = r_d - d_r$, $z = h - d_z \Rightarrow i = n_r$, $j = n_z$

$$\alpha_2 p_{i-1,j} + \alpha_3 p_{i,j} + \alpha_4 p_{i,j-1} = 1/c_b^2 \ddot{p}_{i,j} \tag{15i}$$

In (15*a*)–(15*i*), the differential equation is satisfied only at internal points of the grid modified by constraints on the boundaries.

Applying (15*a*)–(15*i*) at all internal points in the grid (14) produces a set of ordinary differential equations in $p_{i,j}(t)$ cast in the form of tridiagonal blocks as follows:

$$\ddot{\mathbf{p}} = c_b^2 (\mathbf{M}_p \mathbf{p} - \mathbf{F}(t))$$

$$\mathbf{M}_p = \begin{bmatrix} \mathbf{A}_1 & \mathbf{C}_1 & & & & \\ \mathbf{B}_2 & \mathbf{A}_2 & \mathbf{C}_2 & & & \\ & \square & \square & \square & & \\ & & \mathbf{B}_{n_r-1} & \mathbf{A}_{n_r-1} & \mathbf{C}_{n_r-1} & \\ & & & \mathbf{B}_{n_r} & \mathbf{A}_{n_r} \end{bmatrix} \tag{16}$$

$\mathbf{B}_i$ and $\mathbf{C}_i$ are $(n_z \times n_z)$ diagonal matrices, $\mathbf{A}_i$ is the $(n_z \times n_z)$ banded matrix with bandwidth 3, and $\mathbf{F}$ is the global vector of the forcing function in (15*d*) and (15*f*). For each point $j \ni (1 \leq j \leq n_z)$ along an i line in the grid, coefficients of $p_{i,j}$ in the Laplacian define $\mathbf{A}_i$, coefficients of $p_{i-1,j}$ define $\mathbf{B}_i$, and coefficients of $p_{i+1,j}$ define $\mathbf{C}_i$. The time derivative is expressed in the central difference to first order allowing integration in time.

Viscous damping is included following the approximate equation (A12)

$$(1 + \tilde{\nu}/c_b^2 \partial_t)(\partial_{rr} + 1/r \partial_r + \partial_{zz}) p - 1/c_b^2 \partial_{tt} p = 0 \tag{A12}$$

This modifies (16) to the first order system

$$\dot{\mathbf{p}} = \mathbf{q}$$

$$\dot{\mathbf{q}} = c_b^2 \mathbf{M}_p \mathbf{p} + \tilde{\nu} \mathbf{M}_p \mathbf{q} - c_b^2 \mathbf{F}(t) \tag{17}$$

4 Results

The numerical experiments to follow assume a traction-free gelatin disk 12.7 mm (=0.5 in.) thick and 25.4 mm (=1 in.) radius with the boundary $z = h$ bonded to a rigid surface. In the elastic model the gelatin properties are (Eisler [16]):

$$E = 3.1 \times 10^9 \text{ dyn/cm}^2 (= 4.5 \times 10^4 \text{ lb./in.}^2),$$

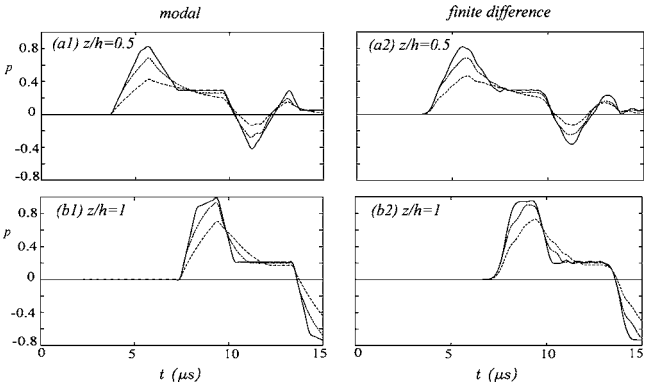


Fig. 1 Acoustic histories from prescribed pressure: —, $r=0$; ---, $r=0.5r_p$; – – –, $r=0.9r_p$. (*a*1), (*b*1) modal; (*a*2), (*b*2) finite difference.

$$\rho = 0.93 \text{ g/cm}^3 (= 8.7 \times 10^{-5} \text{ lb. s}^2/\text{in.}^4), \quad \nu = 0.48 \tag{18a}$$

The data in (13*a*) yield a small ratio of *Lame*' constants $\mu/\lambda = (1-2\nu)/(2\nu) \approx 0.0417$ resulting in reduced shear stresses and in turn large displacements. In the acoustic model, bulk modulus E_b, density ρ, and speed of sound c_b are then

$$E_b = E/(3(1-2\nu)) = 2.73 \times 10^{10} \text{ dyn/cm}^2 (= 3.95 \times 10^5 \text{ lb./in.}^2)$$

$$\rho = 0.93 \text{ g/cm}^3 (= 0.87 \times 10^{-4} \text{ lb. s}^2/\text{in.}^4)$$

$$c_b = \sqrt{E_b/\rho} = 1.71 \text{ km/s} (= 6.74 \times 10^4 \text{ in./s}) \tag{18b}$$

E_b is determined from experimental measurements of c_b.

To confirm the implementation of the complicated analytical approach adopting time dependent influence coefficients, results are first compared to those from the more straightforward numerical finite difference approach derived in the Appendix. Figure 1 compares acoustic pressure histories from the two approaches for a layer forced by a prescribed trapezoidal pressure pulse of unit intensity lasting 8 μs with 2 μs rise and fall times and 4 μs plateau applied over a circular footprint with radius $r_p = 6.35$ mm (=0.25 in.). Figures 1(*a*1), 1(*a*2) plots histories at $z = 0.5h$ and Figs. 1(*b*1), 1(*b*2) at $z = h$. For each z, histories at 3 radial stations $r/r_p = 0$, 0.5, and 0.9 are superimposed. Figures 1(*a*1) and 1(*b*1) show that the prescribed pressure pulse quickly changes profile as the wave travels along z. The flat plateau of the profile acquires a discontinuity in intensity after an interval $\Delta t_1 = 3.5$ μs from the wavefront equal to travel time of the wave over r_p. Over this interval intensity diminishes smoothly with z, while over the remaining interval $\Delta t_2 = 4.5$ μs intensity diminishes steeply with z. At $z = h$, intensity over Δt_1 rises from reflections at the rigid boundary. Histories from the two distinctly different approaches agree confirming the implementation of the analytical model.

The difference in response between the acoustic model and the 3D axisymmetric elastic model is discussed in what follows. Figures 2(*a*) and 2(*b*) plot the eigenfrequency Ω (kHz) versus radial wave number $\lambda_m/\pi = k_{rm} r_d/\pi$ with axial wave number n as parameter for the elastic and acoustic models. For each mode (m,n), Ω of the acoustic model is 5 times higher than that of the elastic model. The reason is that in the acoustic model Ω is proportional to c_b while in the elastic model it is proportional to the flexural phase velocity c_p that is bounded by the shear speed $c_s = \sqrt{E/(2(1+\nu)\rho)}$. For $\nu = 0.48$, $c_b/c_s = 4.97$ consistent with the ratio observed in Fig. 2. This is the fundamental difference distinguishing the two models. Furthermore, the acoustic model cannot capture transverse wave propagation as no shear is included in the model. Although in the elastic model extensional modes exist with frequencies proportional to c_b nevertheless flexural modes dominate the response because of their lower frequencies.

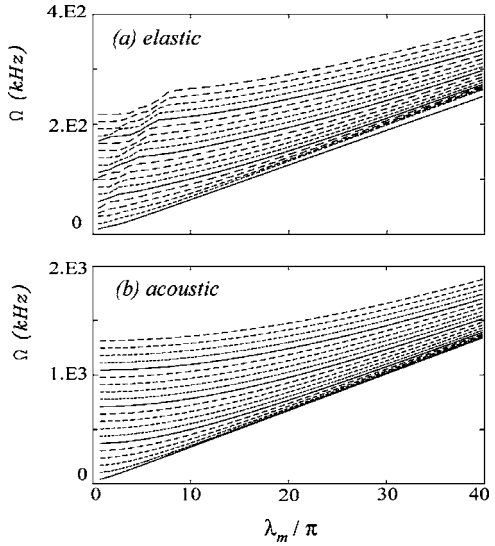


Fig. 2 Frequency spectra of elastic and acoustic models. (*a*) Elastic model; (*b*) acoustic model.

Figure 3 plots prescribed motion $\ddot{f}_w(t)$, $\dot{f}_w(t)$, and $f_w(t)$ when acceleration is prescribed at the footprint. $\ddot{f}_w(t)$ is made of 4 linear segments

(1) Linear acceleration: $\ddot{f}_{w1}(t)=\alpha_1 t,\ 0\leqslant t\leqslant t_1$
(2) Constant acceleration: $\ddot{f}_{w2}(t)=\alpha_1 t_1,\ t_1\leqslant t\leqslant t_2$
(3) Linear deceleration: $\ddot{f}_{w3}(t)=\ddot{f}_{w2}(t_2)-\alpha_2(t-t_2),\ t_2\leqslant t\leqslant t_3$
(4) Constant velocity: $\ddot{f}_{w4}(t)=0,\ \dot{f}_{w4}(t_3)=U_0,\ t_3\leqslant t\leqslant t_4$

Assuming that the first three time intervals are equal ($\Delta t_1=\Delta t_2=\Delta t_3$, $\Delta t_i=t_i-t_{i-1}$) and $\alpha_2=\alpha_1$, then α_1 is determined by assigning the constant velocity U_0 to $\dot{f}_{w4}(t_3)$. In the analysis to follow

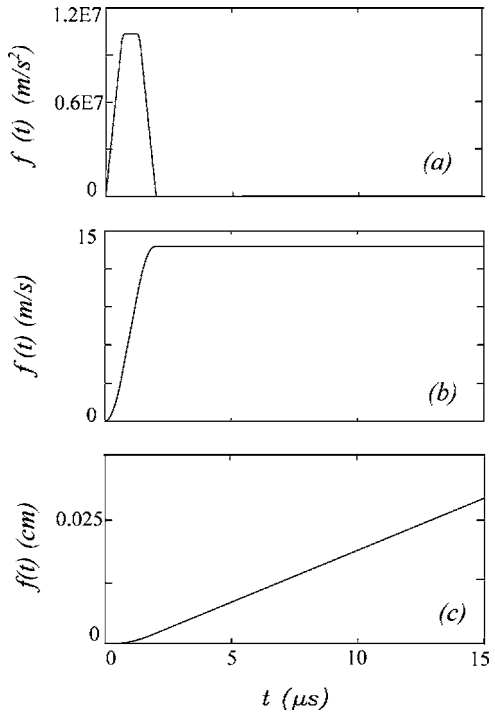


Fig. 3 Prescribed motion at footprint. (*a*) Acceleration; (*b*) velocity; (*c*) displacement.

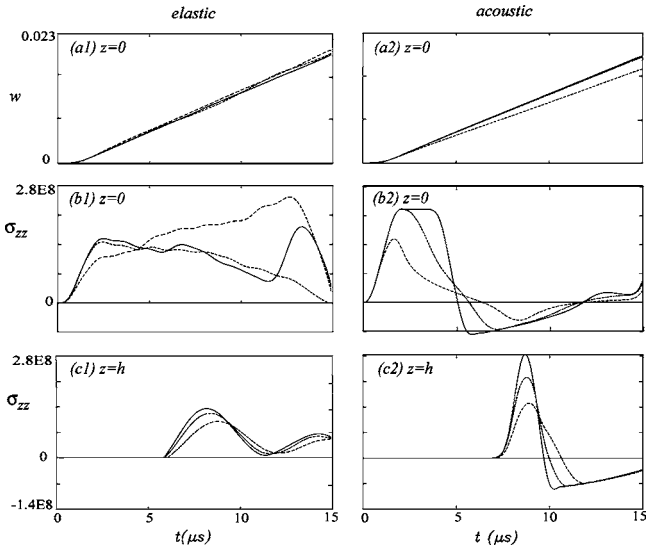


Fig. 4 Comparison of elastic and acoustic histories for prescribed acceleration: —, $r=0$; ---, $r=0.5r_p$; – – –, $r=0.9r_p$. Elastic model: (*a*1) $w(r,0)$, (*b*1) $\sigma_{zz}(r,0)$, (*c*1) $\sigma_{zz}(r,h)$; acoustic model: (*a*2) $w(r,0)$, (*b*2) $\sigma_{zz}(r,0)$, (*c*2) $\sigma_{zz}(r,h)$.

$$\Delta t_{1,3}\equiv\Delta t_1+\Delta t_2+\Delta t_3=2\ \mu\text{s},\quad U_0=14\ \text{m/s}(\equiv 46\ \text{ft./s}) \tag{19}$$

Figure 4 compares histories of the elastic and acoustic models from prescribed acceleration. Displacement at $z=0$ (Figs. 4(a1), 4(a2)) conforms to the prescribed value in Fig. 3(c). At $z=0$, Figs. 4(b1), 4(b2) compare histories of axial stress $-\sigma_{zz}$ from the elastic model to pressure p from the acoustic model. Peak stress, pulse duration, distribution of p over the footprint, and shape differ substantially between the two models. At $z=h$, Fig. 4(c1) and 4 (c2) compare $-\sigma_{zz}$ to p histories. There, magnitude and pulse width also differ. It is evident from this comparison that the two models differ appreciably in spite of the fact that in the elastic model $\nu=0.48$ is close to the transition value 1/2.

The difference between the two models in response from uniform prescribed pressure and prescribed acceleration is demonstrated in the example to follow. A uniform pressure pulse duplicating that at $r=0$ in Fig. 4(b2) is applied at $z=0$ (see Fig. 5(b)). The resulting histories of displacement w and pressure p at $z=h$ are shown in Figs. 5(a) and 5(c). Comparing histories in Figs. 4(a2) and 4(c2) to those in Figs. 5(a) and 5(c) reveals the sensitivity of response to p distribution over the footprint. Further evidence of this sensitivity appears when comparing p and w profiles at $z=0$ of the two cases. For prescribed acceleration p (Fig. 6(a1)) is not uniform while w (Fig. 6(b1)) is almost constant for $r<r_p$ and discontinuous at $r=r_p$. For prescribed pressure, p (Fig. 6(a2)) duplicates the external pulse while w (Fig. 6(b2)) increases with r reaching a maximum at $r=r_p$ with a discontinuity even stronger than that in Fig. 6(b1).

The parameters characterizing the applied acceleration profile are the final constant velocity U_0, and time interval $\Delta t_{1,3}$ of acceleration and deceleration to reach U_0 smoothly from rest. Figure 7 plots $p_{\max}$ against U_0 with $\Delta t_{1,3}$ as the parameter and vice versa. As expected, $p_{\max}$ is linear with U_0 (Figs. 7(a1) and 7(a2)). In contrast, $p_{\max}$ is nonlinear with $\Delta t_{1,3}$ (Figs. 7(b1), 7(b2)) following a relation $p_{\max}\propto U_0\Delta t_{13}^{-\alpha}$, where the α depends on z. $p_{\max}$ approaches a constant value as $\Delta t_{1,3}\rightarrow 0$ when slope of the acceleration profile in Fig. 3(a) becomes infinite. This is the limiting case when U_0 is applied instantaneously. For $\Delta t_{1,3}<3\ \mu$s, $p_{\max}$ goes through a transition when its value at $z=h$ exceeds that at $z=0$. The transition $\Delta t_{1,3}$ is almost independent of U_0.

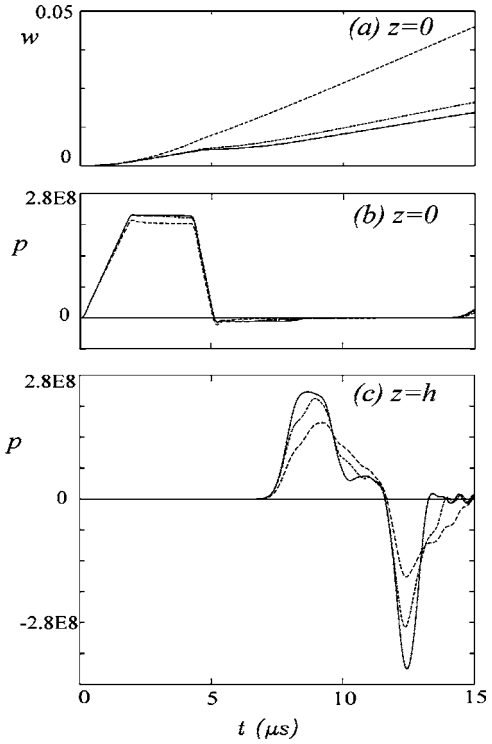


Fig. 5 Acoustic histories from prescribed pressure: —, $r=0$; ---, $r=0.5r_p$; – – –, $r=0.9r_p$. (*a*) $w(r,0;t)$; (*b*) $p(r,0;t)$; (*c*) $p(r,h;t)$.

Figure 8(*a*) shows deformed shapes at $t=8$ μs from the elastic model for $\nu=0.470$ and 0.495 keeping bulk modulus E_b the same. This requires expressing the constitutive law in terms of E_b and ν as in Eq. (A2*b*). Note that bulging of material near the perimeter is more pronounced for $\nu=0.495$ than for $\nu=0.470$. As ν approaches 1/2, material compressibility diminishes followed by a reduction in phase velocity along r near the free surface which delays propagation of the wavefront. In turn, conservation of volume and pressure release beyond the perimeter $r>r_p$ explains the formation and intensification of the bulge. Indeed, the closer ν gets to 1/2 the steeper the displacement gradient $\partial_r w$ along the perimeter reminiscent of the acoustic w profile in Fig. 6(*b*1). The effect on peak elastic stress $(\sigma_{zz})_{\max}$ of ν in the range $0.47\leqslant\nu\leqslant0.498$ is shown in Fig. 8(*b*). Although $(\sigma_{zz})_{\max}$ at $z=0$ is insensitive to ν for $\nu<0.495$, its value at $z=h$ drops by 76% due to a 6% increase in

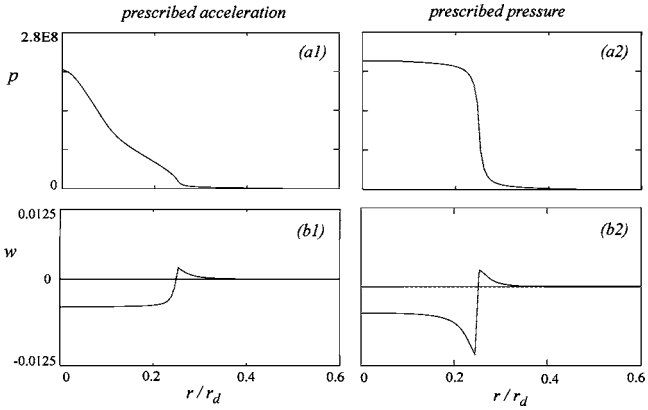


Fig. 6 Acoustic pressure and displacement profiles at $z=0$ and $t=4$ μs. (*a*1), (*b*1) Prescribed acceleration; (*a*2), (*b*2) prescribed pressure.

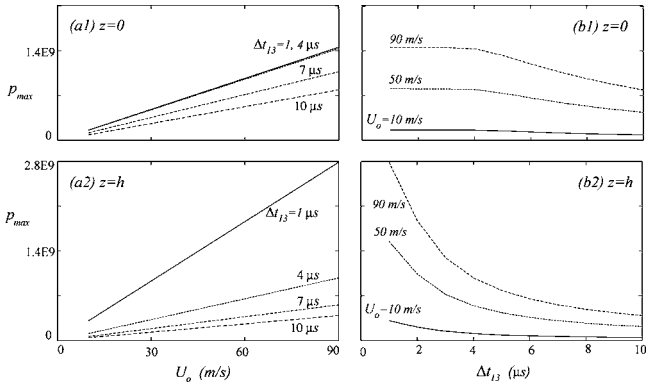


Fig. 7 Variation of p_{mx} with acceleration parameters U_0 and Δt_{13}. (*a*1), (*b*1) $z=0$; (*a*2), (*b*2) $z=h$.

ν. Unfortunately for attempts to use the acoustic model to capture elastic features, this makes the discrepancy between acoustic and elastic results even larger than that in Figs. 4(*c*1) and 4(*c*2).

Convergence of the elastic model with number of modes is paramount in the comparison between elastic and acoustic results. This is especially important since in the elastic model shear drops modal frequencies substantially (see Fig. 2). A larger modal set in the elastic model may be needed for its results to agree with the acoustic model that includes volumetric modes only. To verify convergence of the elastic model, histories from the analysis that produced results in Fig. 4(*b*1) and Fig. 8(*b*) are compared to those from the finite volume model employed by El-Raheb [15] that couples projectile and disk with 40,000 nodes. Properties and geometry of the projectile are

$$E_p = 1.21\times10^{11}\ \text{dyn/cm}^2\ (=1.76\times10^6\ \text{lb./in.}^2),$$

$$\rho_p = 1\ \text{g/cm}^3\ (=9.3\times10^{-5}\ \text{lb. s}^2/\text{in.}^4),\quad \nu_p = 0.3$$

$$r_p = 6.35\ \text{mm}\ (=0.25\ \text{in.}),\quad h_p = 25.4\ \text{mm}\ (=1\ \text{in.}),$$

$$U_p = 20\ \text{m/s}\ (=65\ \text{ft./s})$$

$$c_{bp} = (E_p(1-\nu_p)/((1+\nu_p)(1-2\nu_p)\rho_p))^{1/2}$$
$$= 4.1\ \text{km/s}\ (=1.6\times10^5\ \text{in./s})$$

r_p, h_p are projectile radius and length, U_p is striking velocity, and c_{bp} is dilatational speed of sound. Properties of gelatin are given in (13*a*) and (13*b*). Based on the acoustic impedances (ρc_b) of projectile and gelatin, the velocity of gelatin at the footprint following impact is approximately $U_0=14$ m/s (=45 ft./s). Histories of axial displacement w at the footprint from the two models coincide (Figs. 9(*a*1), 9(*a*2)) since the asymptotic velocity U_0 at the footprint is the same for both models. Figures 9(*b*1) and 9(*b*2) compare histories of axial stress σ_{zz} at the footprint from the two models. In the finite volume model, the drop in σ_{zz} 4 μs after impact (Fig. 9(*b*2)) corresponds to $t_{pr}=c_{bp}/2r_p$ the arrival time at

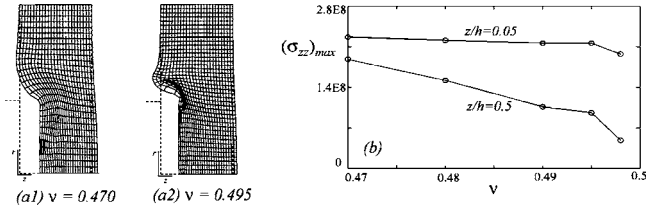


Fig. 8 Effect of Poisson ratio ν on (*a*) deformation snapshots at $t=8$ μs: (*a*1) $\nu=0.470$, (*a*2) $\nu=0.495$; (*b*) variation of peak stress with ν

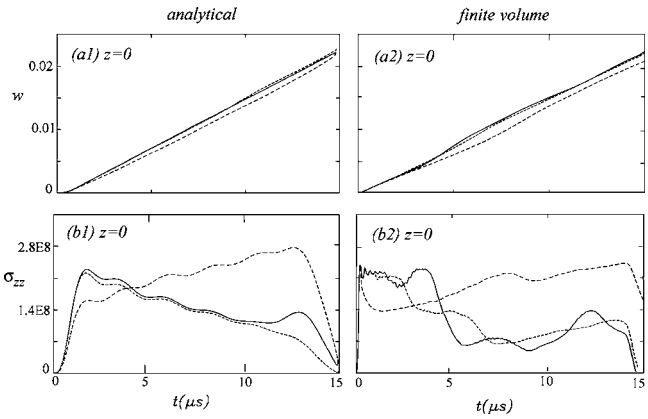


Fig. 9 Comparison of analytical and finite volume elastic models: —, $r=0.02r_p$; ---, $r=0.5r_p$; – – –, $r=0.96r_p$. (*a*1), (*b*1) *w*, σ_{zz} analytical.

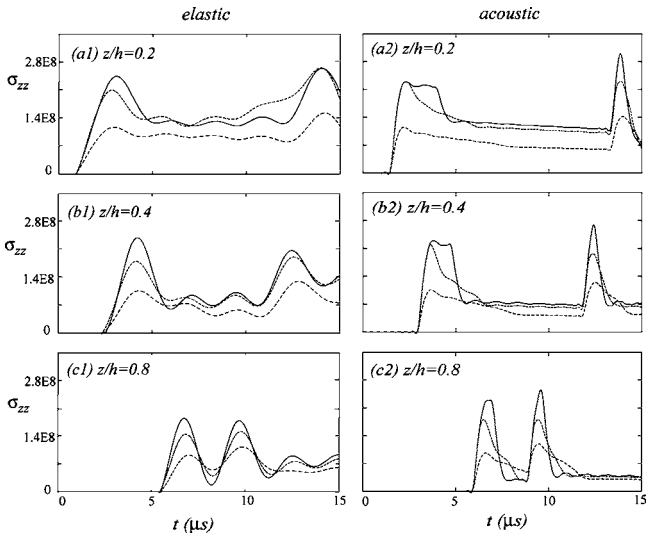


Fig. 11 Comparison of σ_{zz} histories from elastic and acoustic models with prescribed pressure: —, $r=0$; ---, $r=0.5r_p$; – – –, $r=0.9r_p$

$r=0$ of tensile reflections from the projectile's lateral boundary. This is evidenced by the deviation from linearity of the w histories at t_{pr} in Fig. 9(a2). In general, magnitude and shape of the σ_{zz} histories agree suggesting convergence of the analytical elastic model.

For prescribed uniform pressure, w histories from elastic and acoustic models agree (Figs. 10(b1) and 10(b2)) except at the footprint $z=0$ (Figs. 10(a1) and 10(a2)). In Fig. 11, the lead pulse in the σ_{zz} histories from the two models is followed by a plateau with lower magnitude. The wave reflected from the constrained face at $z=h$ appears as a peak following the plateau. In the elastic model,

(i) Risetime is longer;
(ii) History is modulated by a periodic oscillation;
(iii) Magnitude of the reflection dip is reduced.

For prescribed uniform pressure, the two models agree better than for prescribed acceleration implying that mismatch between the two models increases with magnitude of shear stress in the elastic model. Indeed, near the perimeter of the footprint shear stress is lower for prescribed uniform pressure than it is for prescribed acceleration because in the later pressure distribution is not uniform (Ref. [15]).

5 Conclusion

Acoustic wave propagation in a weak layer is treated adopting both a modal and a finite difference approach. The acoustic equation derives from the elastodynamic equations when shear stress vanishes. Two types of excitations are considered at the boundary, prescribed pressure, and prescribed acceleration. In the modal approach, the external excitation is modeled by the static-dynamic superposition method. Noteworthy results are

(1) Acoustic histories from the modal and finite difference approaches coincide.
(2) For prescribed acceleration, histories from the acoustic and elastic models disagree both in magnitude and shape because the resulting pressure is not uniform. However the two models show agreement for prescribed uniform pressure because shear stress is reduced.
(3) Employing the elastic model reveals that remote from the footprint $(\sigma_{zz})_{\max}$ drops sharply as ν approaches 1/2 making the discrepancy between acoustic and elastic results even larger.
(4) Convergence of the elastic model with number of modes is verified by comparing its histories with those from a finite volume model coupling projectile and disk.
(5) For prescribed acceleration at the boundary, rise time in pressure history is proportional to $\Delta t_{1,3}$ while $p_{\max}$ is proportional to $U_0 \Delta t_{13}^{-\alpha}$.
(6) Histories from prescribed pressure and prescribed acceleration differ because of nonuniform pressure distribution over the footprint.
(7) For $\Delta t_{1,3} < (\Delta t_{1,3})_T$, $p_{\max}$ goes through a transition when its value at the boundary $z=h$ exceeds that at the footprint $z=0$. $(\Delta t_{1,3})_T$ is a function of E_b and ρ but is almost independent of U_0.

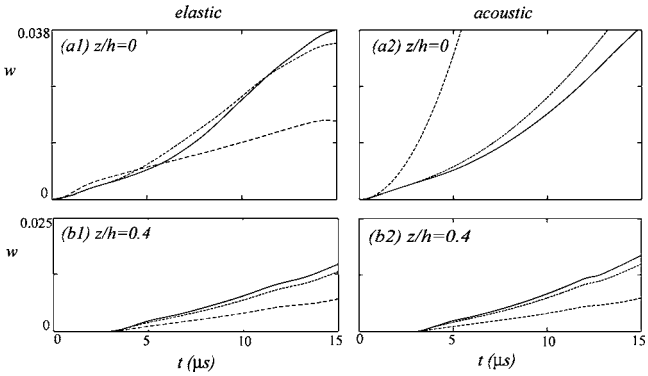


Fig. 10 Comparison of *w* histories from elastic and acoustic models with prescribed pressure: —, $r=0$; ---, $r=0.5r_p$; – – –, $r=0.9r_p$

Acknowledgment

This work was supported by a grant from DARPA, executed by the U.S. Army Medical Research and Materiel Command/TATRC Contract No. W81XWH-04-C-0084.

Appendix: Acoustic Equation in the Limit of Elasto-Dynamic Equations

Consider the linear axisymmetric elasto-dynamic equations in cylindrical coordinates

$$\partial_r \sigma_{rr} + (\sigma_{rr} - \sigma_{\theta\theta})/r + \partial_z \tau_{rz} = \rho \partial_{tt} u$$

$$\partial_z \sigma_{zz} + \partial_r \tau_{rz} + \tau_{rz}/r = \rho \partial_{tt} w \tag{A1}$$

where $(\sigma_{rr}, \sigma_{\theta\theta}, \sigma_{zz}, \tau_{rz})$ are radial, circumferential, axial, and shear stresses, and (u, w) are radial and axial displacements. Bulk modulus E_b relates average normal stress σ_V to volumetric strain ε_V

$$\sigma_V = E_b \varepsilon_V \equiv \rho c_b^2 \varepsilon_V, \quad E_b = (3\lambda + 2\mu)/3 \equiv E/(3(1-2\nu))$$

$$\sigma_V = (\sigma_{rr} + \sigma_{\theta\theta} + \sigma_{zz})/3$$

$$\varepsilon_V = \varepsilon_{rr} + \varepsilon_{\theta\theta} + \varepsilon_{zz} \equiv \nabla \cdot \mathbf{u} = \partial_r u + u/r + \partial_z w \tag{A2a}$$

where (λ, μ) are *Lame*' constants and c_b is bulk speed of sound. In terms of E_b and ν, the constitutive law takes the form

$$\sigma_{ij} = \frac{3\nu}{(1+\nu)} E_b \varepsilon_V \delta_{ij} + \frac{3(1-2\nu)}{(1+\nu)} E_b \varepsilon_{ij} \tag{A2b}$$

As $\nu \to 1/2$, $\sigma_{ij} \to \sigma_V = E_b \varepsilon_V$ recovering the bulk relation in (A2*a*)

$$\nu \to 1/2 \Rightarrow \tau_{rz} = 0, \quad \sigma_{rr} = \sigma_{\theta\theta} = \sigma_{zz} \equiv -p_d \tag{A3}$$

where δ_{ij} is Dirac's delta function. Substituting (A3) in (A1) produces the linear Euler equation

$$\rho \partial_{tt} \mathbf{u} = -\nabla p_d \tag{A4}$$

where $\mathbf{u}$ is the displacement vector. For a homogeneous fluid, conservation of mass takes the form

$$\partial_t \rho + \rho \partial_t (\nabla \cdot \mathbf{u}) = 0 \tag{A5}$$

The equation of state is

$$\frac{dp_d}{d\rho} = c_b^2 \tag{A6}$$

implying that

$$\partial_t p_d = c_b^2 \partial_t \rho \tag{A7}$$

Unlike the elastic solid where deviatoric or shear stresses contribute to material stiffness and reversible strain energy, in a viscous fluid these stresses are dissipative and irreversible. They are related to acoustic velocity by a constitutive law resembling that of an elastic solid

$$\begin{aligned} \tau_{ij} &= (\zeta - 2/3\,\eta)\delta_{ij}\partial_t \varepsilon_{ll} + \eta \partial_t \varepsilon_{ij} \\ &= (\zeta - 2/3\,\eta)\delta_{ij}\partial_t \partial_{x_l} u_l + \eta \partial_t (\partial_{x_i} u_j + \partial_{x_j} u_i) \end{aligned} \tag{A8}$$

x_i, x_j are independent variables and $(\zeta - 2/3\,\eta)$ and η are coefficients of viscosity for dilatational and deviatoric stains (see Landau and Lifshitz [17], p. 48). Equation (A8) resembles the constitutive relation (A2*b*) where $3\nu/(1+\nu)E_b$ and $3(1-2\nu)/(1+\nu)E_b$ are replaced by $(\zeta - 2/3\,\eta)$ and η. The linearized Navier-Stokes equations simplify to

$$\rho \partial_{tt} \mathbf{u} = -\nabla p_d + \partial_t [(\zeta - 1/6\,\eta)\nabla(\nabla \cdot \mathbf{u}) + (\eta/2)\nabla^2 \mathbf{u}] \tag{A9}$$

Conservation of mass and the equation of state are given by (A5) and (A6). Substituting for $\partial_t \rho$ from (A7) into (A5) yields

$$\partial_t [p_d + \rho c_b^2 (\nabla \cdot \mathbf{u})] = 0 \tag{A10}$$

Equation (A10) is the time derivative of (A2*a*) with σ_V replaced by $-p$. For a nonviscous fluid, taking the divergence of (A4), then eliminating $\mathbf{u}$ using (A10) determines the acoustic equation

$$(\partial_{rr} + 1/r\partial_r + \partial_{zz}) p_d - 1/c_b^2 \partial_{tt} p_d = 0 \tag{A11}$$

Equation (A11) is purely hyperbolic nondispersive.

For a viscous fluid, adopting the procedure that led to (A11) on (A9) and assuming that $\zeta = 1/6\,\eta$ yields the approximate viscous acoustic equation

$$(1 + \tilde{\nu}/c_b^2 \partial_t)(\partial_{rr} + 1/r\partial_r + \partial_{zz}) p_d - 1/c_b^2 \partial_{tt} p_d = 0 \tag{A12}$$

where $\tilde{\nu} = \eta/(2\rho)$ $(\mathrm{cm}^2/\mathrm{s})$ is kinematic viscosity.

References

[1] Theil, F., 1998, "Young-Measure Solutions for a Viscoelastically Damped Wave Equation with Nonmonotone Stress-Strain Relation," Arch. Ration. Mech. Anal., **144**, pp. 47–78.

[2] Yserentant, H., 2001, "The Propagation of Sound in Particle Models of Compressible Fluids," Numer. Math., **88**, pp. 581–601.

[3] Yserentant, H., 1997, "A Particle Model of Compressible Fluids," Numer. Math., **76**, pp. 111–142.

[4] Sina, K., and Khashayar, M., 2002, "Analytical Solution of Wave Equation for Arbitrary Non-Homogeneous Media," Proceedings of SPI, The International Society of Optical Engineering 4772, pp. 25–36.

[5] Sujith, R., Bala Subrahmanyam, P. T., and Lieuwen, P., 2003, "Propagation of Sound in Inhomogeneous Media: Exact Solutions in Curvilinear Geometries," ASME J. Vibr. Acoust., **125**, pp. 133–136.

[6] Hamdi, S., Enright, W., Schiesser, W., and Gottlieb, J., 2003, "Exact Solutions of the Generalized Equal Width Wave Equation," Lect. Notes Comput. Sci., **266**, pp. 725–734.

[7] Yang, D., 1994, "Grid Modification for the Wave Equation with Attenuation," Numer. Math., **67**, pp. 391–401.

[8] Narayan, J., 1998, "2.5-D Numerical Simulation of Acoustic Wave Propagation," Pure Appl. Geophys., **151**, pp. 47–61.

[9] Schemann, M., and Bornemann, F., 1998, "An Adaptive Rothe Method for the Wave Equation," Computing and Visualization in Science, **3**, pp. 137–144.

[10] Bailly, C., and Juve, D., 2000, "Numerical solution of acoustic propagation problems using linearized Euler equations," AIAA J., **38**, pp. 22–29.

[11] Wagner, G., Wenzel, M., and Dumont, W., 2001, "Numerical treatment of acoustic problems with the hybrid boundary element method," Int. J. Solids Struct., **38**, pp. 10–13; , **38**, pp. 1871–1888.

[12] Gaul, L., and Wenzel, M., 2001, "Acoustic calculations with the hybrid boundary element method in the time domain," Eng. Anal. Boundary Elem., **25**, pp. 259–265.

[13] Mehdizadeh, O., and Paraschivoiu, M., 2003, "Investigation of a Three-Dimensional Spectral Element Method for Helmholtz's Equation," Lect. Notes Comput. Sci., **2668**, pp. 819–825.

[14] Berry, J., and Naghdi, P., 1956, "On the Vibration of Elastic Bodies Having Time Dependent Boundary Conditions," Q. Appl. Math., **14**, pp. 43–50.

[15] El-Raheb, M., 2004, "Wave Propagation in a Weak Viscoelastic Layer Produced by Prescribed Velocity on the Boundary," J. Sound Vib., **275**(1–2), pp. 89–106.

[16] Eisler, R., 2003, (private communication, Mission Research Corporation, Laguna Hills, CA).

[17] Landau, L., and Lifshitz, E., 1959, *Fluid Mechanics*, 1st English Ed., Pergamon Press, Addison-Wesley, New York.

A Four-Parameter Iwan Model for Lap-Type Joints

Daniel J. Segalman
Fellow ASME
Sandia National Laboratories,
Albuquerque, NM 87185
e-mail: djsegal@sandia.gov

The constitutive behavior of mechanical joints is largely responsible for the energy dissipation and vibration damping in built-up structures. For reasons arising from the dramatically different length scales associated with those dissipative mechanisms and the length scales characteristic of the overall structure, this physics cannot be captured through direct numerical simulation (DNS) of the contact mechanics within a structural dynamics analysis. The difficulties of DNS manifest themselves either in terms of Courant times that are orders of magnitude smaller than that necessary for structural dynamics analysis or as intractable conditioning problems. The only practical method for accommodating the nonlinear nature of joint mechanisms within structural dynamic analysis is through constitutive models employing degrees of freedom natural to the scale of structural dynamics. In this way, development of constitutive models for joint response is a prerequisite for a predictive structural dynamics capability. A four-parameter model, built on a framework developed by Iwan, is used to reproduce the qualitative and quantitative properties of lap-type joints. In the development presented here, the parameters are deduced by matching joint stiffness under low load, the force necessary to initiate macroslip, and experimental values of energy dissipation in harmonic loading. All the necessary experiments can be performed on real hardware or virtually via fine-resolution, nonlinear quasistatic finite elements. The resulting constitutive model can then be used to predict the force/displacement results from arbitrary load histories. [DOI: 10.1115/1.1989354]

1 Introduction

The constitutive behavior of mechanical joints is largely responsible for the energy dissipation and vibration damping in built-up structures. For reasons arising from the dramatically different length scales associated with those dissipative mechanisms and the length scales characteristic of the overall structure, this physics cannot be captured through direct numerical simulation (DNS) of the contact mechanics within a structural dynamics analysis. The difficulties of DNS manifest themselves either in terms of Courant times that are orders of magnitude smaller than that necessary for structural dynamics analysis or as intractable conditioning problems.

The only practical method for accommodating the nonlinear nature of joint mechanisms within structural dynamic analysis is through constitutive models employing degrees of freedom natural to the scale of structural dynamics. In this way, development of constitutive models for joint response is a prerequisite for a predictive structural dynamics capability.

To be useful, such constitutive models must have the following properties:

- They must be capable of reproducing the important features of joint response—primarily the strongly nonlinear dependence of energy dissipation on the amplitude of harmonic loading and the significant but less dramatic decrease in joint stiffness with load amplitude.
- There must be a systematic method to deduce model parameters from joint-level experimental data or from very fine scale finite element modeling of the joint region.
- Model integration into a structural-level finite element code must be practical.

A framework that has potential for providing that balance is that due to Iwan [1,2]. Of his models, the most prominent has been the parallel system of Jenkins elements, sometimes called the parallel-series Iwan model. Such models consist of spring-slider units arranged in a parallel system as indicated in Fig. 1. Though Iwan introduced his constitutive models for metal elastoplasticity, they have since been used to model joints [3,4], and the work reported here addresses how that model form can be exploited in a systematic manner to capture the important responses of mechanical joints.

Mathematically, the constitutive form of the model is [1,6]

$$F(t) = \int_0^\infty k\tilde{\rho}(\tilde{\phi})[u(t) - \tilde{x}(t,\tilde{\phi})]d\tilde{\phi} \tag{1}$$

where u is the imposed displacement, $F(t)$ is the applied force, $\tilde{\rho}(\tilde{\phi})$ is the population density of Jenkins elements of strength $\tilde{\phi}$, k is the stiffness common to all of the Jenkins elements, and $\tilde{x}(t,\tilde{\phi})$ is the current displacement of sliders of strength $\tilde{\phi}$.

The slider displacements, $\tilde{x}(t,\tilde{\phi})$ evolve from the imposed system displacement, $u(t)$

$$\dot{\tilde{x}}(t,\tilde{\phi}) = \begin{cases} \dot{u} & \text{if } \|u - \tilde{x}(t,\tilde{\phi})\| = \tilde{\phi}/k \quad \text{and} \quad \dot{u}(u - \tilde{x}(t,\tilde{\phi})) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

It is assumed $\tilde{x}(0,\tilde{\phi}) = 0$ for all $\tilde{\phi}$. Note that Eq. (2) guarantees that $\|u - \tilde{x}(t,\tilde{\phi})\| \leq \tilde{\phi}/k$ at all times.

Noting that all Iwan models—even those without uniform k—are Masing models and all Masing models can be represented by Iwan models with uniform k, one sees that there is no loss in generality in employing identical stiffnesses among the Jenkins elements [15].

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, July 15, 2004; final revision, February 8, 2005. Associate Editor: O. M. O'Reilly. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

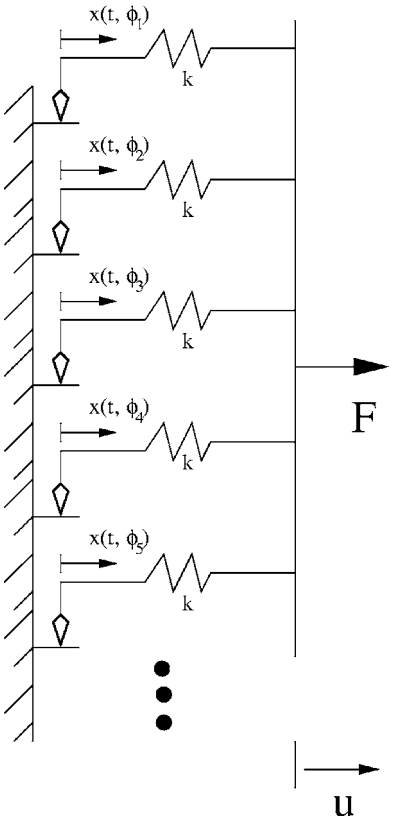


Fig. 1 A parallel-series Iwan system is a parallel arrangement of springs and sliders (Jenkins) elements

The parameter k can be removed from the above equations through the following changes of variable:

$$\phi = \tilde{\phi}/k \tag{3}$$

$$\rho(\phi) = k^2\tilde{\rho}(k\phi) \tag{4}$$

$$x(t,\phi) = \tilde{x}(t,k\phi) \tag{5}$$

Equations (1) and (2) now become

$$F(t) = \int_0^\infty \rho(\phi)[u(t) - x(t,\phi)]d\phi \tag{6}$$

and

$$\dot{x}(t,\phi) = \begin{cases} \dot{u} & \text{if } \|u - x(t,\phi)\| = \phi \quad \text{and} \quad \dot{u}[u - x(t,\phi)] > 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

We are now guaranteed that $\|u - x(t,\phi)\| \leq \phi$.

The new quantities have different dimensions than the originals. Though $\tilde{\phi}$ has dimensions of force, ϕ has dimensions of length. Similarly, $\tilde{\rho}$ has dimensions of 1/force but ρ has dimensions of force/length^2. The dimensions of the external loads and displacements applied to the joint remain unchanged.

Two overall parameters for the interface can be expressed in terms of the above integral system. The force necessary to cause macroslip (slipping of the whole interface) is denoted F_S, and the stiffness of the joint under small applied load (where slip is infinitesimal) is denoted K_T. Macroslip is characterized by every element sliding

$$u(t) - x(t,\phi) = \phi \tag{8}$$

for all ϕ, so Eq. (6) yields

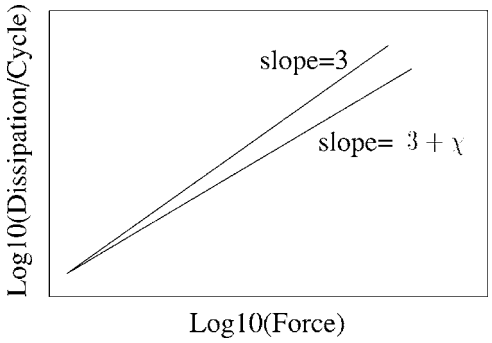


Fig. 2 The dissipation resulting from small amplitude harmonic loading tends to behave as a power of the force amplitude

$$F_S = \int_0^\infty \phi\rho(\phi)d\phi \tag{9}$$

Because no elements have slipped at the inception of loading, (at $t=0$) Eq. (6) yields

$$K_T = \int_0^\infty \rho(\phi)d\phi \tag{10}$$

2 Response to Small and Large Force

Experiments involving large monotonically applied forces can indicate the force necessary to initiate joint macroslip but, for reasons explained below, it is very difficult to obtain any other meaningful detail on joint response from quasistatic experiments. On the other hand, resonance experiments do enable the measurement of dissipation per cycle with reasonable precision even at relatively small loads [7,8]. Additionally, with proper calibration, those experiments can be used to obtain effective stiffness as a function of load amplitude. It is shown below how each sort of experimental data can be used to determine the parameters of a parallel-series Iwan model that can capture both quasistatic and dynamic behaviors.

2.1 Small Amplitude Oscillatory Loads. When a joint is subject to small amplitude oscillatory lateral loads, the dissipation appears to behave as a power of the amplitude of the applied load. Generally, the exponent of that relationship is a number lying between 2.0 and 3.0. Goodman [9] pointed out that the Mindlin solution [10,11] for the energy dissipation resulting from oscillatory lateral loads imposed on two spheres pushed together yields a power-law slope of 3.0 in the regime of small lateral loads.

In Fig. 2, that power-law slope is represented as $3+\chi$ where χ is a negative number of small magnitude ($-1 < \chi \leq 0$). Mathematically, this is expressed as the following:

$$D(F_0) = \upsilon F_0^{3+\chi} \tag{11}$$

where D is the dissipation per cycle resulting from a harmonic load of amplitude F_0 and χ and υ are selected so that Eq. (11) matches experimental data collected at small force amplitudes.

The dissipation per cycle associated with oscillatory displacements $[u(t)=u_0\sin(t)]$ applied to a Jenkins element of strength ϕ is the area within a parallelogram-shaped hysteresis loop having height 2ϕ and base $2(u_0-\phi)$. The dissipation integrated over all Jenkins elements is

$$D = \int_0^{u_0} 4[u_0 - \phi]\phi\rho(\phi)d\phi \tag{12}$$

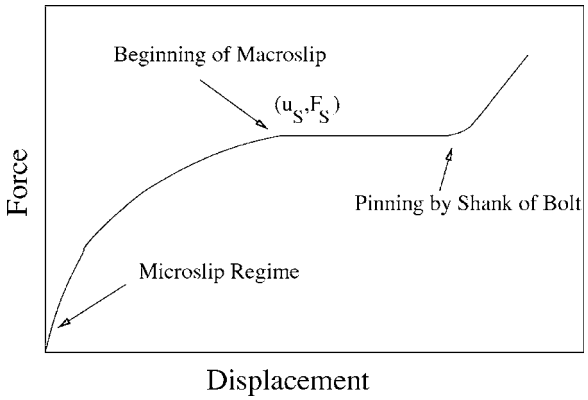


Fig. 3 The monotonic pull of a simple lap joint shows the force saturates at F_S as the displacement passes a critical value

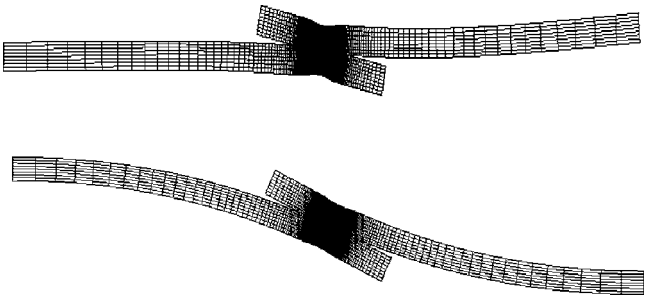


Fig. 4 The numerical predictions of a finely meshed system containing a single lap joint illustrate how interface displacements are obscured by the large compliance of the elastic response of the attached members. In the figure at top, both sides of the system are clamped, and stretched horizontally. In the figure below that, the left side is clamped and a zero-slope boundary condition is imposed on the right.

One major simplification made possible for histories where the displacement is bounded by a small value $u_{\max}$ (i.e. $|u(t)| < u_{\max}$) is that the integral of Eqs. (6) and (7) can be simplified to

$$F(t) = \int_0^{u_{\max}} \rho(\phi)[u(t) - x(t,\phi)]d\phi + u(t)\int_{u_{\max}}^{\infty} \rho(\phi)d\phi \quad (13)$$

$$= K_T u(t) + O(u_{\max}^2) \quad (14)$$

where the $O(\ )$ is the notation for quantities that are on the order of their argument as the argument goes to zero [12].

Expanding $\rho(\phi)$ as a

$$\rho(\phi) = \phi^{\chi}(a_0 + a_1\phi + a_2\phi^2 \cdots), \quad (15)$$

and substituting Eqs. (11), (14), and (15) into Eq. (12), and matching leading terms, we find

$$\rho(\phi) = \frac{\upsilon(K_T)^{3+\chi}(2+\chi)(3+\chi)}{4}\phi^{\chi} + O(\phi^{1+\chi}) \quad (16)$$

for small ϕ [6]. This provides something of the character of the population distribution that is necessary to yield a power-law type dissipation behavior.

2.2 Large Monotonic Loads. Let us consider large monotonic pulls ($0 < \dot{u}$). Equations (6) and (7) show that

$$F(t) = \int_0^{u(t)} \phi\rho(\phi)d\phi + u(t)\int_{u(t)}^{\infty} \rho(\phi)d\phi \quad (17)$$

from which Iwan derived

$$\frac{\partial^2 F(u)}{\partial u^2} = -\rho(u) \quad (18)$$

Because the second derivative of force cannot be measured with any resolution for most joints at small displacements, the above is at best only useful for large-displacement experiments.

Figure 3 sketches the monotonic force-displacement curve for a canonical lap joint. We anticipate that the force saturates at F_S and interface displacement u_S, corresponding to complete breaking of interface bonds.

Some comment should be made about why we have to guess at the force displacement curves for joints in structures such as we usually encounter. The key is that the interface mechanics cannot be viewed directly. The interface region is acted on by external loads conveyed through an elastic structure. Additionally, kinematic measurements are of the net displacements of that composite system—not of the joint. Particularly vexing is that the elastic subsystem is generally much more compliant than the interface until the interface has been forced into the vicinity of macroslip.

This insight is illustrated in Fig. 4, showing large elastic deformations taking place in a lap joint specimen long before serious slip occurs at the joint. As suggested by the drawing in Fig. 5, the force-displacement plot looks nearly linear until the applied force almost reaches that necessary to induce macroslip of the joint. The nearly linear region is dominated by the compliance of the elastic part of the system and the response of the interface is almost entirely obscured. Once the force is nearly sufficient to cause macroslip, it is the (near infinite) compliance of the interface which dominates. Though such experiments do inform us of the force necessary to initiate macroslip of the joint, one can conclude that they are not very useful to achieve resolution on the force-displacement response of the interface itself.

It should be said that for some structures for which the joints represent a major source of stiffness degradation of the structure, Levine and White [3] were able to deduce Iwan parameters by examining distortion of nominal frequency response curves as excitation frequency increased. This is an illustration of deducing joint properties indirectly through observation of the integrated behavior of the full structural dynamic response.

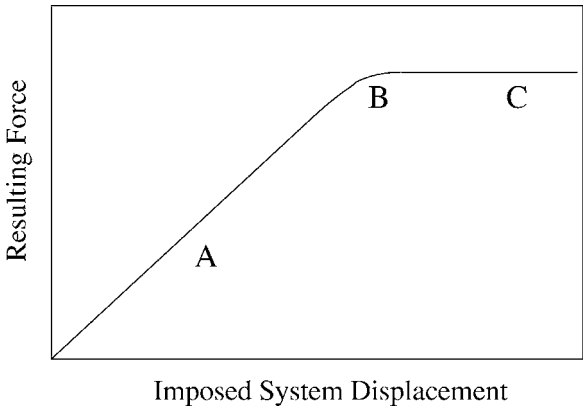


Fig. 5 Typically the force displacement conditions on elastic systems containing joints is as dominated at low loads by the elastic compliance (Region A). As the applied load approaches that necessary to initiate macroslip the force displacement curve begins to flatten (Region B). In macroslip the force-displacement curve is exactly flat (Region C). The only useful information about the joint available from such experiments is identification of the force necessary to initiate macroslip.

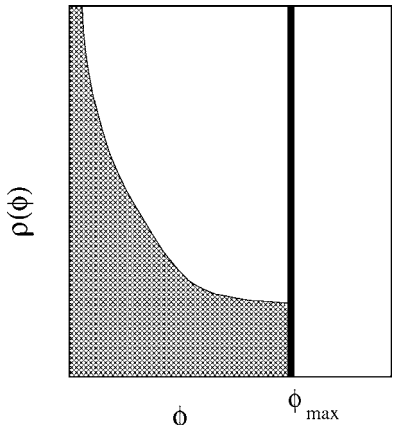


Fig. 6 A spectrum that is the sum of a truncated power law distribution and a Dirac delta function can be selected to satisfy asymptotic behavior at small and large force amplitudes

3 Truncated Power-Law Spectra

The above observations may be summarized:

- At small displacements the population density $\rho(\phi)$ behaves as in Eq. (16);
- Initiation of macroslip occurs at finite displacement, indicating that $\rho(\phi)$ has finite support;
- Equation (18) implies that a discontinuous slope of the force-displacement curve at the initiation of macroslip corresponds to a delta function in ρ;

motivating us to consider parallel Iwan systems having a power-law population distribution terminated by a Dirac delta

$$\rho(\phi) = R\phi^{\chi}[H(\phi) - H(\phi - \phi_{\max})] + S\delta(\phi - \phi_{\max}) \tag{19}$$

where $H(\,)$ is the Heaviside step function and $\phi_{\max}$ is numerically equal to u_S. The coefficient S accounts for the potential discontinuity in the slope of the force displacement curve at the inception of macroslip: S is the slope of the force-displacement curve just before macroslip. This form of population distribution is shown graphically in Fig. 6.

Substitution of Eq. (19) into Eq. (6) yields

$$F(t) = \int_0^{\phi_{\max}} [u(t) - x(t,\phi)]R\phi^{\chi}d\phi + S[u(t) - x(t,\phi_{\max})] \tag{20}$$

Referring to Eq. (9), the macroslip force for the system becomes

$$F_S = \int_0^{\phi_{\max}} \phi\rho(\phi)d\phi \tag{21}$$

$$= \frac{R\phi_{\max}^{\chi+2}}{(\chi+2)} + S\phi_{\max} \tag{22}$$

$$= \phi_{\max}\left(\frac{R\phi_{\max}^{\chi+1}}{\chi+1}\right)\left[\frac{\chi+1}{\chi+2} + \beta\right] \tag{23}$$

where

$$\beta = S \Big/ \left(\frac{R\phi_{\max}^{\chi+1}}{\chi+1}\right) \tag{24}$$

The dimensionless quantity β is the ratio of the joint stiffness due to the ϕ^{χ} term in Eq. (19) to that due to the δ function term. It will be shown below that β does strongly influence the shape of the log-log curve of dissipation versus force amplitude in harmonic loading and the shape of the force-displacement curve in monotonic loading.

One notes that R and S each have fractional dimension—not desirable qualities in constitutive parameters. On the other hand F_S does have the desirable features of having integral dimension and being measurable. One may invert Eq. (23) to solve for R and employ Eq. (24) to express S in terms of F_S

$$R = \frac{F_S(\chi+1)}{\phi_{\max}^{\chi+2}\left(\beta + \dfrac{\chi+1}{\chi+2}\right)} \tag{25}$$

and

$$S = \left(\frac{F_S}{\phi_{\max}}\right)\left(\frac{\beta}{\beta + \left(\dfrac{\chi+1}{\chi+2}\right)}\right) \tag{26}$$

Referring to Eq. (10), the interface stiffness could be computed as

$$K_T = \int_0^{\infty} \rho(\phi)d\phi = \frac{R\phi_{\max}^{\chi+1}}{(\chi+1)} + S = \frac{R\phi_{\max}^{\chi+1}}{(\chi+1)}(1+\beta)$$

$$= \frac{F_S(1+\beta)}{\phi_{\max}\left(\beta + \dfrac{\chi+1}{\chi+2}\right)} \tag{27}$$

The stiffness K_T can be estimated from resonance experiments in a manner described below. Because K_T can be estimated from experiment and involves no fractional units, it is also a desirable parameter.

Equation (27) can be solved for $\phi_{\max}$

$$\phi_{\max} = \frac{F_S(1+\beta)}{K_T\left(\beta + \dfrac{\chi+1}{\chi+2}\right)} \tag{28}$$

which is substituted into Eqs. (25) and (26) to define our model completely in terms of a preferred system of parameters: $\{F_S, K_T, \chi, \beta\}$. The first two of these are measurable and of integral dimension and the last two are dimensionless.

3.1 Monotonic Pull and Hysteresis. Though, as discussed above, performing meaningful monotonic pull experiments on a jointed structure is not feasible, it is worthwhile to express the force-displacement curve that would result if the experiment could be performed on the joint alone. This is particularly true because the original parallel-series Iwan model was presented in terms of the properties of such a curve.

Substituting the equations for monotonic pull into Eqs. (19) and (20), transforming to our preferred parameter set, and normalizing by F_S we find that

$$f(s) = F(s\phi_{\max})/F_S = \frac{s\left(\beta\chi + 2\beta + \chi + 2 - s^{\chi+1}\right)}{\beta\chi + 2\beta + \chi + 1} \tag{29}$$

where $s = u/\phi_{\max}$. Note that the only parameters of this dimensionless curve are χ and β, themselves dimensionless. The above function is plotted for $\chi = -1/2$ and for three values of β in Fig. 7. This figure shows that smaller values of β correspond to larger amounts of curvature prior to macro-slip, but greater discontinuity in slope as the imposed force approaches that necessary to initiate macroslip. In fact, examination of Eq. (20) shows that as $\beta \to \infty$ the response of the model to monotonic load approaches that of a single Jenkins element.

The smoothness—or lack of smoothness—of the transition from partial slip to macroslip is indicated by

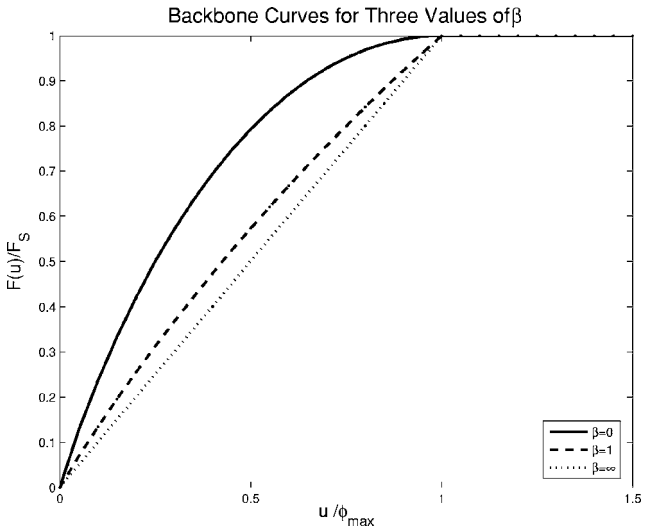


Fig. 7 The dimensionless force-displacement curve for monotonic pull for the four-parameter model for $\chi=-1/2$ and for three values of β

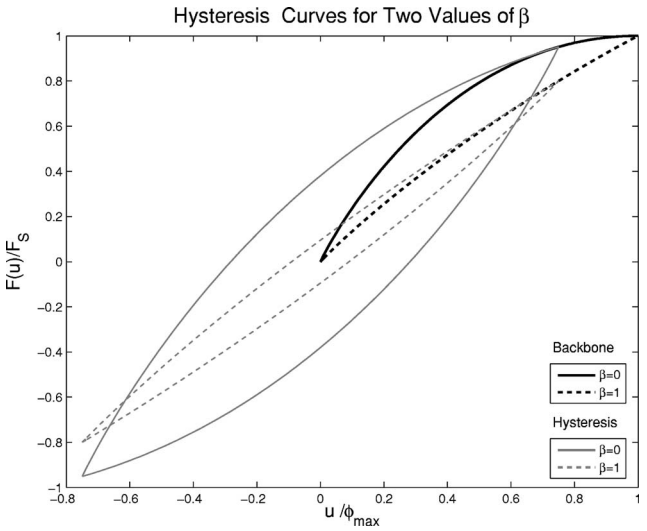


Fig. 8 The dimensionless hysteresis curves for the four-parameter model for $\chi=-1/2$ and for two values of β are shown in gray. The maximum and minimum extensions are set to 3/4 of that associated with the inception of macroslip. The corresponding curves for the unidirectional extension of a virgin material (backbone curves) are shown in black.

$$f'(s=1)=\frac{\beta(\chi+2)}{\beta\chi+2\beta+\chi+1} \tag{30}$$

and we see that the transition is sudden unless $\beta=0$.

As Peng [4] points out, parallel-series Iwan models satisfy the Masing conditions postulated for plastic materials. The first Masing condition, which actually is due to Masing, is that if the virgin material is monotonically deformed to some state and then subject to cyclic deformation at that same amplitude, the following symmetries will exist:

- The "forward" part of the hysteresis curve will look like the loading ("backbone") curve, but stretched along each of the displacement (strain) and force (stress) axes by a factor of two and translated to terminate at the tip of the backbone curve.
- The "return" part of the hysteresis curve will look like the forward part, but reflected along each of the displacement (strain) and force (stress) axes.

The second condition states that the equation of any hysteretic response curve is obtained by (1) applying the Masing hypothesis using the latest point of loading reversal and (2) requiring that if an active curve crosses a curve described in a previous cycle, the current curve follows that of the previous cycle. These two conditions constitute the extended Masing rules [5]. With these rules the response to any load history can be computed from the backbone curve and a record of all load reversals. (Among the ramifications is that K_T is twice the slope of the hysteresis curve just after reversal.)

Mathematically, if the joint is cycled between dimensionless extensions $u/u_{\max}=s_0$ and $-s_0$, then on the extensional branch, the force/displacement curve will behave as

$$f_e(s)=-f(s_0)+2f\left(\frac{s+s_0}{2}\right) \tag{31}$$

and will behave on the compression branch as

$$f_c(s)=f(s_0)-2f\left(\frac{s_0-s}{2}\right) \tag{32}$$

Using the extension curve for the four-parameter Iwan model of Eq. (29), maximum extension $s_0=3/4$, $\chi=1/2$, and two values of β, one obtains hysteresis loops indicated in Fig. 8.

3.2 Oscillatory Response. For reasons discussed above, quasistatic experiments alone provide nowhere near sufficient data to characterize joints. Good supplemental data can be obtained by dynamic resonance experiments [7,8]. In these experiments a jointed specimen is anchored on one end by a large sprung mass and is excited on the other end by an electromagnetic shaker acting through a force transducer. The shaker is driven to excite the system through resonance and to do so at various levels of force amplitude. Because this is a resonance experiment, the energy dissipation per cycle can be deduced from the force amplitude measured at the force transducer and the acceleration measured at the reaction mass [7].

Additionally, the joint stiffness at each force amplitude can be estimated in the following manner. From the resonance frequency and the known mass, one may deduce an effective compliance of the mechanical system. That compliance is the sum of the compliance of the force transducer, the compliance of the elastic part of the specimen, various attachment compliances, and the compliance of the joint. The sum of all but the joint compliance is deduced by performing a resonance experiment on a nearly identical unjointed specimen (machined from a single piece of metal). The compliance measured with the unjointed specimen is nearly always force independent and is assumed to be elastic. One subtracts that elastic compliance from the effective compliance measured with the jointed specimen to identify the compliance of the joint

$$1/K(F_0)=C_{\exp}(F)-C_E \tag{33}$$

where $K(F_0)$ is the effective joint stiffness at force amplitude F_0, $C_{\exp}(F)$ is the compliance deduced for the system with the jointed specimen, and C_E is the compliance deduced for the system with the unjointed specimen. The parameter K_T is the stiffness of the joint under zero load and is approximated in a practical manner by the $K(F_0)$ found at the lowest driving force employed.

Direct solution of Eqs. (7) and (20) for a problem specified by $F=F_0\sin(t)$ would involve the solution of a difficult nonlinear integral equation. An alternative approach is to specify $u(t)=u_0\sin(t)$ and then to solve for the resulting dissipation and peak force.

Noting that the maximum displacement of Jenkins elements of strength ϕ is $x(t,\phi)=u_0-\phi$, we observe that for $u_0\leq\phi_{\max}$ the

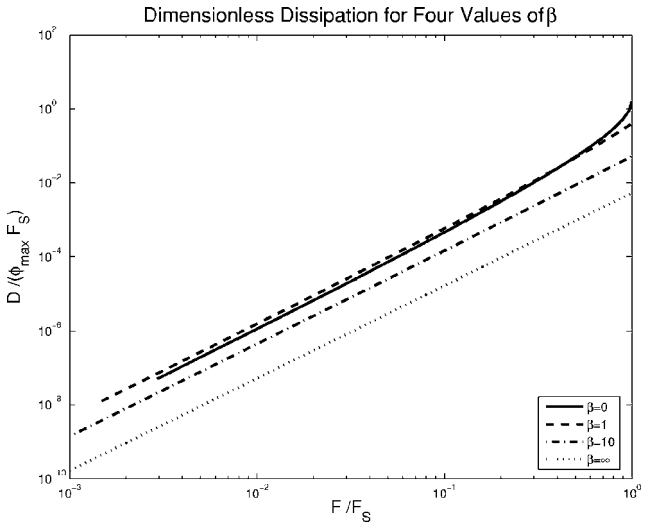


Fig. 9 The dimensionless dissipation per cycle as a function of normalized force for the four-parameter model for $\chi=-1/2$ and for three values of β

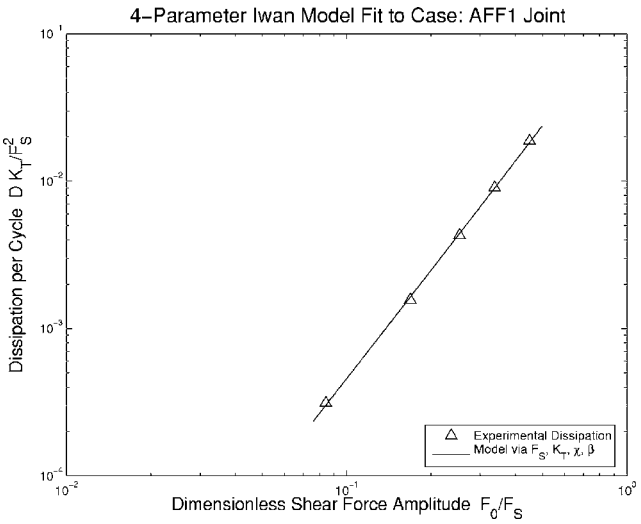


Fig. 10 Fit to dissipation data from a single leg of a component mass mockup. In this case, there is almost no curvature in the log-log plot of dissipation per cycle versus force amplitude, consistent with a power-law relationship. The dimensionless parameters employed were $\chi=-0.632$ and $\beta=3.68$.

dissipation per cycle of such elements is $4(u_0-\phi)\phi$. The net dissipation per cycle is exactly that given by Eq. (12). For the density function of Eq. (20) and for $u_0 \leqslant \phi_{\max}$, the dissipation per cycle is

$$D = \frac{4Ru_0^{\chi+3}}{(\chi+3)(\chi+2)} \tag{34}$$

$$= 4r^{\chi+3}\left(\frac{F_S^2}{K_T}\right)\left(\frac{(\beta+1)(\chi+1)}{\left(\beta+\dfrac{\chi+1}{\chi+2}\right)^2(\chi+2)(\chi+3)}\right) \tag{35}$$

where $r=u_0/\phi_{\max}$.

Next, observing that the maximum force seen in each cycle is that force current during the maximum displacement in the cycle

$$F_0 = \int_0^{u_0} \phi\rho(\phi)d\phi + u_0\int_{u_0}^{\phi_{\max}} \rho(\phi)d\phi \tag{36}$$

$$= u_0\left(S + R\frac{\phi_{\max}^{\chi+1}}{\chi+1}\right) - \frac{Ru_0^{\chi+2}}{(\chi+1)(\chi+2)} \tag{37}$$

Equation (37) is made nondimensional by dividing by F_S

$$F_0/F_S = r\frac{(\beta+1)-r^{\chi+1}/(\chi+2)}{\beta+(\chi+1)/(\chi+2)} \tag{38}$$

The experimental quantity most easily measured is dissipation (D) as a function of applied lateral load amplitude (F_0). Examination of Eqs. (35) and (38) permits us to observe the following:

- As $u_0 \to 0$, $F_0 \to r(\beta+1)\big/\beta+(\chi+1)/(\chi+2)=u_0K_T$, and $\partial\log(D)/\partial\log(F_0)\to\chi+3$.
- As $u_0\to\phi_{\max}$, $F_0\to F_S$, and $\partial\log(D)/\partial\log(F_0)\to(\chi+3)(\beta+(\chi+1)/(\chi+2))/\beta$.

A plot of dimensionless dissipation per cycle ($D\,K_T/F_S^2$) versus normalized force amplitude (F_0/F_S) is shown in Fig. 9 for $\chi=-1/2$ and three values of β. As expected, we see that for small force amplitudes ($F_0<F_S/2$) and all values of β the dissipation per cycle behaves as a power law. Also for values of β substantially greater than 1, the dissipation appears to have power-law behavior over the full range of force amplitude. Asymptotic analysis of Eqs. (35) and (38) as $\beta\to\infty$ shows that this should be the case.

Equation (37) permits us to express the secant stiffness at large amplitude oscillation

$$K(r) = \frac{F_0}{u_0} = K_T\left(1 - \frac{r^{\chi+1}}{(\chi+2)(\beta+1)}\right) \tag{39}$$

4 Identifying Parameters

Of the four parameters F_S, K_T, χ, and β, the first can be determined experimentally or estimated via statics and an assumed value for Coulomb friction. The second K_T might be estimated by extrapolating specimen stiffnesses obtained from resonance down to a zero-load state, though that approach magnifies the experimental uncertainty. It is the determination of χ and β which is particularly challenging.

The following approach is the one that the author has found most robust and most reliable at giving unique parameters. In this approach one attempts simultaneously to fit the dissipation data over the whole experimental range in a least-squares sense and to match the measured stiffness at some point in the midst of the range of applied load. This is achieved by employing inner and outer iteration loops. In the inner loop one employs Eqs. (35) and (38) F_S, and an estimate for K_T to deduce $\chi(K_T)$ and $\beta(K_T)$ that will reproduce the measured dissipation per cycle over the experimental load range. This is done most easily with a canned optimization tool, such as that available through MATLAB'S *fminsearch* tool [13]. (One subtlety is that each comparison of the four-parameter model with the experimental data requires a solution for $r(\chi,\beta,F_0/F_S)$.) The outer loop consists of an Newton iteration to identify K_T so that when the parameters F_S, K_T, $\chi(K_T)$, and $\beta(K_T)$ are employed in Eqs. (38) and (39), the stiffness $K(F_0)$ is correctly predicted at some load in the middle of the experimental data.

Figures 10 and 11 show reasonable fits to experimental dissipation when the automated method is exploited. In the first case, (Fig. 10) the fit is to data from a bolted leg of a component mass mockup (Fig. 12) of a mock W76 warhead substructure. We see that the dissipation data appear to lie on a straight line when plotted in a log-log manner. The parameters used to fit that data

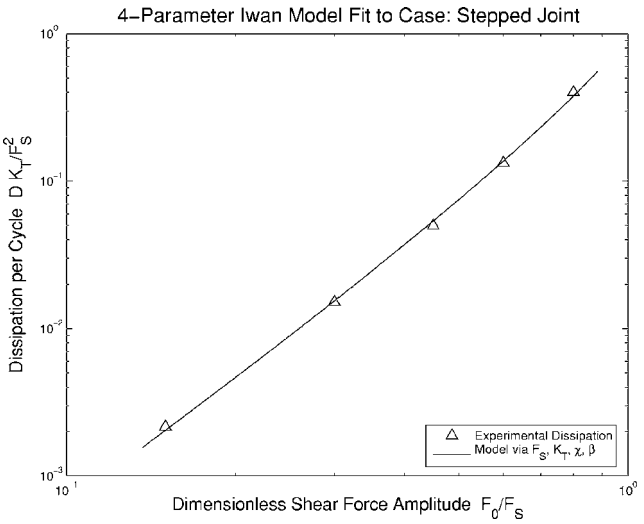


Fig. 11 Fit to dissipation data from a stepped specimen. In this case, there is appreciable curvature in the log-log plot of dissipation per cycle versus force amplitude. The dimensionless parameters employed were χ=−0.304 and β=0.613.

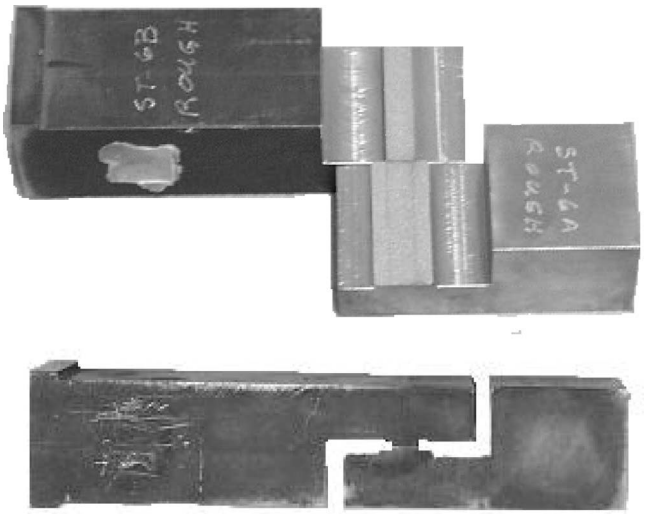


Fig. 13 A stepped specimen shows qualitatively different dissipation than a simple half-lap joint. The difference may be due to the near singular traction that develops at the edges of the contact patch.

are indicated in the figure caption. The value of F_S employed was deduced from statics and a postulated coefficient of friction of 0.5.

The log-log plot of dissipation versus lateral force for the mockup leg problem is nearly linear, but that is not the case for the stepped specimen shown in Fig. 13. Though this is a geometrically simple specimen, the dissipation data (shown in Fig. 11) shows substantially more curvature. The qualitatively different response might be due to the nearly singular normal tractions at the edges of the contact patch. In this case, there was no unjointed specimen constructed and it is impossible to know how much of the specimen compliance to ascribe to the joint, so K_T was arbitrarily assumed to be three times the specimen stiffness and only the inner iteration was employed. The joint parameters of the fitting curve are indicated in the figure caption.

In the case that all data has been collected well below the macroslip force, the dissipation behavior will appear to be power law in nature. A good fit to that dissipation can be obtained with any value for F_S substantially above the experimental loads and appropriate values of β and χ. Of course, if one plans to use a joint model for loads that may approach macroslip, it is appropriate to employ the best possible estimate for F_S.

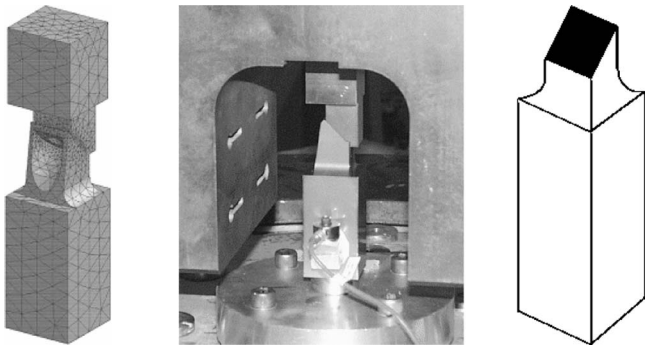


Fig. 12 The leg section of the mockup. To the left is a finite-element mesh of the full leg section, in the middle is the actual leg section in the test apparatus, and to the right is a sketch indicating the interface being modeled by the four-parameter model.

5 Discretization

Equations (6) and (7) are sufficient to solve for the force response of the above Iwan system once one has the constitutive parameters (F_S, K_T, χ, and β). It is useful to discretize the integral in Eq. (6) in the following manner. One breaks up the interval $(0, \phi_{\max})$ into N intervals whose lengths form a geometric series

$$\Delta\phi_{m+1} = \alpha\Delta\phi_m \quad \text{for all } m+1 < N \tag{40}$$

where α is a number slightly greater than one ($1<\alpha$) and $\phi_{\max}$ is determined from Eq. (28). That the sum of the intervals must be the whole interval

$$\sum_{m=1}^{N} \Delta\phi_m = \phi_{\max} \tag{41}$$

permits us to solve

$$\Delta\phi_m = \alpha^{m-1}\Delta\phi_1 \tag{42}$$

where

$$\Delta\phi_1 = \left[\phi_{\max}\frac{\alpha-1}{\alpha^N-1}\right] \tag{43}$$

We consider one sample point, characterized by slide strength ϕ_m, at the midpoint of each interval $\Delta\phi_m$. At that sample point, the evolution of $x_m(t)$ is computed per Eq. (7). For quadrature purposes, we refer to the coordinates of the left and right hand of each subinterval as $\phi_{l,m}$ and $\phi_{r,m}$, respectively.

The force is evaluated by a discrete version of Eq. (6).

$$F(t) = \sum_{m=1}^{N} F_m(t) + F_\delta(t) \tag{44}$$

where

$$F_m(t) = \begin{cases} R\dfrac{\phi_{r,m}^{2+\chi} - \phi_{l,m}^{2+\chi}}{2+\chi}\operatorname{sgn}[u(t)-x_m(t)] & \text{if } \|u(t)-x_m(t)\| = \phi_m \\ R\dfrac{\phi_{r,m}^{1+\chi} - \phi_{l,m}^{1+\chi}}{1+\chi}[u(t)-x_m(t)] & \text{if } \|u(t)-x_m(t)\| < \phi_m \end{cases} \tag{45}$$

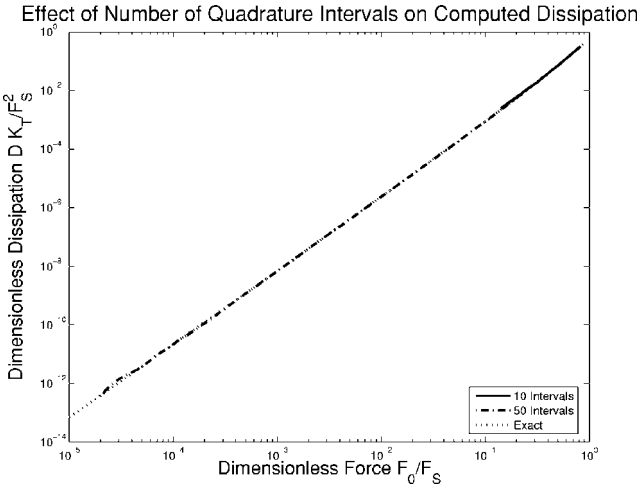


Fig. 14 Comparison of dissipation prediction of Eq. (35) with the quadrature of Eqs. (44) through (46)

$$F_\delta = S\phi_{\max}[u(t) - x_\delta(t)], \tag{46}$$

$\phi_\delta = \phi_{\max}$, and x_δ and each $x_m(t)$ evolve per Eq. (7). Appropriate values for R and S are determined from Eqs. (25) and (28) and Eqs. (26) and (28), respectively.

Note that the above quadrature reproduces the values for F_S in Eq. (22) exactly.

The discretization discussed here is illustrated by the results of a C++ code that imposes cyclic deformation on a four-parameter Iwan system and calculates the energy dissipation once steady state is achieved (always on the second cycle). Those numerical calculations are compared with the analytical expressions of Eq. (35). In Fig. 14 we see that for the amplitude range $0.1F_S < F_0 < F_S$ integration over the responses of as few as ten Jenkins elements ($N=10$) appears to be sufficient. Satisfactory results were achieved in all exercises using values of $\alpha=1.2$ and $N=50$. This is certainly overly conservative.

The question arises as to whether there is analytical guidance on how many Jenkins elements are necessary. The simplest criteria are

- In a monotonic pull, the stiffness degradation from K_T down to zero at macroslip should occur without too much discontinuity in the stiffness slope

$$\max_m\{\rho(\phi_m)\Delta\phi_m\} \ll K_T \tag{47}$$

  The maximum term in the above sequence is that associated with the last increment, so the condition is

$$R\left[\frac{\alpha^{N-1}\left(\frac{1+\alpha}{2}\right)-1}{\alpha^N-1}\right]^\chi \frac{\alpha^{N-1}(\alpha-1)}{\alpha^N-1}\phi_{\max}^{1+\chi} \ll K_T \tag{48}$$

  For large N, this condition becomes

$$\left(\frac{\chi+1}{\beta+1}\right)\left(\frac{(\alpha+1)^\chi(\alpha-1)}{2^\chi\alpha^{\chi+1}}\right) \ll 1 \tag{49}$$

  For ranges of χ and β considered above ($0<\chi<1$ and $0<\beta$), values of α on the order of 1.1 or 1.2 appear to cause Eq. (49) to be satisfied adequately.
- The sliding forces associated with the weakest element should slide at a force well below the smallest increment of force $\Delta F_{\min}$ between reversals that one wants to capture

$$R\frac{\phi_{r,1}^{2+\chi} - \phi_{l,1}^{2+\chi}}{2+\chi} \ll \Delta F_{\min} \tag{50}$$

This becomes a condition that

$$R\left(\phi_{\max}\frac{\alpha-1}{\alpha^N-1}\right)^{\chi+2}/(\chi+2) \ll \Delta F_{\min} \tag{51}$$

The quantity on the left goes as $\alpha^{-(\chi+2)N}$, explaining why Eq. (51) appears to be satisfied with fairly modest values of N.

6 Conclusion

The four-parameter model presented here appears to be capable of capturing the dissipation behavior found from harmonically loaded experiments on lap-type joints conducted so far. Further, the tools have been demonstrated to deduce the necessary model parameters with only modest effort.

Though the results presented here provide some reason for optimism, comparison with more sophisticated experiments should be made. Among those experiments could be multifrequency experiments such as discussed by Segalman [6] or random vibration experiments as performed by Smallwood for his hysteretic model [14]. Such experiments would be necessary to validate this four-parameter model in ranges and types of loadings other than those used to calibrate it.

The delta function in Eq. (20) causes the force-displacement curve to have a discontinuity in slope as the joint approaches macroslip. Because there does not appear to be any precise data on this transition, it is unclear whether the slope discontinuity is a problem. The discontinuity can be removed by replacing the delta function by an integrable singularity at the expense of adding one more parameter. A model of this sort has been suggested in Ref. [15].

Finally, one should note that constitutive equations of the sort developed here are "whole-joint" models. Such models may capture the response of the joint for the class of loads from which model parameters were deduced, but they give little insight into the microphysics taking place. Also, it is not yet clear how one would integrate joint data taken from qualitatively different load types—such as tension and moment loads—into models of this sort.

Over that longer term, more sophisticated approaches must be developed that better incorporate the distribution of tractions and displacements that develop dynamically around the joint and that do not presume a specific nature to the joint loading.

Acknowledgment

The author thanks the whole joints research team at Sandia National Laboratories for providing insight, conjecture, and experimental data that have led to the developments presented here. In particular the author thanks his colleague Danny Gregory for providing all the experimental data shown here and his colleague Todd Simmermacher for leading him by the hand through the mysteries of elementary MATLAB optimization. Dr. Simmermacher also deserves appreciation for reading multiple drafts of this document, making helpful suggestions each time. The author appreciates the collegial and supportive environment provided by his friends.

References

[1] Iwan, W. D., 1966, "A Distributed-Element Model for Hysteresis and Its Steady-State Dynamic Response," ASME J. Appl. Mech., **33**, pp. 893–900.
[2] Iwan, W. D., 1967, "On a Class of Models for the Yielding Behavior of Continuous and Composite Systems," ASME J. Appl. Mech., **34**, pp. 612–617.
[3] Levine, M. B., and White, C., 2001, "Microdynamic Analysis for Establishing Nanometric Stability Requirements of Jointed Precision Space Structures," Paper No. 325, *Proceedings of the International Modal Analysis Conference*, Kissimmee, FL, Society of Experimental Mechanics, Bethel, CT.
[4] Peng, C.-Y., 1988, Generalized Model Identification of Linear and Nonlinear

Dynamic Systems, Ph.D. thesis, California Institute of Technology, Pasadena, CA.
[5] Jayakumar, P., 1987, "Modeling and Identification in Structural Dynamics," Report No. EERL 87-01, California Institute of Technology, Pasadena, CA.
[6] Segalman, D. J., 2001, An Initial Overview of Iwan Modeling for Mechanical Joints, Report No. SAND2001-0811, Sandia National Laboratories, Albuquerque, NM.
[7] Gregory, D. L., Smallwood, D. O., Nusser, M. A., and Coleman, R. G., 1999, "Experimental Device to Study the Damping in Bolted Joints," *Proceedings of the 1999 ASME Design Engineering Technical Conferences*, Las Vegas, NV, ASME, New York.
[8] Gaul, L., and Lenz, J., 1997, "Nonlinear Dynamics of Structures Assembled by Bolted Joints," Acta Mech., **125**, pp. 169–181.
[9] Goodman, L. E., 1959, "A Review of Progress in Analysis of Interfacial Slip Damping," *Structural Damping*, papers presented at a colloquium on structural damping held at the ASME annual meeting in Atlantic City, NJ, Jerome E. Ruzicka, ed., ASME, New York, pp. 35–48.
[10] Mindlin, R. D., 1949, "Compliance of Elastic Bodies in Contact," ASME J. Appl. Mech., **16**, pp. 259–268.
[11] Mindlin, R. D., 1952, "Effects of an Oscillating Tangential Force on the Contact Surfaces of Elastic Spheres," *Proc. 1st US National Congress of Applied Mechanics*, ASME, New York, p. 203.
[12] Abromowitz, M., and Stegun, I. A., 1964, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*, Dover, New York, p. 1045.
[13] *Matlab Optimization Toolbox Guide*, 2002, The Mathworks, Inc., Natick, MA, p. 2.3
[14] Smallwood, D. O., Gregory, D. L., and Coleman, R. G., 2001, "A Three Parameter Constitutive Model for a Joint which Exhibits a Power Law Relationship Between Energy Loss and Relative Displacement," *72nd Shock and Vibration Symposium*, Destin, FL, SAVIAC, Columbia, MD.
[15] Segalman, D. J., and Starr, M. J., 2004, Relationships Among Certain Joints Constitutive Models, Report No. SAND2004-4321, Sandia National Laboratories, Albuquerque, NM.

Model Order Reduction of Viscously Damped Vibration Systems Using Accelerated Iterative Dynamic Condensation

Zu-Qing Qu
Michelin America R&D Corporation,
515 Michelin Road,
Greenville, SC 29605
e-mail: zuqing.qu@us.michelin.com

Panneer Selvam
Department of Civil Engineering,
University of Arkansas,
4190 Bell Engineering Center,
Fayetteville, AR 72701
e-mail: rps@engr.uark.edu

An accelerated iterative dynamic condensation method for the model order reduction of vibration systems with viscous damping is proposed. A group of governing equations for dynamic condensation matrix are derived from the eigenvalue equations defined in the state space. Two iterative schemes for solving these governing equations are provided. Based on different state space formulations, two more groups of governing equations are developed. A comparison of the present method with three iterative approaches proposed recently is provided. The present approach is implemented into two practical vibration systems, a tall building with one tuned mass damper and a floating raft isolation system. The results show that the proposed method has much higher accuracy than the other three approaches while the computational effort is almost the same.
[DOI: 10.1115/1.1993668]

1 Introduction

The development of increasing complex structures and mechanical systems demands sophisticated simulation techniques for design, control, and optimization. Due to the complex nature of these systems, the discrete models resulted from the finite element method are usually very large. If these models are directly used in the simulation, the computational effort will be extremely expansive. Therefore, the development of efficient model reduction methods for creating accurate low-order dynamic models has recently become a major goal of simulation and modeling research.

Various reduction techniques have been proposed to reduce the size of the full model or the dimension of the structural matrices involved in the formulations. Dynamic condensation, as an efficient technique for model reduction, has received much attention in recent years [1–15]. One of its important features is that the resulted, small size model is defined in the subspace of the original full-size physical space. The coordinates on which the reduced model is based have their inherent physical meanings. Thus, the reduced model is very convenient to be used in further dynamic analyses. Furthermore, one major by-product of the dynamic condensation is the partial eigensolution of the full model which is very useful in many dynamic problems. Therefore, this technique has been successfully applied to the eigenvalue problem analysis, test-analysis model correlation, active vibration control, finite element model updating, parameter identification, dynamic simulation, and modeling. The detailed review and complete list of publications on the dynamic condensation technique can be found in Qu [16].

Currently, most of the research in this area concentrates on undamped systems. However, the undamped assumptions are invalid in many situations. Examples of such cases are the structures made up of materials with different damping characteristics in different parts, structures equipped with passive and active control systems, and structures with layers of damping materials. The condensation methods for undamped models may be valid for the systems with proportional damping because this damping does not affect the eigenvectors of the full model [13]. If the damping is nonproportional, the accuracy of reduced models computed from these condensation methods is very low. For the iterative schemes, the dynamic characteristics of the reduced model do not converge to those of the full model even though these results may converge [13]. Therefore, the dynamic condensation approaches directly valid for viscous damping should be developed.

Based on modal reduction method [17], an extended modal reduction method was proposed by Kane and Torby [18] and applied to rotor dynamic problems in 1991. However, one difficulty of the method is that the interested eigenvectors should be calculated before the dynamic condensation. Three iterative methods for the model reduction of viscously damped systems were, respectively, proposed by Qu [13,19], Rivera et al. [20], and Qu et al. [21]. The dynamic characteristics of the reduced models resulted from these methods are usually very close to those of the full model in a given frequency range. Guyan condensation was used as an initial approximation in the iterative scheme proposed by Rivera et al. [20] to improve the accuracy. Very recently the iterative methods have been extended by Rao [22] to the models with unsymmetrical structural matrices.

The theory of viscously damped systems is reviewed in Sec. 2. In Sec. 3, the basic version of the governing equation of dynamic condensation matrix is derived from the eigenvalue problem in the state space. Three forms of the reduced model corresponding to the three state space formulations are defined for the eigenvalue problems and the dynamic equations of equilibrium. The relations between two different definitions of dynamic condensation matrix are shown using the complex modal superposition theory. Several other versions of the governing equations will also be provided. Because the governing equations are implicit, it is difficult to solve for the dynamic condensation matrix directly. The iterative forms of three of the governing equations are given in Sec. 4. Two iterative schemes are also provided. The iterative approaches previously developed are reviewed in Sec. 5. A detailed comparison of these approaches with the currently proposed method is made theoretically in this section and numerically in Sec. 6. The advantages of the proposed method will be shown in two practical applications.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, May 23, 2003; final revision, May 23, 2003. Associate Editor: A. A. Ferri. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

2 State Space Formulations of Viscously Damped Systems

2.1 Dynamic Equations of Equilibrium. The dynamic equations of equilibrium of a viscously damped system with n degrees of freedom (dofs) can be expressed in a matrix form as

$$\boldsymbol{M}\ddot{\boldsymbol{X}}(t)+\boldsymbol{C}\dot{\boldsymbol{X}}(t)+\boldsymbol{K}\boldsymbol{X}(t)=\boldsymbol{f}(t) \tag{1}$$

where $\boldsymbol{M}$, $\boldsymbol{C}$, and $\boldsymbol{K}\in\boldsymbol{R}^{n\times n}$ are the mass, damping and stiffness matrices of the full model. They are assumed to be symmetrical. $\boldsymbol{f}(t)\in\boldsymbol{R}^{n\times 1}$ is an external force vector. $\boldsymbol{X}(t)$, $\dot{\boldsymbol{X}}(t)$, and $\ddot{\boldsymbol{X}}(t)\in\boldsymbol{R}^{n}$ are the displacement, velocity, and acceleration response vector of the system.

Introduce the supplemental identity equations

$$-\boldsymbol{M}\dot{\boldsymbol{X}}(t)+\boldsymbol{M}\dot{\boldsymbol{X}}(t)=\boldsymbol{0} \tag{2}$$

Equations (1) and (2) can be rewritten in a compact form as

$$\boldsymbol{A}^{I}\boldsymbol{Y}(t)-\boldsymbol{B}^{I}\dot{\boldsymbol{Y}}(t)=\boldsymbol{F}^{I}(t) \tag{3}$$

in which $\boldsymbol{A}^{I}$, $\boldsymbol{B}^{I}\in\boldsymbol{R}^{2n\times 2n}$, $\boldsymbol{Y}(t)$, $\dot{\boldsymbol{Y}}(t)$, $\boldsymbol{F}^{I}(t)\in\boldsymbol{R}^{2n}$ are defined as

$$\boldsymbol{A}^{I}=\begin{bmatrix}\boldsymbol{K} & \boldsymbol{0}\\ \boldsymbol{0} & -\boldsymbol{M}\end{bmatrix},\quad \boldsymbol{B}^{I}=-\begin{bmatrix}\boldsymbol{C} & \boldsymbol{M}\\ \boldsymbol{M} & \boldsymbol{0}\end{bmatrix},\quad \boldsymbol{Y}(t)=\begin{Bmatrix}\boldsymbol{X}(t)\\ \dot{\boldsymbol{X}}(t)\end{Bmatrix},$$

$$\dot{\boldsymbol{Y}}(t)=\begin{Bmatrix}\dot{\boldsymbol{X}}(t)\\ \ddot{\boldsymbol{X}}(t)\end{Bmatrix},\quad \boldsymbol{F}^{I}(t)=\begin{Bmatrix}\boldsymbol{f}(t)\\ 0\end{Bmatrix} \tag{4}$$

$\boldsymbol{A}^{I}$ and $\boldsymbol{B}^{I}$ are referred to as state matrices. Clearly, the dynamic equations of equilibrium (3) are defined in the state space which is a combination of the displacement and velocity spaces.

If we introduce another supplemental identity, that is,

$$\boldsymbol{K}\dot{\boldsymbol{X}}(t)-\boldsymbol{K}\dot{\boldsymbol{X}}(t)=\boldsymbol{0} \tag{5}$$

the dynamic equations of equilibrium in Eq. (1) may be expressed in the state space as

$$\boldsymbol{A}^{II}\boldsymbol{Y}(t)-\boldsymbol{B}^{II}\dot{\boldsymbol{Y}}(t)=\boldsymbol{F}^{II}(t) \tag{6}$$

where the state matrices $\boldsymbol{A}^{II}$, $\boldsymbol{B}^{II}\in\boldsymbol{R}^{2n\times 2n}$, load vector $\boldsymbol{F}^{II}(t)\in\boldsymbol{R}^{2n}$ are defined as

$$\boldsymbol{A}^{II}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{K}\\ \boldsymbol{K} & \boldsymbol{C}\end{bmatrix},\quad \boldsymbol{B}^{II}=\begin{bmatrix}\boldsymbol{K} & \boldsymbol{0}\\ \boldsymbol{0} & -\boldsymbol{M}\end{bmatrix},\quad \boldsymbol{F}^{II}(t)=\begin{Bmatrix}\boldsymbol{0}\\ \boldsymbol{f}(t)\end{Bmatrix} \tag{7}$$

Differentiating both sides of Eq. (1) with respect to time once yields

$$\boldsymbol{M}\dddot{\boldsymbol{X}}(t)+\boldsymbol{C}\ddot{\boldsymbol{X}}(t)+\boldsymbol{K}\dot{\boldsymbol{X}}(t)=\dot{\boldsymbol{f}}(t) \tag{8}$$

The compact form of Eqs. (1) and (8) is expressed as

$$\boldsymbol{A}^{III}\boldsymbol{Y}(t)-\boldsymbol{B}^{III}\ddot{\boldsymbol{Y}}(t)=\boldsymbol{F}^{III}(t) \tag{9}$$

where the state matrices and vector are given by

$$\boldsymbol{A}^{III}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{K}\\ \boldsymbol{K} & \boldsymbol{C}\end{bmatrix},\quad \boldsymbol{B}^{III}=-\begin{bmatrix}\boldsymbol{C} & \boldsymbol{M}\\ \boldsymbol{M} & \boldsymbol{0}\end{bmatrix},\quad \boldsymbol{F}^{III}(t)=\begin{Bmatrix}\dot{\boldsymbol{f}}(t)\\ \boldsymbol{f}(t)\end{Bmatrix} \tag{10}$$

This is the third form of the dynamic equations of equilibrium in the state space. Due to the symmetry of mass, damping and stiffness matrices, the state matrices shown in Eqs. (4), (7), and (10) are all symmetrical.

2.2 Eigenvalue Problem. The eigenvalue problem corresponding to Eq. (1) is given by

$$\lambda_i^2\boldsymbol{M}\boldsymbol{\psi}_i+\lambda_i\boldsymbol{C}\boldsymbol{\psi}_i+\boldsymbol{K}\boldsymbol{\psi}_i=\boldsymbol{0},\quad (i=1,2,\ldots n) \tag{11a}$$

in which λ_i and $\boldsymbol{\psi}_i\in C^n$ are the ith eigenvalue (or complex frequency) and eigenvector. They generally have complex values. It can be simply verified that if λ_i and $\boldsymbol{\psi}_i$ is a solution of Eq. (11a), its complex conjugate pair, λ_i^* and $\boldsymbol{\psi}_i^*$, is also a solution of this equation, that is,

$$(\lambda_i^*)^2\boldsymbol{M}\boldsymbol{\psi}_i^*+\lambda_i^*\boldsymbol{C}\boldsymbol{\psi}_i^*+\boldsymbol{K}\boldsymbol{\psi}_i^*=\boldsymbol{0},\quad (i=1,2,\ldots n) \tag{11b}$$

where the superscript * denotes the complex conjugate. Therefore, the eigenvalue equation actually has $2n$ pairs of solution that appears in conjugate couples.

The compact forms of an eigenvalue problem of the n eigenpairs in Eqs. (11a) and (11b) are, respectively, expressed as

$$\boldsymbol{M}\boldsymbol{\Psi}\boldsymbol{\Omega}^2+\boldsymbol{C}\boldsymbol{\Psi}\boldsymbol{\Omega}+\boldsymbol{K}\boldsymbol{\Psi}=\boldsymbol{0} \tag{12a}$$

$$\boldsymbol{M}\boldsymbol{\Psi}^*(\boldsymbol{\Omega}^*)^2+\boldsymbol{C}\boldsymbol{\Psi}^*\boldsymbol{\Omega}^*+\boldsymbol{K}\boldsymbol{\Psi}^*=\boldsymbol{0} \tag{12b}$$

in which the eigenvector matrix $\boldsymbol{\Psi}$, $\boldsymbol{\Psi}^*\in C^{n\times n}$ and eigenvalue matrix (diagonal) $\boldsymbol{\Omega}$, $\boldsymbol{\Omega}^*\in C^{n\times n}$ are given by

$$\boldsymbol{\Psi}=[\boldsymbol{\psi}_1\ \boldsymbol{\psi}_2\ldots\boldsymbol{\psi}_n],\quad \boldsymbol{\Omega}=\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n) \tag{13a}$$

$$\boldsymbol{\Psi}^*=[\boldsymbol{\psi}_1^*\ \boldsymbol{\psi}_2^*\ldots\boldsymbol{\psi}_n^*],\quad \boldsymbol{\Omega}^*=\mathrm{diag}(\lambda_1^*,\lambda_2^*,\ldots,\lambda_n^*) \tag{13b}$$

The compact form of Eq. (12) is given by

$$\boldsymbol{M}\tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}^2+\boldsymbol{C}\tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}+\boldsymbol{K}\tilde{\boldsymbol{\Psi}}_U=\boldsymbol{0} \tag{14}$$

where the eigenvector matrix $\tilde{\boldsymbol{\Psi}}_U\in C^{n\times 2n}$ and eigenvalue or spectral matrix $\tilde{\boldsymbol{\Omega}}\in C^{2n\times 2n}$ are defined as

$$\tilde{\boldsymbol{\Psi}}_U=[\boldsymbol{\Psi}\ \boldsymbol{\Psi}^*],\quad \tilde{\boldsymbol{\Omega}}=\begin{bmatrix}\boldsymbol{\Omega} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Omega}^*\end{bmatrix} \tag{15}$$

Introduce the supplemental identity:

$$-\boldsymbol{M}\tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}+\boldsymbol{M}\tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}=\boldsymbol{0} \tag{16}$$

The eigenvalue Eq. (14) may be rewritten in the state space as

$$\boldsymbol{A}^{I}\tilde{\boldsymbol{\Psi}}=\boldsymbol{B}^{I}\tilde{\boldsymbol{\Psi}}\tilde{\boldsymbol{\Omega}} \tag{17}$$

The complex conjugate eigenvector matrix $\tilde{\boldsymbol{\Psi}}\in C^{2n\times 2n}$ has the form

$$\tilde{\boldsymbol{\Psi}}=\begin{bmatrix}\tilde{\boldsymbol{\Psi}}_U\\ \tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\Psi} & \boldsymbol{\Psi}^*\\ \boldsymbol{\Psi}\boldsymbol{\Omega} & \boldsymbol{\Psi}^*\boldsymbol{\Omega}^*\end{bmatrix} \tag{18}$$

The orthogonalities of the system are given by

$$\tilde{\boldsymbol{\Psi}}^T\boldsymbol{A}\tilde{\boldsymbol{\Psi}}=\tilde{\boldsymbol{\Omega}},\quad \tilde{\boldsymbol{\Psi}}^T\boldsymbol{B}\tilde{\boldsymbol{\Psi}}=\boldsymbol{I} \tag{19}$$

where $\boldsymbol{I}\in R^{2n\times 2n}$ is an identity matrix.

If the identities

$$\boldsymbol{K}\tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}-\boldsymbol{K}\tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}=\boldsymbol{0} \tag{20}$$

$$\boldsymbol{M}\tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}^3+\boldsymbol{C}\tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}^2+\boldsymbol{K}\tilde{\boldsymbol{\Psi}}_U\tilde{\boldsymbol{\Omega}}=\boldsymbol{0} \tag{21}$$

are introduced to Eq. (14), the eigenvalue equations in the state space become

$$\boldsymbol{A}^{II}\tilde{\boldsymbol{\Psi}}=\boldsymbol{B}^{II}\tilde{\boldsymbol{\Psi}}\tilde{\boldsymbol{\Omega}} \tag{22}$$

$$\boldsymbol{A}^{III}\tilde{\boldsymbol{\Psi}}=\boldsymbol{B}^{III}\tilde{\boldsymbol{\Psi}}\tilde{\boldsymbol{\Omega}}^2 \tag{23}$$

Equations (17), (22), and (23) are the three forms of an eigenvalue problem of viscously damped models.

3 Dynamic Condensation Matrix and Reduced Models of the Proposed Method

3.1 Basic Version of Governing Equation. If only the lowest m groups of eigenpairs are considered in Eq. (23), this equation can be rewritten as

$$\boldsymbol{A}^{III}\tilde{\boldsymbol{\Psi}}_m = \boldsymbol{B}^{III}\tilde{\boldsymbol{\Psi}}_m\tilde{\boldsymbol{\Omega}}^2_{mm} \tag{24}$$

In the dynamic condensation, the total dofs (n) of a full model are divided into the master dofs (masters) (m) and the slave dofs (slaves) (s). Based on the division, Eq. (24) can be expressed in a partitioned form as

$$\begin{bmatrix} \boldsymbol{A}^{III}_{mm} & \boldsymbol{A}^{III}_{ms} \\ \boldsymbol{A}^{III}_{sm} & \boldsymbol{A}^{III}_{ss} \end{bmatrix}\begin{bmatrix} \tilde{\boldsymbol{\Psi}}_{mm} \\ \tilde{\boldsymbol{\Psi}}_{sm} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}^{III}_{mm} & \boldsymbol{B}^{III}_{ms} \\ \boldsymbol{B}^{III}_{sm} & \boldsymbol{B}^{III}_{ss} \end{bmatrix}\begin{bmatrix} \tilde{\boldsymbol{\Psi}}_{mm} \\ \tilde{\boldsymbol{\Psi}}_{sm} \end{bmatrix}\tilde{\boldsymbol{\Omega}}^2_{mm} \tag{25}$$

where the submatrices are defined as

$$\boldsymbol{A}^{III}_{mm} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{K}_{mm} \\ \boldsymbol{K}_{mm} & \boldsymbol{C}_{mm} \end{bmatrix}, \quad \boldsymbol{A}^{III}_{ms} = (\boldsymbol{A}^{III}_{sm})^T = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{K}_{ms} \\ \boldsymbol{K}_{ms} & \boldsymbol{C}_{ms} \end{bmatrix},$$

$$\boldsymbol{A}^{III}_{ss} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{K}_{ss} \\ \boldsymbol{K}_{ss} & \boldsymbol{C}_{ss} \end{bmatrix} \tag{26a}$$

$$\boldsymbol{B}^{III}_{mm} = -\begin{bmatrix} \boldsymbol{C}_{mm} & \boldsymbol{M}_{mm} \\ \boldsymbol{M}_{mm} & \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{B}^{III}_{ms} = (\boldsymbol{B}^{III}_{sm})^T = -\begin{bmatrix} \boldsymbol{C}_{ms} & \boldsymbol{M}_{ms} \\ \boldsymbol{M}_{ms} & \boldsymbol{0} \end{bmatrix},$$

$$\boldsymbol{B}^{III}_{ss} = -\begin{bmatrix} \boldsymbol{C}_{ss} & \boldsymbol{M}_{ss} \\ \boldsymbol{M}_{ss} & \boldsymbol{0} \end{bmatrix} \tag{26b}$$

$$\tilde{\boldsymbol{\Psi}}_{mm} = \begin{bmatrix} \boldsymbol{\Psi}_{mm} & \boldsymbol{\Psi}^*_{mm} \\ \boldsymbol{\Psi}_{mm}\boldsymbol{\Omega}_{mm} & \boldsymbol{\Psi}^*_{mm}\boldsymbol{\Omega}^*_{mm} \end{bmatrix},$$

$$\tilde{\boldsymbol{\Psi}}_{sm} = \begin{bmatrix} \boldsymbol{\Psi}_{sm} & \boldsymbol{\Psi}^*_{sm} \\ \boldsymbol{\Psi}_{sm}\boldsymbol{\Omega}_{mm} & \boldsymbol{\Psi}^*_{sm}\boldsymbol{\Omega}^*_{mm} \end{bmatrix},$$

$$\tilde{\boldsymbol{\Omega}}^2_{mm} = \begin{bmatrix} \boldsymbol{\Omega}^2_{mm} & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{\Omega}^*_{mm})^2 \end{bmatrix} \tag{26c}$$

Expanding the second equation in the partitioned Eq. (25) produces

$$\boldsymbol{A}^{III}_{sm}\tilde{\boldsymbol{\Psi}}_{mm} + \boldsymbol{A}^{III}_{ss}\tilde{\boldsymbol{\Psi}}_{sm} = \boldsymbol{B}^{III}_{sm}\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}^2_{mm} + \boldsymbol{B}^{III}_{ss}\tilde{\boldsymbol{\Psi}}_{sm}\tilde{\boldsymbol{\Omega}}^2_{mm} \tag{27}$$

Equation (27) leads to

$$\tilde{\boldsymbol{\Psi}}_{sm} = (\boldsymbol{A}^{III}_{ss})^{-1}(\boldsymbol{B}^{III}_{sm}\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}^2_{mm} + \boldsymbol{B}^{III}_{ss}\tilde{\boldsymbol{\Psi}}_{sm}\tilde{\boldsymbol{\Omega}}^2_{mm} - \boldsymbol{A}^{III}_{sm}\tilde{\boldsymbol{\Psi}}_{mm}) \tag{28}$$

According to the definition [13], the dynamic condensation matrix of viscously damped systems is given by

$$\tilde{\boldsymbol{\Psi}}_{sm} = \boldsymbol{R}\tilde{\boldsymbol{\Psi}}_{mm} \tag{29}$$

The dynamic condensation matrix $\boldsymbol{R} \in R^{2s\times 2m}$ represents the relations of eigenvectors between the masters and slaves. Substituting Eq. (29) into both sides of Eq. (28) and then postmultiplying its both sides by the inverse of matrix $\tilde{\boldsymbol{\Psi}}^{-1}_{mm}$, the dynamic condensation matrix is obtained as

$$\boldsymbol{R} = (\boldsymbol{A}^{III}_{ss})^{-1}[(\boldsymbol{B}^{III}_{sm} + \boldsymbol{B}^{III}_{ss}\boldsymbol{R})\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}^2_{mm}\tilde{\boldsymbol{\Psi}}^{-1}_{mm} - \boldsymbol{A}^{III}_{sm}] \tag{30}$$

Equation (30) is the basic version of governing equation of dynamic condensation matrix.

3.2 Reduced Models. Using the dynamic condensation matrix, the eigenvector matrix $\tilde{\boldsymbol{\Psi}}_m$ of the full model may be expressed as

$$\tilde{\boldsymbol{\Psi}}_m = \begin{bmatrix} \tilde{\boldsymbol{\Psi}}_{mm} \\ \tilde{\boldsymbol{\Psi}}_{sm} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{R} \end{bmatrix}\tilde{\boldsymbol{\Psi}}_{mm} \equiv \boldsymbol{T}\tilde{\boldsymbol{\Psi}}_{mm} \tag{31}$$

where $\boldsymbol{T} \in R^{2n\times 2m}$ is referred to as the coordinate transformation matrix. Introducing Eq. (31) into both sides of Eq. (25) and premultiplying it by the transpose of the transformation matrix $\boldsymbol{T}$ lead to

$$\boldsymbol{A}^{III}_R\tilde{\boldsymbol{\Psi}}_{mm} = \boldsymbol{B}^{III}_R\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}^2_{mm} \tag{32}$$

$\boldsymbol{A}^{III}_R,\ \boldsymbol{B}^{III}_R \in \boldsymbol{R}^{2m\times 2m}$ are the state matrices of the reduced model. They are given by

$$\boldsymbol{A}^{III}_R = \boldsymbol{T}^T\boldsymbol{A}^{III}\boldsymbol{T} = \boldsymbol{A}^{III}_{mm} + \boldsymbol{R}^T\boldsymbol{A}^{III}_{sm} + \boldsymbol{A}^{III}_{ms}\boldsymbol{R} + \boldsymbol{R}^T\boldsymbol{A}^{III}_{ss}\boldsymbol{R} \equiv \Re(\boldsymbol{A}^{III},\boldsymbol{R}) \tag{33a}$$

$$\boldsymbol{B}^{III}_R = \boldsymbol{T}^T\boldsymbol{B}^{III}\boldsymbol{T} = \boldsymbol{B}^{III}_{mm} + \boldsymbol{R}^T\boldsymbol{B}^{III}_{sm} + \boldsymbol{B}^{III}_{ms}\boldsymbol{R} + \boldsymbol{R}^T\boldsymbol{B}^{III}_{ss}\boldsymbol{R} \equiv \Re(\boldsymbol{B}^{III},\boldsymbol{R}) \tag{33b}$$

Clearly, the state matrices of the reduced model depend upon the state matrices of full model and the dynamic condensation matrix $\boldsymbol{R}$. They are symbolically expressed as $\Re(\boldsymbol{A}^{III},\boldsymbol{R})$ and $\Re(\boldsymbol{B}^{III},\boldsymbol{R})$ respectively. Equation (32) is the eigenvalue equation of the reduced model. The corresponding orthogonality equations can be expressed in the state space as

$$\tilde{\boldsymbol{\Psi}}^T_{mm}\boldsymbol{A}^{III}_R\tilde{\boldsymbol{\Psi}}_{mm} = \tilde{\boldsymbol{\Omega}}^2_{mm}, \tag{34a}$$

$$\tilde{\boldsymbol{\Psi}}^T_{mm}\boldsymbol{B}^{III}_R\tilde{\boldsymbol{\Psi}}_{mm} = \boldsymbol{I} \tag{34b}$$

Equations (32) and (34) show that the reduced model $(\boldsymbol{A}^{III}_R,\boldsymbol{B}^{III}_R)$ computed from Eq. (33) may exactly keep the given m eigenpairs of the full model if the dynamic condensation matrix, as a result of the coordinate transformation matrix, is exact.

Similarly, if the transformation defined in Eq. (31) is, respectively, performed on both sides of Eqs. (17) and (22), we have

$$\boldsymbol{A}^{I}_R\tilde{\boldsymbol{\Psi}}_{mm} = \boldsymbol{B}^{I}_R\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}_{mm} \tag{35}$$

$$\boldsymbol{A}^{II}_R\tilde{\boldsymbol{\Psi}}_{mm} = \boldsymbol{B}^{II}_R\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}_{mm} \tag{36}$$

in which the state matrices $\boldsymbol{A}^{I}_R$, $\boldsymbol{B}^{I}_R$, $\boldsymbol{A}^{II}_R$, and $\boldsymbol{B}^{II}_R$ of reduced model may be computed by analogy with Eq. (33) and symbolically given by $\Re(\boldsymbol{A}^{I},\boldsymbol{R})$, $\Re(\boldsymbol{B}^{I},\boldsymbol{R})$, $\Re(\boldsymbol{A}^{II},\boldsymbol{R})$, and $\Re(\boldsymbol{B}^{II},\boldsymbol{R})$, respectively. The orthogonality equations for these two reduced models are given by

$$\tilde{\boldsymbol{\Psi}}^T_{mm}\boldsymbol{A}^{I}_R\tilde{\boldsymbol{\Psi}}_{mm} = \tilde{\boldsymbol{\Omega}}_{mm}, \tag{37a}$$

$$\tilde{\boldsymbol{\Psi}}^T_{mm}\boldsymbol{B}^{I}_R\tilde{\boldsymbol{\Psi}}_{mm} = \boldsymbol{I} \tag{37b}$$

$$\tilde{\boldsymbol{\Psi}}^T_{mm}\boldsymbol{A}^{II}_R\tilde{\boldsymbol{\Psi}}_{mm} = \tilde{\boldsymbol{\Omega}}_{mm}, \tag{38a}$$

$$\tilde{\boldsymbol{\Psi}}^T_{mm}\boldsymbol{B}^{II}_R\tilde{\boldsymbol{\Psi}}_{mm} = \boldsymbol{I} \tag{38b}$$

According to the complex modal theory, the state vector $\boldsymbol{Y}(t)$ may be expressed as

$$\boldsymbol{Y}(t) = \tilde{\boldsymbol{\Psi}}\boldsymbol{q}(t) \tag{39}$$

in which $\boldsymbol{q}(t) \in C^{2n}$ are modal coordinates. Generally, it is unnecessary to include all modes in the modal superposition. Thus, if the m eigenpairs are considered, the modal truncated form of Eq. (39) is rewritten as

$$\boldsymbol{Y}(t) = \tilde{\boldsymbol{\Psi}}_m\boldsymbol{q}_m(t) \tag{40}$$

where $\boldsymbol{q}_m(t) \in C^{2m}$. Using the same division of the total dofs, the partitioned form of Eq. (40) is given by

$$\begin{Bmatrix} \boldsymbol{Y}_m(t) \\ \boldsymbol{Y}_s(t) \end{Bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{\Psi}}_{mm} \\ \tilde{\boldsymbol{\Psi}}_{sm} \end{bmatrix}\boldsymbol{q}_m(t) \tag{41}$$

Expanding Eq. (41) leads to two equations

$$\boldsymbol{Y}_m(t) = \tilde{\boldsymbol{\Psi}}_{mm}\boldsymbol{q}_m(t), \quad \boldsymbol{Y}_s(t) = \tilde{\boldsymbol{\Psi}}_{sm}\boldsymbol{q}_m(t) \tag{42}$$

Introducing Eq. (31) into the right-hand side of Eq. (40) and using Eq. (42) give

$$\boldsymbol{Y}(t) = \boldsymbol{T}\tilde{\boldsymbol{\Psi}}_{mm}\boldsymbol{q}_m(t) = \boldsymbol{T}\boldsymbol{Y}_m(t) \tag{43}$$

which is equivalent to

$$\boldsymbol{Y}_s(t) = \boldsymbol{R}\boldsymbol{Y}_m(t) \tag{44}$$

Equation (44) shows that the dynamic condensation matrix also represents the relation of the responses between the masters and slaves. This is another physical meaning of the dynamic condensation matrix. However, this relation is approximate due to the truncation of modes. Because the dynamic condensation matrix and the coordinate transformation matrix are time independent, the differentiation of Eq. (43) with respect to time once gives

$$\dot{\boldsymbol{Y}}(t) = \boldsymbol{T}\dot{\boldsymbol{Y}}_m(t) \tag{45}$$

Inserting Eqs. (43) and (45) into Eq. (3) and premultiplying both sides of the resulted equation by the transpose of $\boldsymbol{T}$ result in

$$\boldsymbol{A}_R^I\boldsymbol{Y}_m(t) - \boldsymbol{B}_R^I\dot{\boldsymbol{Y}}_m(t) = \boldsymbol{F}_R^I(t) \tag{46}$$

Equation (46) is the dynamic equations of equilibrium of the reduced model defined in the state space. The equivalent force vector of the reduced model is defined as

$$\boldsymbol{F}_R^I(t) = \boldsymbol{T}^T\boldsymbol{F}^I(t) = \boldsymbol{F}_m^I(t) + \boldsymbol{R}^T\boldsymbol{F}_s^I(t) \tag{47}$$

If all dofs at which the external loads act are selected as the masters, the equivalent load vector may be simplified as

$$\boldsymbol{F}_R^I(t) = \boldsymbol{F}_m^I(t) \tag{48}$$

The dynamic equations of equilibrium of the other two cases can be similarly derived. Generally, the responses computed from Eq. (46) are very close to those from the full model if the dynamic condensation matrix is selected properly.

It can be seen from the definitions of reduced state matrices that the reduced model resulted from the dynamic condensation is defined in the subspace of the original state space. Thus, each coordinate has its physical meaning. It is very convenient to use them in further dynamic analysis.

3.3 Other Versions of Governing Equations. Postmultiplying both sides of Eq. (34*a*) by the inverse of matrix $\tilde{\boldsymbol{\Psi}}_{mm}$ results in

$$\tilde{\boldsymbol{\Psi}}_{mm}^T\boldsymbol{A}_R^{III} = \tilde{\boldsymbol{\Omega}}_{mm}^2\tilde{\boldsymbol{\Psi}}_{mm}^{-1} \tag{49}$$

Premultiplying both sides of Eq. (32) by the inverse of matrix $\boldsymbol{B}_R^{III}$ and then postmultiplying by the inverse of matrix $\tilde{\boldsymbol{\Psi}}_{mm}$, one obtains

$$(\boldsymbol{B}_R^{III})^{-1}\boldsymbol{A}_R^{III} = \tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}_{mm}^2\tilde{\boldsymbol{\Psi}}_{mm}^{-1} \tag{50}$$

Using Eqs. (49) and (50), the governing Eq. (30) can be rewritten as

$$\boldsymbol{R} = (\boldsymbol{A}_{ss}^{III})^{-1}[(\boldsymbol{B}_{sm}^{III} + \boldsymbol{B}_{ss}^{III}\boldsymbol{R})\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Psi}}_{mm}^T\boldsymbol{A}_R^{III} - \boldsymbol{A}_{sm}^{III}] \tag{51}$$

$$\boldsymbol{R} = (\boldsymbol{A}_{ss}^{III})^{-1}\lfloor(\boldsymbol{B}_{sm}^{III} + \boldsymbol{B}_{ss}^{III}\boldsymbol{R})(\boldsymbol{B}_R^{III})^{-1}\boldsymbol{A}_R^{III} - \boldsymbol{A}_{sm}^{III}\rfloor \tag{52}$$

Equations (51) and (52) are also the governing equations of dynamic condensation matrix. Only the third form of the state matrices, $\boldsymbol{A}^{III}$ and $\boldsymbol{B}^{III}$, are included in these two equations.

Postmultiplying both sides of Eq. (37*a*) by the inverse of matrix $\tilde{\boldsymbol{\Psi}}_{mm}$ results in

$$\tilde{\boldsymbol{\Psi}}_{mm}^T\boldsymbol{A}_R^I = \tilde{\boldsymbol{\Omega}}_{mm}\tilde{\boldsymbol{\Psi}}_{mm}^{-1} \tag{53}$$

Premultiplying both sides of Eq. (35) by $\boldsymbol{A}_R^I(\boldsymbol{B}_R^I)^{-1}$ and with consideration of this equation, one has

$$\boldsymbol{A}_R^I(\boldsymbol{B}_R^I)^{-1}\boldsymbol{A}_R^I\tilde{\boldsymbol{\Psi}}_{mm} = \boldsymbol{A}_R^I\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}_{mm} = \boldsymbol{B}_R^I\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}_{mm}^2 \tag{54}$$

Thus, we have the relation

$$(\boldsymbol{B}_R^I)^{-1}\boldsymbol{A}_R^I(\boldsymbol{B}_R^I)^{-1}\boldsymbol{A}_R^I = \tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}_{mm}^2\tilde{\boldsymbol{\Psi}}_{mm}^{-1} \tag{55}$$

Introducing Eqs. (53) and (55) into Eq. (30) produces

$$\boldsymbol{R} = (\boldsymbol{A}_{ss}^{III})^{-1}[(\boldsymbol{B}_{sm}^{III} + \boldsymbol{B}_{ss}^{III}\boldsymbol{R})\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}_{mm}\tilde{\boldsymbol{\Psi}}_{mm}^T\boldsymbol{A}_R^I - \boldsymbol{A}_{sm}^{III}] \tag{56}$$

$$\boldsymbol{R} = (\boldsymbol{A}_{ss}^{III})^{-1}[(\boldsymbol{B}_{sm}^{III} + \boldsymbol{B}_{ss}^{III}\boldsymbol{R})((\boldsymbol{B}_R^I)^{-1}\boldsymbol{A}_R^I)^2 - \boldsymbol{A}_{sm}^{III}] \tag{57}$$

Equations (56) and (57) are two governing equations based on the first and third forms of the state matrices. Two forms of state matrices are used in each of these two governing equations. Thus, they are referred to as hybrid governing equations of dynamic condensation matrix. Similarly, another two hybrid governing equations using the second and third forms of the state matrices are given by

$$\boldsymbol{R} = (\boldsymbol{A}_{ss}^{III})^{-1}[(\boldsymbol{B}_{sm}^{III} + \boldsymbol{B}_{ss}^{III}\boldsymbol{R})\tilde{\boldsymbol{\Psi}}_{mm}\tilde{\boldsymbol{\Omega}}_{mm}\tilde{\boldsymbol{\Psi}}_{mm}^T\boldsymbol{A}_R^{II} - \boldsymbol{A}_{sm}^{III}] \tag{58}$$

$$\boldsymbol{R} = (\boldsymbol{A}_{ss}^{III})^{-1}[(\boldsymbol{B}_{sm}^{III} + \boldsymbol{B}_{ss}^{III}\boldsymbol{R})((\boldsymbol{B}_R^{II})^{-1}\boldsymbol{A}_R^{II})^2 - \boldsymbol{A}_{sm}^{III}] \tag{59}$$

4 Iterative Schemes for Dynamic Condensation Matrix

It can be seen from the preceding section that the reduced model fully depends on the dynamic condensation matrix. A series of governing equations have been derived for the dynamic condensation matrix. Because these equations are implicit, it is very difficult to solve them directly. The iterative forms of Eqs. (30), (51), and (52) are given by

$$\boldsymbol{R}^{(i)} = (\boldsymbol{A}_{ss}^{III})^{-1}\lfloor(\boldsymbol{B}_{sm}^{III} + \boldsymbol{B}_{ss}^{III}\boldsymbol{R}^{(i-1)})\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)}(\tilde{\boldsymbol{\Omega}}_{mm}^{(i-1)})^2(\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)})^{-1} - \boldsymbol{A}_{sm}^{III}\rfloor \tag{60}$$

$$\boldsymbol{R}^{(i)} = (\boldsymbol{A}_{ss}^{III})^{-1}\lfloor(\boldsymbol{B}_{sm}^{III} + \boldsymbol{B}_{ss}^{III}\boldsymbol{R}^{(i-1)})\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)}(\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)})^T(\boldsymbol{A}_R^{III})^{(i-1)} - \boldsymbol{A}_{sm}^{III}\rfloor \tag{61}$$

$$\boldsymbol{R}^{(i)} = (\boldsymbol{A}_{ss}^{III})^{-1}[(\boldsymbol{B}_{sm}^{III} + \boldsymbol{B}_{ss}^{III}\boldsymbol{R}^{(i-1)})((\boldsymbol{B}_R^{III})^{(i-1)})^{-1}(\boldsymbol{A}_R^{III})^{(i-1)} - \boldsymbol{A}_{sm}^{III}] \tag{62}$$

where $i = 1, 2, 3, \ldots$. The initial approximation of the dynamic condensation matrix is given by

$$\boldsymbol{R}^{(0)} = -(\boldsymbol{A}_{ss}^{III})^{-1}\boldsymbol{A}_{ss}^{III} \tag{63}$$

For concise purpose, the iterative forms of hybrid governing equations are not provided. It can be proven that the dynamic condensation matrix computed from Eqs. (60) or (61) has real values although the matrices with complex values are used on the right-hand sides of the two equations.

Obviously, when an iterative scheme is utilized to solve Eq. (60), the reduced state matrices as well as the eigenvalue matrix, eigenvector matrix and its inverse of the reduced model should be calculated at every iteration. Because the eigenproblem analysis of the viscously damped model is usually time consuming the solution time used in the iteration will rise rapidly as the dimension of the reduced model increases. The governing Eq. (61) is a little better than Eq. (60) because there is not any matrix inverting process. To reduce the computational work, an alternative governing equation is given in Eq. (62). Because the eigenvectors and eigenvalues of reduced model are not included in this governing equation, it is unnecessary to compute the reduced eigenproblem during every iteration. Consequently, if the solution of the reduced eigenproblem is required at each iteration, governing Eq. (61) is a better selection, or Eq. (62) is better.

4.1 Iterative Scheme I. The major iterative steps for governing Eq. (61) are listed below.

(1) Rearrange the matrices $\boldsymbol{A}^{III}$ and $\boldsymbol{B}^{III}$ according to the division of the masters and slaves; or formulate the submatrices $\boldsymbol{A}_{mm}^{III}$, $\boldsymbol{A}_{ms}^{III}$, $\boldsymbol{A}_{ss}^{III}$, $\boldsymbol{B}_{mm}^{III}$, $\boldsymbol{B}_{ms}^{III}$, and $\boldsymbol{B}_{ss}^{III}$ according to Eq. (26);
(2) Factorize the matrix $\boldsymbol{A}_{ss}^{III}$: $\boldsymbol{A}_{ss}^{III} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T$, where $\boldsymbol{L}$ is a low triangular matrix with all diagonal elements equal to zero; $\boldsymbol{D}$ is a diagonal matrix;

(3) Calculate the initial approximation of dynamic condensation matrix $\boldsymbol{R}^{(0)}$ in Eq. (63);
(4) Compute the initial approximation of the reduced state matrices $(\boldsymbol{A}_R^{III})^{(0)}$ and $(\boldsymbol{B}_R^{III})^{(0)}$ and the eigenpairs $\tilde{\boldsymbol{\Psi}}_{mm}^{(0)}$, $\tilde{\boldsymbol{\Omega}}_{mm}^{(0)}$; and
(5) For $i=1,2,3,\ldots$, begin the iteration as follows: (a) Compute the ith approximation of the dynamic condensation matrix in Eq. (61); (b) Calculate the reduced state matrices $(\boldsymbol{A}_R^{III})^{(i)}$, $(\boldsymbol{B}_R^{III})^{(i)}$ and the corresponding eigenpairs $\tilde{\boldsymbol{\Psi}}_{mm}^{(i)}$ and $\tilde{\boldsymbol{\Omega}}_{mm}^{(i)}$; (c) Check the convergence of the iteration using the criterion $|(\alpha_j^{(i)}-\alpha_j^{(i-1)})/\alpha_j^{(i)}|\leqslant\varepsilon$, $(j=1,2,\ldots,p\leqslant m)$ in which α denotes the damping ratio and damped frequency of the complex frequencies (eigenvalues), respectively. The complex frequencies are the diagonal elements of the matrices $\tilde{\boldsymbol{\Omega}}_{mm}^{(i-1)}$ and $\tilde{\boldsymbol{\Omega}}_{mm}^{(i)}$, respectively. ε is a given error tolerance. If the given p complex frequencies are convergent, exit the loop.
(6) Output the results $\boldsymbol{R}=\boldsymbol{R}^{(i)}$, $\boldsymbol{A}_R^{III}=(\boldsymbol{A}_R^{III})^{(i)}$, $\boldsymbol{B}_R^{III}=(\boldsymbol{B}_R^{III})^{(i)}$ and compute the reduced state matrices $(\boldsymbol{A}_R^{I},\boldsymbol{B}_R^{I})$ or $(\boldsymbol{A}_R^{II},\boldsymbol{B}_R^{II})$. Compute the eigenvectors at the slaves using Eq. (31) if necessary.

4.2 Iterative Scheme II. The major iterative steps for the governing Eq. (62) are listed below. Steps 1 through 4 are similar to those in Scheme I;

(5) For $i=0,k,2k,\ldots(k\geqslant 1)$, begin the iteration as follows: (a) Iterate Eq. (62) for k times and obtain the matrix $\boldsymbol{R}^{(i+k)}$; (b) Calculate the reduced state matrices $(\boldsymbol{A}_R^{III})^{(i+k)}$, $(\boldsymbol{B}_R^{III})^{(i+k)}$ and the corresponding eigenpairs $\tilde{\boldsymbol{\Psi}}_{mm}^{(i+k)}$ and $\tilde{\boldsymbol{\Omega}}_{mm}^{(i+k)}$; and (c) Check the convergence of the iteration using the criterion: $|(\alpha_j^{(i+k)}-\alpha_j^{(i)})/\alpha_j^{(i+k)}|\leqslant\varepsilon$, $(j=1,2,\ldots,p\leqslant m)$. If the given p complex frequencies are convergent, exit the loop.

(6) Output the results $\boldsymbol{R}=\boldsymbol{R}^{(i+k)}$, $\boldsymbol{A}_R^{III}=(\boldsymbol{A}_R^{III})^{(i+k)}$, $\boldsymbol{B}_R^{III}=(\boldsymbol{B}_R^{III})^{(i+k)}$ and compute the reduced state matrices $(\boldsymbol{A}_R^{I},\boldsymbol{B}_R^{I})$ or $(\boldsymbol{A}_R^{II},\boldsymbol{B}_R^{II})$. Compute the eigenvectors at the slaves using Eq. (31) if necessary.

In this iterative scheme, when the integer k is larger than 1, $k-1$ times calculation of the reduced eigenvalue problems is avoided. This makes the scheme more computationally efficient.

In the earlier two iterative schemes, the damping ratios and damped frequencies are used to check the convergence of the reduced model. If the errors of the damping ratios and damped frequencies resulted from the reduced models in the two successive iterations are less than the error tolerance, we say the reduced model is close to the full model at that frequency range. This conclusion is generally guaranteed. Another advantage of the convergent criterion is that the by-products, damping ratios and damped frequencies, are usually of interest in dynamic analysis. Of course, another parameter such as dynamic condensation matrix itself may be used as a measure for the convergence.

5 Comparison With Other Iterative Approaches

5.1 Other Iterative Methods. The efficient iterative approaches for the dynamic condensation of viscously damped models were developed by Qu [13,19], Rivera [20], and Qu et al. [21]. The iterative form of governing equation derived by Qu [13,19], referred to method I, from the dynamic equations of equilibrium is given by

$$\begin{cases}\boldsymbol{R}_1^{(i)}=(\boldsymbol{A}_{ss}^{I})^{-1}\lfloor(\boldsymbol{B}_{sm}^{I}+\boldsymbol{B}_{ss}^{I}\boldsymbol{R}_1^{(i-1)})\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)}(\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)})^{T}(\boldsymbol{A}_R^{I})^{(i-1)}-\boldsymbol{A}_{sm}^{I}\rfloor\\ \boldsymbol{R}_1^{(0)}=-(\boldsymbol{A}_{ss}^{I})^{-1}\boldsymbol{A}_{sm}^{I}\end{cases} \tag{64}$$

or

$$\begin{cases}\boldsymbol{R}_1^{(i)}=(\boldsymbol{A}_{ss}^{I})^{-1}[(\boldsymbol{B}_{sm}^{I}+\boldsymbol{B}_{ss}^{I}\boldsymbol{R}_1^{(i-1)})((\boldsymbol{B}_R^{I})^{(i-1)})^{-1}(\boldsymbol{A}_R^{I})^{(i-1)}-\boldsymbol{A}_{sm}^{I}]\\ \boldsymbol{R}_1^{(0)}=-(\boldsymbol{A}_{ss}^{I})^{-1}\boldsymbol{A}_{sm}^{I}\end{cases} \tag{65}$$

The state matrices in Eqs. (64) and (65) may be expressed as $\Re(\boldsymbol{A}^{I},\boldsymbol{R}_1^{(i-1)})$ and $\Re(\boldsymbol{B}^{I},\boldsymbol{R}_1^{(i-1)})$. Rivera's, referred to as method II, iterative equations were developed from the eigenvalue equations of the full model. Using the denotations in this paper, the iterative form of the governing equations can be expressed as

$$\begin{cases}\boldsymbol{R}_2^{(i)}=(\boldsymbol{A}_{ss}^{I})^{-1}\lfloor(\boldsymbol{B}_{sm}^{I}+\boldsymbol{B}_{ss}^{I}\boldsymbol{R}_2^{(i-1)})\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)}(\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)})^{T}(\boldsymbol{A}_R^{I})^{(i-1)}-\boldsymbol{A}_{sm}^{I}\rfloor\\ \boldsymbol{R}_2^{(0)}=\begin{bmatrix}\boldsymbol{R}_G & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}_G\end{bmatrix}\end{cases} \tag{66}$$

in which $\boldsymbol{R}_G=-\boldsymbol{K}_{ss}^{-1}\boldsymbol{K}_{sm}$ is the Guyan condensation matrix of undamped models.

An alternative iterative approach was developed by Qu et al. [21], referred to as method III, to improve the convergence of the iterative schemes in Eqs. (64) or (65). The iterative form is given by

$$\begin{cases}\boldsymbol{R}_3^{(i)}=(\boldsymbol{A}_{ss}^{II})^{-1}\lfloor(\boldsymbol{B}_{sm}^{II}+\boldsymbol{B}_{ss}^{II}\boldsymbol{R}_3^{(i-1)})\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)}(\tilde{\boldsymbol{\Psi}}_{mm}^{(i-1)})^{T}(\boldsymbol{A}_R^{II})^{(i-1)}-\boldsymbol{A}_{sm}^{II}\rfloor\\ \boldsymbol{R}_3^{(0)}=-(\boldsymbol{A}_{ss}^{II})^{(-1)}\boldsymbol{A}_{sm}^{II}\end{cases} \tag{67}$$

or

$$\begin{cases}\boldsymbol{R}_3^{(i)}=(\boldsymbol{A}_{ss}^{II})^{-1}[(\boldsymbol{B}_{sm}^{II}+\boldsymbol{B}_{ss}^{II}\boldsymbol{R}_3^{(i-1)})((\boldsymbol{B}_R^{II})^{(i-1)})^{-1}(\boldsymbol{A}_R^{II})^{(i-1)}-\boldsymbol{A}_{sm}^{II}]\\ \boldsymbol{R}_3^{(0)}=-(\boldsymbol{A}_{ss}^{II})^{-1}\boldsymbol{A}_{sm}^{II}\end{cases} \tag{68}$$

The state matrices in Eqs. (67) and (68) are given by $\Re(\boldsymbol{A}^{II},\boldsymbol{R}_3^{(i-1)})$ and $\Re(\boldsymbol{B}^{II},\boldsymbol{R}_3^{(i-1)})$.

5.2 Comments on Convergence. Generally, starting vector and convergent rate are two important features of an iterative dynamic condensation approach. In the following the present iterative method will be compared with the approaches in Refs. [13,19–21] from these two viewpoints.

The governing equation of dynamic condensation matrix in methods I and II are actually the same as shown in Eqs. (64) and (66). Using Eq. (4), the initial approximation in Eqs. (64) and (65) is given by

$$\boldsymbol{R}_1^{(0)}=-(\boldsymbol{A}_{ss}^{I})^{-1}\boldsymbol{A}_{sm}^{I}=\begin{bmatrix}-\boldsymbol{K}_{ss}^{-1}\boldsymbol{K}_{sm} & \boldsymbol{0}\\ \boldsymbol{0} & -\boldsymbol{M}_{ss}^{-1}\boldsymbol{M}_{sm}\end{bmatrix}\equiv\begin{bmatrix}\boldsymbol{R}_G & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}_M\end{bmatrix} \tag{69}$$

With the introduction of Eq. (69) into Eq. (29) and consideration of the partitioned form of eigenvector matrix, shown in Eq. (26*c*), we have

$$\begin{bmatrix}\boldsymbol{\Psi}_{sm} & \boldsymbol{\Psi}_{sm}^{*}\\ \boldsymbol{\Psi}_{sm}\boldsymbol{\Omega}_{mm} & \boldsymbol{\Psi}_{sm}^{*}\boldsymbol{\Omega}_{mm}^{*}\end{bmatrix}=\begin{bmatrix}\boldsymbol{R}_G & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}_M\end{bmatrix}\begin{bmatrix}\boldsymbol{\Psi}_{mm} & \boldsymbol{\Psi}_{mm}^{*}\\ \boldsymbol{\Psi}_{mm}\boldsymbol{\Omega}_{mm} & \boldsymbol{\Psi}_{mm}^{*}\boldsymbol{\Omega}_{mm}^{*}\end{bmatrix} \tag{70}$$

Expanding Eq. (70) gives

$$\boldsymbol{\Psi}_{sm}=\boldsymbol{R}_G\boldsymbol{\Psi}_{mm}, \tag{71a}$$

$$\boldsymbol{\Psi}_{sm}\boldsymbol{\Omega}_{mm}=\boldsymbol{R}_M\boldsymbol{\Psi}_{mm}\boldsymbol{\Omega}_{mm} \tag{71b}$$

$$\boldsymbol{\Psi}_{sm}^{*}=\boldsymbol{R}_G\boldsymbol{\Psi}_{mm}^{*}, \tag{72a}$$

$$\boldsymbol{\Psi}_{sm}^{*}\boldsymbol{\Omega}_{mm}^{*}=\boldsymbol{R}_M\boldsymbol{\Psi}_{mm}^{*}\boldsymbol{\Omega}_{mm}^{*} \tag{72b}$$

Clearly, the two equations in Eqs. (71) and (72) are, respectively,

Table 1 Lowest five damping ratios and damped frequencies (rad/s)

Mode	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
Ratio	0.02740	0.13041	0.02368	0.03875	0.05394
Freq.	1.07187	1.07709	3.23865	5.38783	7.52560

contradictive. Thus, this initial approximation is unreasonable.

The initial approximation used by Rivera [20] gives the following relations:

$$\boldsymbol{\Psi}_{sm} = \boldsymbol{R}_G \boldsymbol{\Psi}_{mm}, \tag{73a}$$

$$\boldsymbol{\Psi}_{sm}\boldsymbol{\Omega}_{mm} = \boldsymbol{R}_G \boldsymbol{\Psi}_{mm}\boldsymbol{\Omega}_{mm} \tag{73b}$$

$$\boldsymbol{\Psi}^*_{sm} = \boldsymbol{R}_G \boldsymbol{\Psi}^*_{mm}, \tag{74a}$$

$$\boldsymbol{\Psi}^*_{sm}\boldsymbol{\Omega}^*_{mm} = \boldsymbol{R}_G \boldsymbol{\Psi}^*_{mm}\boldsymbol{\Omega}^*_{mm} \tag{74b}$$

They are consistent to each other. Thus, this approximation has higher accuracy than that in Eqs. (64) and (65). This will be shown in the numerical examples.

Using Eq. (7), the initial approximation in Eqs. (67) and (68) are given by

$$\boldsymbol{R}_3^{(0)} = -\begin{bmatrix} -\boldsymbol{K}_{ss}^{-1}\boldsymbol{C}_{ss}\boldsymbol{K}_{ss} & \boldsymbol{K}_{ss}^{-1} \\ \boldsymbol{K}_{ss}^{-1} & \boldsymbol{0} \end{bmatrix}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{K}_{sm} \\ \boldsymbol{K}_{sm} & \boldsymbol{C}_{sm} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{K}_{ss}^{-1}\boldsymbol{K}_{sm} & \boldsymbol{K}_{ss}^{-1}\boldsymbol{C}_{ss}\boldsymbol{K}_{ss}^{-1}\boldsymbol{K}_{sm} - \boldsymbol{K}_{ss}^{-1}\boldsymbol{C}_{sm} \\ \boldsymbol{0} & -\boldsymbol{K}_{ss}^{-1}\boldsymbol{K}_{sm} \end{bmatrix} \tag{75}$$

As one knows, the norm of the damping matrix is usually much smaller than that of the stiffness matrix. Thus, one has

$$\|\boldsymbol{K}_{ss}^{-1}\boldsymbol{C}_{ss}\boldsymbol{K}_{ss}^{-1}\boldsymbol{K}_{sm} - \boldsymbol{K}_{ss}^{-1}\boldsymbol{C}_{sm}\| \ll \|\boldsymbol{K}_{ss}^{-1}\boldsymbol{K}_{sm}\| \tag{76}$$

Therefore, the inconsistency and contradiction in the initial approximation $\boldsymbol{R}_3^{(0)}$ are much lighter than those in $\boldsymbol{R}_1^{(0)}$. Moreover, if the damping matrix is proportional to the stiffness matrix, $\boldsymbol{R}_3^{(0)}$ is identical to $\boldsymbol{R}_2^{(0)}$. Since the state matrices $\boldsymbol{A}^{II}$ has the same form of the state matrix $\boldsymbol{A}^{III}$ as shown in Eqs. (7) and (10), the initial approximation of dynamic condensation matrix used in Eqs. (67) and (68) are identical to that used in the proposed method.

A full proof of the convergence of the present iterative method and those in Refs. [19–21] is very complex. Fortunately, the convergence of an iterative dynamic condensation approach of undamped models, called iterative improved reduced system, has been proven by Friswell et al. [23]. The convergent rate is approximately in the order of λ_j/λ_k [23] at the first several iterations. j denotes the jth eigenvalue of full model. k runs over the modes that are not included in the reduced model. It can be simply verified that the iterative approaches of damped models aforementioned have very similar convergent features.

The governing equations of dynamic condensation matrix in methods I and II are the same and they are based on the same eigenvalue equation, i.e., Eq. (17). Of course, they have the same convergent rate. Therefore, although the unreasonable initial approximation is used in method I, it has been shown and will be shown later that the reduced models resulted from these two approaches are very close after several iterations.

Method III is based on eigenvalue Eq. (22) which has the same eigenvalues as those in Eq. (17). Therefore, this approach has the same order of convergent rate as that in methods I and II at the first several iterations. Research shows [21] that this approach converges much faster than methods I and II after several iterations. The initial approximation used in this method is shown in Eq. (75) which is very close to the approximation in Eq. (66).

The present iteration scheme is based on eigenvalue Eq. (23). It has quadratic eigenvalues of the real model. Thus, the convergent rate is in the order of (λ_j/λ_k). The initial approximation is the same as that defined in Eq. (67). Therefore, this approach is expected to be convergent faster than the three iterative approaches previously proposed.

It is necessary to note that all four iterative approaches aforementioned have very similar iterative formats, the computational efforts required for each iteration of these methods is very close.

6 Applications

The iterative methods given in Secs. 4 and 5 will be tested on two practical dynamic systems which have been used in Refs. [19,21]. The details of these two examples may be found in any of these two references and will not be iterated herein. In these two examples, if the frequencies and eigenvectors (mode shapes) resulted from the reduced model is close to those of the full model, we will say that the reduced model are close to the full model within that frequency range.

6.1 40-Story Building With Tuned Mass Damper (TMD). The lowest five damping ratios and damped frequencies resulted from the full model are listed in Table 1 for comparison purposes. Six cases, shown in Table 2, for the dynamic condensation of the full model will be considered. In case 1, for example, Eqs. (64) or (65) is used to compute the dynamic condensation matrix and $\Re(\boldsymbol{A}^I, \boldsymbol{R}^{(i)})$ and $\Re(\boldsymbol{B}^I, \boldsymbol{R}^{(i)})$ are used to calculate the reduced state matrices from which the corresponding eigenpairs of the reduced model are obtained.

The dofs pertaining to the 10th, 20th, 30th, 40th floors and TMD are selected as the masters when the dynamic condensation approaches are applied to the tall building. The relative errors of the five damping ratios and damped frequencies of the reduced models are listed in Tables 3–12. Ten iterations are considered for the reduced model. "-" in these tables indicates that the relative error is less than 5×10^{-6}. The convergent features of the eigenvectors are not shown because they are very similar to those of the damped frequencies and this takes lots of space.

The reduced models obtained from all four iterative approaches are convergent to the full model after sufficient iterations. Several iterations are enough to compute the lower order of eigenpairs. For example, the first and second damped frequencies in the first approximation are accurate enough for engineering applications. One or two iterations are also enough for the accuracy of these two damping ratios.

Due to the contradiction of the initial approximation used in case 1, the accuracy is generally very low. Most of the relative errors are greater than one as shown clearly in Tables 3–12. They are reduced significantly after the first or the second iteration. The results in these tables show that this iterative scheme converges

Table 2 Six cases considered in the first example

Cases	Condensation matrix	Reduced state matrices
1	Eqs. (64)/(65)	$\Re(\boldsymbol{A}^I, \boldsymbol{R}^{(i)})$, $\Re(\boldsymbol{B}^I, \boldsymbol{R}^{(i)})$
2	Eqs. (66)	$\Re(\boldsymbol{A}^I, \boldsymbol{R}^{(i)})$, $\Re(\boldsymbol{B}^I, \boldsymbol{R}^{(i)})$
3	Eqs. (67)/(68)	$\Re(\boldsymbol{A}^{II}, \boldsymbol{R}^{(i)})$, $\Re(\boldsymbol{B}^{II}, \boldsymbol{R}^{(i)})$
4	Eqs. (61)/(62)	$\Re(\boldsymbol{A}^{III}, \boldsymbol{R}^{(i)})$, $\Re(\boldsymbol{B}^{III}, \boldsymbol{R}^{(i)})$
5	Eqs. (61)/(62)	$\Re(\boldsymbol{A}^I, \boldsymbol{R}^{(i)})$, $\Re(\boldsymbol{B}^I, \boldsymbol{R}^{(i)})$
6	Eqs. (61)/(62)	$\Re(\boldsymbol{A}^{II}, \boldsymbol{R}^{(i)})$, $\Re(\boldsymbol{B}^{II}, \boldsymbol{R}^{(i)})$

Table 3 Relative errors of the first damped frequency

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	0.00942	0.00412	0.00412	0.00696	0.00412	0.00412
1	0.00412	0.00002	0.00004	-	-	-
2	0.00002	0.00001	-	-	-	-
4	0.00001	-	-	-	-	-

although the accuracy of the initial approximation is very low.

The accuracy of the initial approximation based on Guyan condensation matrix is much higher than that in case 1, especially for the second through fifth damped frequencies and the third through fifth damping ratios. However, the difference reduces with the increase of number of iterations. After several iterations, the accuracy of the reduced models resulted from these two iterative methods, methods I and II, is very close.

The relative errors of the initially approximate eigenpairs in case 3 are the same as those in case 2 because the damping matrix in this example is almost proportional to the stiffness matrix. Generally, the iterative approach in case 3 converges faster than the approach in case 2. However, this is not true for the fifth damping ratio of the reduced model as shown in Table 12.

The proposed iterative method converges much faster than the three iterative schemes in cases 1–3. The relative errors of the eigenpairs resulted from the former are generally less than one tenth of those resulted from the latter three methods. As anticipated, the reduced models in cases 4, 5, and 6 have very close accuracy. Due to this, the latter two cases will not be considered in the second example.

6.2 Floating Raft Isolation System. The isolation system has a total of 179 dofs. The lowest 15 frequencies are listed in Table 13. They are considered as exact for comparison purposes. The modal damping matrix, which is transformed by the lowest five undamped modes, is listed in Table 14. It can be seen that the maximum off-diagonal elements are very close to or greater than the corresponding diagonal elements in the lowest three modes. Therefore, the nonproportionality of damping is heavy.

The lowest ten eigenpairs are of interest. As we know the ratio of the number of masters to the number of modes interested is generally between 1 and 2 [13]. The ratio is selected as 1.5, that is, the number of masters is 15. The dofs associated with the two machines and the translational dofs at nodes 1, 7, 15, 17, 19, 21, 29, 35 in raft and at nodes 7, 8, 9, 10, 16 in base are selected as

Table 4 Relative errors of the second damped frequency

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	1.85339	0.00207	0.00207	0.00075	0.00207	0.00207
1	0.00207	-	0.00002	-	-	-
2	-	-	-	-	-	-

Table 5 Relative errors of the third damped frequency

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	1.71425	0.05433	0.05433	0.05433	0.05433	0.05433
1	0.05433	0.00176	0.00205	0.00009	0.00008	0.00010
2	0.00176	0.00067	0.00040	-	-	-
4	0.00047	0.00035	-	-	-	-
6	0.00026	0.00020	-	-	-	-
8	0.00015	0.00012	-	-	-	-
10	0.00009	0.00007	-	-	-	-

Table 6 Relative errors of the fourth damped frequency

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	1.49221	0.13902	0.13902	0.13902	0.13902	0.13902
1	0.13902	0.01816	0.02009	0.00362	0.00331	0.00394
2	0.01816	0.00680	0.00839	0.00039	0.00032	0.00046
4	0.00378	0.00249	0.00068	0.00001	0.00001	0.00001
6	0.00176	0.00128	0.00006	-	-	-
8	0.00095	0.00072	0.00001	-	-	-
10	0.00055	0.00042	-	-	-	-

Table 7 Relative errors of the fifth damped frequency

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	1.18343	0.16483	0.16483	0.16483	0.16483	0.16483
1	0.16483	0.05496	0.05854	0.03288	0.03189	0.03385
2	0.05496	0.03587	0.03986	0.01426	0.01322	0.01528
4	0.02412	0.01644	0.01729	0.00314	0.00269	0.00360
6	0.01155	0.00829	0.00784	0.00082	0.00067	0.00097
8	0.00609	0.00456	0.00367	0.00023	0.00018	0.00028
10	0.00346	0.00266	0.00178	0.00007	0.00005	0.00008

Table 8 Relative errors of the first damping ratio

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	4.35952	3.72752	3.72752	0.03491	3.72752	3.72752
1	3.72752	0.00013	0.00371	0.00001	-	0.00002
2	0.00013	0.00006	0.00002	-	-	-
4	0.00004	0.00003	-	-	-	-
6	0.00003	0.00002	-	-	-	-
8	0.00002	0.00001	-	-	-	-
10	0.00001	0.00001	-	-	-	-

the masters. The former four cases in Table 2 are considered in this example. They are denoted by A, B, C, and D, respectively. Since higher modes generally converge more slowly than the lower modes, only the higher five damping ratios and damped frequencies are considered to check the convergence. The relative errors of the damping ratios and damped frequencies of the reduced models resulted from these four iterative schemes are plotted in Figs. 1 and 2.

The damped frequencies of the reduced models approach to those of the full model steadily with the implementation of itera-

Table 9 Relative errors of the second damping ratio

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	0.78947	0.78254	0.78254	0.00662	0.78254	0.78254
1	0.78254	0.00002	0.00005	-	-	-
2	0.00002	0.00001	-	-	-	-
4	0.00001	0.00001	-	-	-	-
6	-	-	-	-	-	-

Table 10 Relative errors of the third damping ratio

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	1.70258	0.05672	0.05672	0.05672	0.05672	0.05672
1	0.05672	0.00170	0.10333	0.00330	0.00005	0.00665
2	0.00170	0.00012	0.00426	0.00008	-	0.00016
4	0.00010	0.00010	0.00005	-	-	-
6	0.00008	0.00006	-	-	-	-
8	0.00005	0.00004	-	-	-	-
10	0.00003	0.00002	-	-	-	-

Table 11 Relative errors of the fourth damping ratio

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	1.49350	0.14210	0.14210	0.14210	0.14210	0.14210
1	0.14210	0.01735	0.23383	0.02711	0.00259	0.05589
2	0.01735	0.00616	0.02546	0.00471	0.00002	0.00941
4	0.00339	0.00185	0.00362	0.00013	0.00001	0.00024
6	0.00108	0.00069	0.00051	-	-	0.00001
8	0.00048	0.00035	0.00007	-	-	-
10	0.00026	0.00020	0.00001	-	-	-

Table 12 Relative errors of the fifth damping ratio

Iteration	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
0	1.18903	0.16744	0.16744	0.16744	0.16744	0.16744
1	0.16744	0.05400	0.16806	0.01446	0.03050	0.05722
2	0.05400	0.01575	0.01988	0.02146	0.00838	0.05012
4	0.00393	0.00463	0.01907	0.00902	0.00205	0.01588
6	0.00803	0.00896	0.01225	0.00316	0.00171	0.00460
8	0.00817	0.00697	0.00758	0.00111	0.00084	0.00139
10	0.00567	0.00455	0.00472	0.00039	0.00035	0.00044

Table 13 Lower 15 damping ratios and frequencies of the isolation system (rad/s)

Mode	Ratio	Freq.	Mode	Ratio	Freq.	Mode	Ratio	Freq.
1	0.01216	27.4784	6	0.02818	238.643	11	0.00331	698.030
2	0.01478	30.7270	7	0.02230	335.931	12	0.00779	783.472
3	0.00477	67.4671	8	0.03093	418.320	13	0.02380	1051.91
4	0.02850	226.759	9	0.03348	533.934	14	0.02292	1124.94
5	0.04282	227.902	10	0.03021	542.827	15	0.00098	1244.86

Table 14 Modal damping matrix

.668501E+00	−.473408E−01	−.531268E+00	−.178491E+00	−.393805E+00
−.473408E−01	.908089E+00	−.315532E+00	.396007E+00	−.217935E+00
−.531268E+00	−.315532E+00	.643560E+00	−.640672E−02	.140491E+01
−.178491E+00	.396007E+00	−.640672E−02	.128987E+02	.845864E−02
−.393805E+00	−.217935E+00	.140491E+01	.845864E−02	.195278E+02

tion as shown in Fig. 1. The convergence is very fast at the first several iterations. After that, the convergence becomes much slower. Particularly, the relative errors of the damped frequencies computed from cases A and B reduce very slowly after several iterations, five for example. The error curves A and B are almost horizontal after that. Therefore, very small error tolerance for the damped frequency is required in the iterative scheme I in Sec. 4.1. It could be slightly big in iterative scheme II in Sec. 4.2. For both iterative schemes, a large number of iterations is usually necessary if the accuracy of interest is high. These problems do not happen

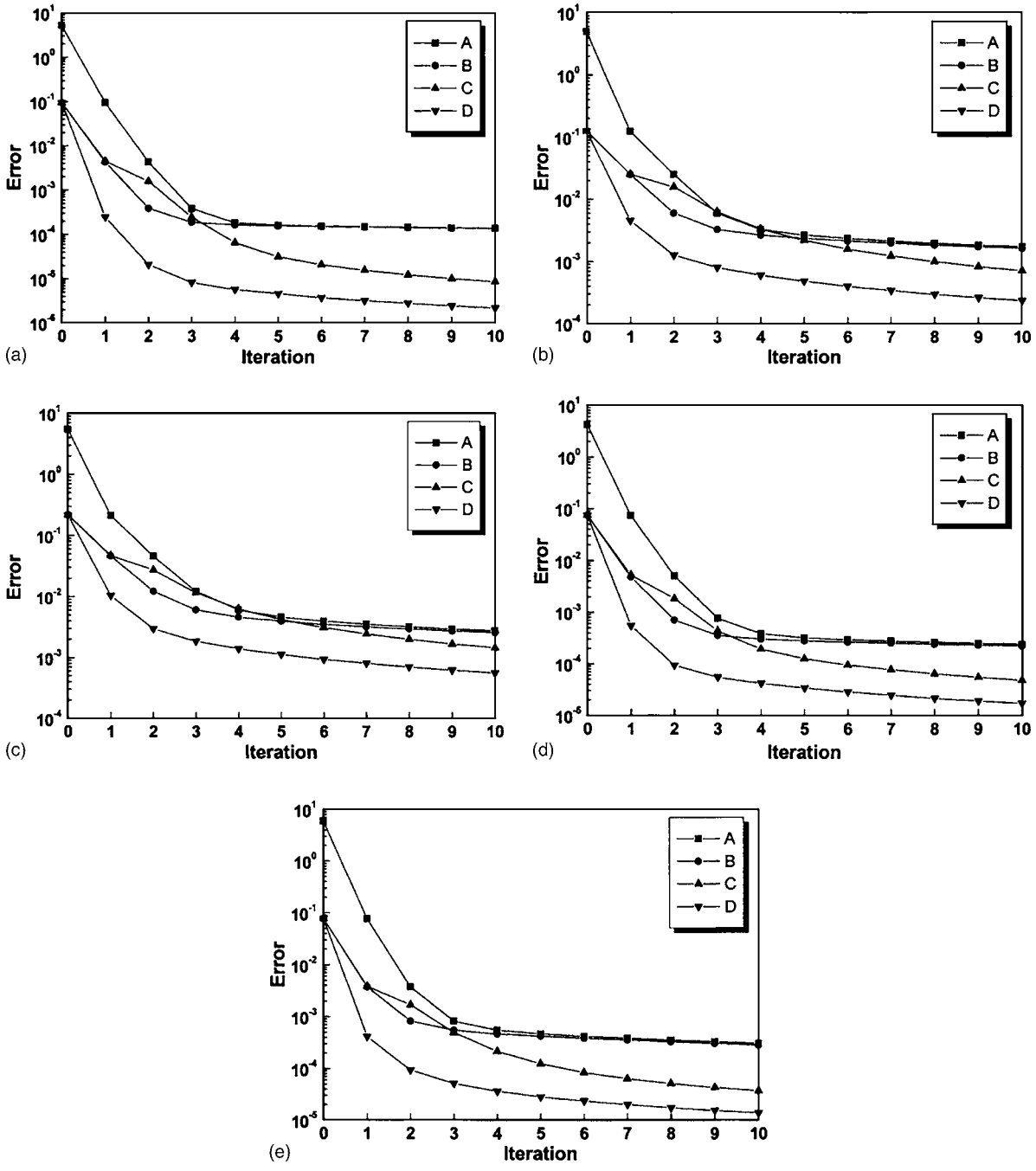


Fig. 1 Comparison of relative errors of damped frequencies: (a) sixth damped frequency; (b) seventh damped frequency; (c) eighth damped frequency; (d) ninth damped frequency; and (e) tenth damped frequency

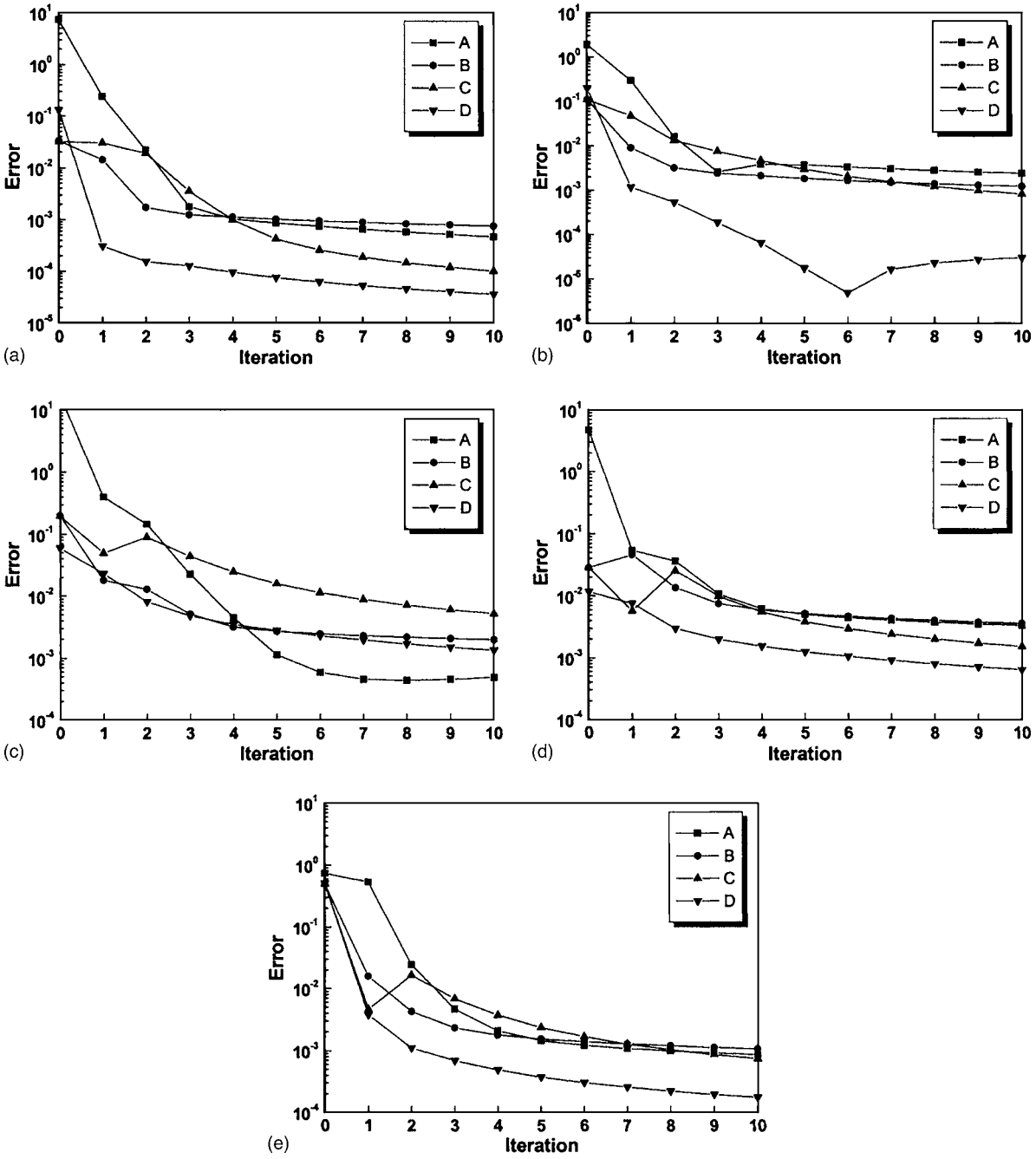


Fig. 2 Comparison of relative errors of damping ratios: (a) sixth damping ratio; (b) seventh damping ratio; (c) eighth damping ratio; (d) ninth damping ratio; and (e) tenth damping ratio

in cases C and D.

Although the relative errors of the damped frequencies of the reduced model resulted from case B are much smaller than those from case A at the first several iterations, they are very close in the following iterations. The reason is that both iterative approaches are based on the same governing equations of the dynamic condensation matrix. The initial approximation of the dynamic condensation matrix may make the two schemes different at the first several iterations. However, the final results should be very close.

The accuracy of the damped frequencies obtained from case C is higher than that of cases A and B after several iterations. This means that the advantages of the iterative method in case C only can be shown after these iterations. Except the initial approximation, the proposed method has much higher accuracy than all these three iterative approaches.

The convergent properties of the damping ratios shown in Fig. 2 are not as clear as those of damped frequencies, although they are generally similar. Therefore, it is better to select the damped frequency as the major parameter to check the convergence and the damping ratio as the minor.

7 Conclusions

An accelerated iterative method for the dynamic condensation of viscously damped models has been developed. Three groups of governing equations for dynamic condensation matrix were derived from the eigenvalue equations defined in the state space. Two iterative schemes for solving these governing equations were

provided. A detailed comparison of the present method with three approaches previously proposed was presented. Two numerical examples were included to show the convergent features of each iterative method.

Based on different supplemental identities, three state space formulations for the dynamic equations of equilibrium and eigenvalue problems of viscously damped models were provided. It has been shown that although these three formulations are very similar, the convergent features of the corresponding iterative methods are different.

Two definitions for the dynamic condensation matrix of viscously damped models were reviewed. They represent the relations of eigenvectors and the state vector between the masters and the slaves, respectively. The relations between these two definitions have been demonstrated using the complex modal superposition theory. If the proper modal truncation scheme is applied, the two definitions become the same.

The reduced models obtained from all four iterative approaches are convergent to the full model after sufficient iterations. Several iterations are enough to compute the lower orders of eigenpairs. The convergence is very fast at the first several iterations. After that, the convergence becomes much slower.

Due to the strong contradiction in the initial approximation of dynamic condensation matrix used in iterative method I, the accuracy is generally very low. Most of the relative errors are greater than 1. Fortunately, the low accuracy of initial approximation does not affect the convergence. The accuracy of the initial approximation based on Guyan condensation matrix, used in the iterative method II, is much higher. However, the difference reduces with the increase of number of iterations. After several iterations, the accuracy of the reduced models resulted from these two iterative methods are very close. Generally, the iterative method III converges faster than the two approaches, methods I and II. The proposed iterative method converges much faster than the earlier three iterative algorithms.

Due to the very slow convergence of methods I and II after several iterations, very small error tolerance for the damped frequency is usually required in the iterative scheme 1. For this case, the iterative scheme 2 is better than iterative scheme 1. The convergent properties of the damping ratios are not as clear as those of damped frequencies, although they are generally similar. Therefore, it is better to select the damped frequency as the major parameter to check the convergence and the damping ratio as the minor.

References

[1] Suarez, L. E., and Singh, M. P., 1992, "Dynamic Condensation Method for Structural Eigenvalue Analysis," AIAA J., **30**, pp. 1046–1054.

[2] Rothe, K., and Voss, H., 1994, "Improving Condensation Methods for Eigenvalue Problems via Rayleigh Functions," Comput. Methods Appl. Mech. Eng., **111**, pp. 169–183.

[3] Noor, A. K., 1994, "Recent Advances and Applications of Reduction Methods," Appl. Mech. Rev., **47**, pp. 125–146.

[4] Friswell, M. I., Garvey, S. D., and Penny, J. E. T., 1995, "Model Reduction using Dynamic and Iterative IRS Technique," J. Sound Vib., **186**, pp. 311–323.

[5] Kim, K.-O., 1995, "Hybrid Dynamic Condensation for Eigenproblems," Comput. Struct., **56**, pp. 105–112.

[6] Hou, Z., and Chen, S., 1997, "Iterated Dynamic Condensation Technique and Its Applications in Modal Testing," Shock Vib., **4**, pp. 143–151.

[7] Mokeyev, V. V., 1998, "A Frequency Condensation Method for the Eigenvalue Problem," Commun. Numer. Methods Eng., **14**, pp. 1–8.

[8] Qu, Z., and Fu, Z., 1998, "An Iterative Method for Dynamic Condensation of Finite Element Models, Part I: Basic Method," Journal of Shanghai Jiao Tong University (English Edition), **3**, pp. 83–88.

[9] Qu, Z.-Q., and Fu, Z.-F, 1998, "New Structural Dynamic Condensation Method for Finite Element Models," AIAA J., **36**, pp. 1320–1324.

[10] Qu, Z.-Q., 1998, "A Multi-Step Method for Matrix Condensation of Finite Element Models," J. Sound Vib., **214**, pp. 965–971.

[11] Friswell, M. I., Garvey, S. D., and Penny, J. E. T., 1998, "The Convergence of the Iterated IRS Methods," J. Sound Vib., **211**, pp. 123–132.

[12] Kim, K.-O., 1998, "Perturbation Method for Condensation of Eigenproblems," AIAA J., **36**, pp. 1537–1543.

[13] Qu, Z.-Q., 1998, "Structural Dynamic Condensation Techniques: Theory and Applications," Ph. D. dissertation, State Key Laboratory of Vibration, Shock and Noise, Shanghai Jiao Tong University.

[14] Qu, Z.-Q., and Fu, Z.-F., 2000, "An Iterative Method for Dynamic Condensation of Structural Matrices," Mech. Syst. Signal Process. **14**, pp. 667–678.

[15] Kim, K.-O., and Kang, M.-K., 2001, "Convergence Acceleration of Iterative Modal Reduction Methods," AIAA J. **39**, pp. 134–140.

[16] Qu, Z.-Q., *Model Order Reduction Techniques: With Applications in Finite Element Analysis*, Springer-Verlag, London, 2004.

[17] Kammer, D. C., 1987, "Test-Analysis Model Development using an Exact Modal Reduction," Int. J. Anal. Exp. Modal Anal. **2**, pp. 174–179.

[18] Kane, K., and Torby, B. J., 1991, "The Extended Modal Reduction Method Applied to Rotor Dynamic Problems," ASME J. Vibr. Acoust. **113**, pp. 79–84.

[19] Qu, Z.-Q., and Chang, W., 2000, "Dynamic Condensation Method for Viscous Damped Vibration Systems in Engineering," Eng. Struct. **22**, pp. 1426–1432.

[20] Rivera, M. A., Singh, M. P., and Suarez, L. E., 1999, "Dynamic Condensation Approach for Nonclassically Damped Structures," AIAA J. **37**, pp. 564–571.

[21] Qu, Z.-Q., Selvam, R. P., and Jung, Y., 2003, "Model Condensation for Nonclassically Damped Systems, Part II: Iterative Schemes for Dynamic Condensation," Mech. Syst. Signal Process. **17**, pp. 1017–1032.

[22] Rao, G. V., 2002, "Dynamic Condensation and Synthesis of Unsymmetric Structural Systems," ASME J. Appl. Mech. **69**, pp. 610–616.

[23] Friswell, M. I., Garvey, S. D., and Penny, J. E. T., 1998, "The Convergence of the Iterated IRS Methods," J. Sound Vib. **211**, pp. 123–132.

Dynamics of Plate Generated by Moving Harmonic Loads

Lu Sun
Assistant Professor
Department of Civil Engineering,
The Catholic University of America,
Washington, DC 20064
e-mail: sunl@cua.edu

A thin plate resting on a Winkler elastic foundation subject to a moving harmonic load can be used as the model for highway and airport pavement under moving vehicle load and many other applications. The study of dynamic response of the plate thus becomes very important. In this paper we study the dynamic displacement of a plate caused by a moving harmonic line and point load. The solution is represented by the convolution of dynamic Green's function of plate. An approximate relationship between critical load velocity and critical frequency is established analytically. It is found that the maximum displacement response occurs at the center of the moving load and travels at the same speed with the load. [DOI: 10.1115/1.1993669]

1 Introduction

Dynamic response of a thin plate subject to a moving load is a very important problem not only to pavement design [1–5], but also to other applications such as optimum vehicle suspension design [6,7]. To prevent the stress applied by a moving vehicle exceeding a certain level on a landmine, it is paramount to know the dynamic displacement and stress field caused by a moving load. Measured pavement dynamic response information can also be used for pavement nondestructive evaluation [8–10]. In Ref. [11] the full-scale force-vibration test resulted from a moving load is used to evaluate the before and after effect of bridge structure repairs.

The most significant difference between static theory and its dynamic counterpart is that inertial effect plays an important role in the latter case [12]. To better understand the dynamic response of a thin plate to a moving load, it is indispensable to analyze the dynamic effect of an impulse loading on a plate. Finite element procedures have been developed to carry out the response of a thin plate to dynamic loads with applications in airport design and pavement nondestructive testing and evaluation [3,13–15]. Kennedy [16] studied the steady-state response of a moving load on a beam on elastic foundation. Regarding analytical solutions, Deshun [17] applies the vibrational principle to solve the vibration of thick plates. The vertical vibration of an elastic plate on a fluid-saturated porous half space subjected to a harmonic load is investigated by Bo [18], in which Hankel transform is used to convert the governing equation to Fredholm integral equation of the second order and numerical calculation can then be carried out. The monograph by Fryba [19] contains a large body of work on moving load problem and establishes an extensive bibliography in this field.

This paper studies dynamic displacement of a thin plate caused by a moving harmonic load. The remainder of this paper is organized as follows. In Sec. 2, the governing equation is developed. In Sec. 3 the Green's function is obtained as a fundamental solution to construct dynamic response of the plate to moving harmonic loads. In Sec. 4, dynamic displacement of plate is established based on the convolution of the Green's function over time and space. In Sec. 5 the characteristic function of the thin plate is investigated. In Sec. 6 maximum response and condition under which resonance may occur are analyzed. The concluding remarks are made in Sec. 7.

2 The Governing Equation

Suppose that a thin plate is setting in an orthogonal x-y-z coordinate system. Denote $W(x,y,t)$ the displacement of the plate in z direction. Let $\mathbf{x}=(x,y,t)$ be the spatial and temporal coordinate. Three assumptions are often made to simplify mathematical description of the plate: (l) the strain component ε_z in the perpendicular direction of the plate 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 plate is zero. Under these assumptions, the well-known partial differential equation governing the displacement of the Kirchhoff plate on an elastic Winkler foundation is given by [3,15,16,20–23]

$$D\nabla^2\nabla^2 W(\mathbf{x}) + KW(\mathbf{x}) + \rho h\frac{\partial^2}{\partial t^2}W(\mathbf{x}) = F(\mathbf{x}) \tag{1}$$

where $F(\mathbf{x})$ is the impressed force, $\nabla^2=\partial^2/\partial x^2+\partial^2/\partial y^2$ is the Laplace operator; $D=Eh^3/[12(1-\nu^2)]$ is the stiffness of the plate; h is thickness of the plate, ρ is density of the plate; E is Young's modulus of elasticity, ν is Poisson ratio of the plate, and K is the modulus of subgrade reaction and is assumed to be constant to represent linear elasticity of the subgrade.

3 The Fundamental Solution

According to the theory of mathematical-physical equation, the fundamental solution of a partial differential equation is given by the Green's function, a solution corresponding to the situation where the right-hand side of Eq. (1) is in the form of the Dirac-delta function [24]

$$F(\mathbf{x}) = \boldsymbol{\delta}(\mathbf{x}-\mathbf{x}_0) \tag{2}$$

where $\mathbf{x}_0=(x_0,y_0,t_0)$ represents the source position in space and time; $\delta(\mathbf{x}-\mathbf{x}_0)=\delta(x-x_0)\delta(y-y_0)\delta(t-t_0)$ and $\delta(\cdot)$ is the Dirac-delta function, defined by

$$\int_{-\infty}^{\infty}\delta(x-x_0)f(x)dx = f(x_0) \tag{3}$$

Define the three-dimensional Fourier transform pair

$$\tilde{f}(\xi)\mathbf{F}[f(\mathbf{x})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\mathbf{x})\exp(-i\xi\mathbf{x})d\mathbf{x} \tag{4a}$$

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, October 17, 2003; final revision, May 20, 2005. Associate Editor: N. Sri Namachchivaya. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

 Copyright © 2005 by ASME

$$f(\mathbf{x}) = \mathbf{F}^{-1}[\tilde{f}(\xi)] = (2\pi)^{-3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{f}(\xi)\exp(i\xi\mathbf{x})d\xi \quad (4b)$$

where $\xi=(\xi,\eta,\omega)$, $\mathbf{F}[\cdot]$ and $\mathbf{F}^{-1}[\cdot]$ are Fourier transform and its inversion, respectively. To solve for Green's function, apply Fourier transform to both sides of Eq. (1)

$$D(\xi^2+\eta^2)^2\tilde{G}(\xi;\mathbf{x}_0) + K\tilde{G}(\xi;\mathbf{x}_0) - \rho h\omega^2\tilde{G}(\xi;\mathbf{x}_0) = \tilde{F}(\xi) \quad (5)$$

in which $\tilde{F}(\xi)$ is the Fourier transform of $F(\mathbf{x})$, and the displacement response $W(\mathbf{x})$ is replaced by the notation $G(\mathbf{x};\mathbf{x}_0)$ to indicate the Green's function. In the derivation of Eq. (5) the following property of Fourier transform is used

$$\mathbf{F}[f^{(n)}(t)] = (i\omega)^n\mathbf{F}[f(t)] \quad (6)$$

Also $\tilde{F}(\xi)$, the representation of $F(\mathbf{x})$ in the frequency domain can be obtained by taking Fourier transform of both sides of Eq. (2)

$$\tilde{F}(\xi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\mathbf{x}-\mathbf{x}_0)\exp(-i\xi\mathbf{x})d\mathbf{x} = \exp(-i\xi\mathbf{x}_0) \quad (7)$$

Here, the property of the Dirac-delta function, i.e., Eq. (3), is utilized in for evaluating the above integral. Substitute Eq. (7) into Eq. (5) and rearrange terms give

$$\tilde{G}(\xi;\mathbf{x}_0) = \exp(-i\xi\mathbf{x}_0)[D(\xi^2+\eta^2)^2 + K - \rho h\omega^2]^{-1} \quad (8)$$

The Green's function given by Eq. (8) is in the frequency domain and can be converted to the time domain by taking the inverse Fourier transform

$$\tilde{G}(\mathbf{x};\mathbf{x}_0) = (2\pi)^{-3}D^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp[(i\xi\mathbf{x}-\mathbf{x}_0)][(\xi^2+\eta^2)^2 + \bar{K} - \bar{m}\omega^2]^{-1}d\xi \quad (9)$$

where $\bar{K}=K/D$, $\bar{m}=\rho h/D$. Equation (9) is the Green's function of a plate resting on a Winkler elastic foundation. It serves as a fundamental solution of a partial differential equation and can be very useful when dealing with linear systems.

4 Dynamic Response to Moving Harmonic Load

According to the theory of linear partial differential equation, the dynamic response of a slab to an external load $F(\mathbf{x})$ can be constructed by integrating the Green's function in all dimensions [24]

$$W(\mathbf{x}) = \int_{-\infty}^{t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(\mathbf{x}_0)G(\mathbf{x};\mathbf{x}_0)d\mathbf{x}_0 \quad (10)$$

For a moving point load, $F(\mathbf{x})$ can be expressed as

$$F_e(\mathbf{x}) = P\delta(y)\delta(x-vt)\exp(i\Omega t) \quad (11)$$

For a moving line load, $F(\mathbf{x})$ can be expressed as

$$F_l(\mathbf{x}) = (2r_0)^{-1}P\delta(y)[H(x-vt+r_0) - H(x-vt-r_0)]\exp(i\Omega t) \quad (12)$$

where P is the magnitude of the load; v the load velocity; r_0 the half length of the line load; $H(x)$ the Heaviside step function, defined as $H(x)=1$ for $x>0$, $H(x)=0$ for $x<0$, and $H(x)=1/2$ for $x=0$.

The case of moving point load will be considered first. The dynamic displacement response of the slab under a moving point load can be given by Eq. (10) in which the load $F(\mathbf{x})$ is replaced by Eq. (11)

$$W_p(\mathbf{x}) = P\int_{-\infty}^{t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(y_0)\delta(x_0-vt_0)\exp(i\Omega t_0)G(\mathbf{x};\mathbf{x}_0)d\mathbf{x}_0 \quad (13)$$

Applying the properties of Dirac-delta function as showed in Eq. (3), the earlier equation becomes

$$W_p(\mathbf{x}) = P\int_{-\infty}^{t}\exp(i\Omega t_0)G(\mathbf{x};vt_0,0,t_0)dt_0 \quad (14)$$

Substituting the Green's function Eq. (9) into Eq. (14) and performing some mathematical manipulation, the displacement of the slab becomes

$$W_p(\mathbf{x}) = \bar{P}\int_{-\infty}^{t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\times\frac{\exp(i\Omega t_0)\exp\{i[\xi(x-vt_0)+\eta y+\omega(t-t_0)]\}}{(2\pi)^3[(\xi^2+\eta^2)^2+\bar{K}-\bar{m}\omega^2]}d\xi dt_0 \quad (15)$$

where $\bar{P}=P/D$. Substituting Eq. (3) and $\int_{-\infty}^{t}\exp[-i(v\xi+\omega-\Omega)t_0]dt_0 = 2\pi\delta(v\xi+\omega-\Omega)$ into Eq. (15) give

$$W_p(\mathbf{x}) = (2\pi)^{-2}\bar{P}\exp(i\Omega t)\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp\{i[\xi(x-vt)+\eta y]\}}{(\xi^2+\eta^2)^2+\bar{K}-\bar{m}(\Omega-v\xi)^2}d\xi d\eta \quad (16)$$

For a moving line load substituting Eqs. (9) and (12) into Eq. (10) and performing similar mathematical manipulation, we have

$$W_l(\mathbf{x}) = (2\pi)^{-2}\bar{P}\exp(i\Omega t)\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\sin\xi r_0\exp\{i[\xi(x-vt)+\eta y]\}}{\xi r_0[(\xi^2+\eta^2)^2+\bar{K}-\bar{m}(\Omega-v\xi)^2]}d\xi d\eta \quad (17)$$

In the derivation of Eq. (17) the Euler formula $\exp(i\varsigma)=\cos\varsigma + i\sin\varsigma$ and the following integral is used

$$\int_{-\infty}^{\infty}\frac{[H(x_0-vt+r_0)-H(x_0-vt-r_0)]\exp(-i\xi x_0)}{2r_0}dx_0 = \int_{vt-r_0}^{vt+r_0}\frac{\exp(-i\xi x_0)}{2r_0}dx_0 = \frac{\sin\xi r_0\exp(-i\xi vt)}{\xi r_0} \quad (18)$$

Because a moving point load is a special case of a moving line load, we should be able to derive Eq. (16) from Eq. (17) as a limiting case. Note that $\lim_{r_0\to 0}(\sin\xi r_0/\xi r_0)=1$, it is straightforward to verify this corollary by taking the limit on both sides of Eq. (17).

5 Characteristic Equation and Analytical Displacement

Dynamic displacement response of a plate to a moving line load is significantly affected by the denominator of the integrand in Eq. (17). Define characteristic equation of the plate as

$$z(\xi,\eta) = (\xi^2+\eta^2)^2 + \bar{K} - \bar{m}(\Omega-v\xi)^2 = 0 \quad (19)$$

Four zeros of this equation are

$$\eta_j(\xi) = (\pm)\{-\xi^2(\pm)[\bar{m}(\Omega-v\xi)^2-\bar{K}]^{1/2}\}^{1/2},\ j=1,\ldots,4 \quad (20)$$

It should be noted that here we require ξ to be real valued. To evaluate Eq. (17), complex theory is adopted, and therefore, these four roots of the characteristic equation become poles in the η

plane. We construct a closed contour to surround these poles. For $\eta \geqslant 0$ we select the closed contour in the upper half η plane, while for $\eta < 0$ the lower half η plane. Only the case of $\eta \leqslant 0$ is presented here due to limitation of paper length. The case of $\eta \geqslant 0$ can be derived in a similar way. Appling the theorem of residue to Eq. (17) gives dynamic displacement response of a plate subject to a moving line harmonic load

$$W_l(\mathbf{x}) = (2\pi)^{-2}\bar{P}\exp(i\Omega t)\int_{-\infty}^{\infty}\exp[i\xi(x-vt)] \times\left\{2\pi i\sum_{\mathrm{Im}(\eta)>0}\mathrm{res}\left[\frac{\sin\xi r_0\exp(i\eta y)}{\xi r_0[(\xi^2+\eta^2)^2+\bar{K}-\bar{m}(\Omega-v\xi)^2]}\right] + \pi i\sum_{\mathrm{Im}(\eta)=0}\mathrm{res}\left[\frac{\sin\xi r_0\exp(i\eta y)}{\xi r_0[(\xi^2+\eta^2)^2+\bar{K}-\bar{m}(\Omega-v\xi)^2]}\right]\right\}d\xi = (8\pi)^{-1}i\bar{P}\exp(i\Omega t)\sum_{\mathrm{Im}(\eta_j)>0}\int_{-\infty}^{\infty} \times\frac{\sin\xi r_0\exp[i\xi(x-vt)]\exp[i\eta_j(\xi)y]}{\xi r_0\eta_j(\xi)[\eta_j^2(\xi)+\xi^2]}d\xi \tag{21}$$

where res(·) stands for residue [25].

To evaluate Eq. (21) one needs to deal with another characteristic equations

$$\xi r_0\eta_j(\xi)[\eta_j^2(\xi)+\xi^2]=0 \tag{22}$$

Clearly $\xi=0$, or $\eta_j(\xi)=0$, or $\eta_j^2(\xi)+\xi^2=0$ could all lead to the existence of Eq. (22). According to Eq. (19), $\eta_j^2(\xi)+\xi^2=0$ is equivalent to $\bar{K}-\bar{m}(\Omega-v\xi)^2=0$, resulting two first-order poles in a complex ξ plane, i.e., $\xi_{1,2}=[\Omega\pm(\bar{K}/\bar{m})^{1/2}]/v$. To seek for zeros of $\eta_j(\xi)=0$, note that the application of the theorem of residue in Eq. (21) implies that ξ has to be real in the first place.

With this constraint in mind, one needs to identify two real roots from the equation $\eta_j(\xi)=0$ or equivalently

$$\xi^4-\bar{m}v^2\xi^2+2\bar{m}v\Omega\xi-\bar{m}\Omega^2+\bar{K}=0 \tag{23}$$

To understand the nature of this characteristic equation, define two quantities $S_+=\bar{K}-\bar{m}\Omega^2$ and $S_-=\bar{K}-\bar{m}(\Omega^2+v^4/4)$. Furthermore, define two functions

$$f_1(\xi)=\xi^4-\bar{m}v^2\xi^2-\bar{m}\Omega^2+\bar{K}=(\xi^2-\bar{m}v^2/2)^2+\bar{K}-\bar{m}(\Omega^2+v^4/4) \tag{24}$$

$$f_2(\xi)=-2\bar{m}v\Omega\xi \tag{25}$$

Equation (23) is identical to $f_1(\xi)=f_2(\xi)$. In order to analyze the effect of various parameters on characteristic Eq. (23), it is necessary to make some assumptions about parameters. Here, it is required that none of the parameters $\bar{K}$, $\bar{m}$, v, Ω be zero. For degraded situations (i.e., one of these parameters becomes zero), discussion will be given elsewhere. Under this assumption, $S_+=0$ and $S_-=0$ cannot simultaneously occur. First order derivatives of functions $f_1(\xi)$ and $f_2(\xi)$ are plotted in Fig. 1. Different S_+ and S_- scenarios are presented in Table 1.

Define $c=-\bar{m}v^2$, $d=2\bar{m}v\Omega$, $e=K-\bar{m}\Omega^2$. According to the theory of algebraic equation, four roots of Eq. (23) are identical to the roots of two quadratic equations

$$\xi^2+(2z-c)^{1/2}\xi+z-[2(2z-c)^{1/2}]^{-1}d=0 \tag{26}$$

$$\xi^2-(2z-c)^{1/2}\xi+z+[2(2z-c)^{1/2}]^{-1}d=0 \tag{27}$$

Here z is a real toot of the following equation [7]

$$Z(z)=z^3-cz^2/2-ez+ce/2-d^2/8=0 \tag{28}$$

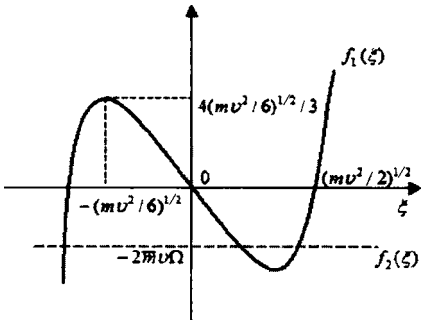


Fig. 1 First order derivatives of functions $f_1(\xi)$ and $f_2(\xi)$

Since $Z(c/2)=-d^2/8<0$ and $Z(+\infty)\to+\infty>0$, it is clear that there exists at least one real root satisfying $z>c/2$. Two approaches can be used to identify one real root of Eq. (28). Given the existence of at least one root in $(c/2, \infty)$, one approach relies on numerical iteration using search method such as Newton's algorithm [26]. Another approach gives analytical representation of the real root of Eq. (24) after carefully analyzing the distribution of various parameter combinations. For the time being, it is assume that z in Eqs. (26) and (27) has been obtained as known value. Define respectively determinants of Eqs. (26) and (27)

$$\Delta_1=-(2z+c)+2d(2z-c)^{-1/2} \tag{29}$$

$$\Delta_2=-(2z+c)-2d(2z-c)^{-1/2} \tag{30}$$

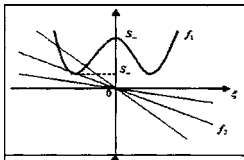
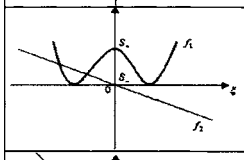
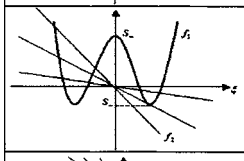
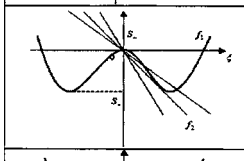
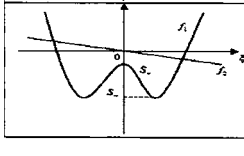
The four roots of Eqs. (26) and (27) can thus be given by

$$\xi_{1,2}=[-(2z-c)^{1/2}\pm\Delta_1^{1/2}]/2 \tag{31}$$

$$\xi_{3,4}=[(2z-c)^{1/2}\pm\Delta_2^{1/2}]/2 \tag{32}$$

Whether $\xi_{1,2}$ and $\xi_{3,4}$ are real or complex depends on the distribution of parameters. Two real roots of $\xi_{1,2}$ and $\xi_{3,4}$ are true zeros for $\eta_j(\xi)=0$ in Eq. (22).

Table 1 Features of characteristic equation

	Low Frequency, Low Speed ($S_+>0$ and $S_-<0$) $S_+>0$ results in $\Omega<(\bar{K}/\bar{m})^{1/2}$. $S_-<0$ results in $v>[4(\bar{K}-\bar{m}\Omega^2)/\bar{m}^2]^{1/4}$ Eq. (23) may have (i) two real negative roots plus two complex roots, or (ii) one duplicated real negative root plus two complex roots, or (iii) four complex roots
	Low Frequency, Critical Speed ($S_+>0$ and $S_-=0$) $S_+>0$ results in $\Omega<(\bar{K}/\bar{m})^{1/2}$ $S_-=0$ results in $v=[4(\bar{K}-\bar{m}\Omega^2)/\bar{m}^2]^{1/4}$ Eq. (23) has two real negative roots plus two complex roots
	Low Frequency, High Speed ($S_+>0$ and $S_-<0$) $S_+>0$ results in $\Omega<(\bar{K}/\bar{m})^{1/2}$ $S_-<0$ results in $v>[4(\bar{K}-\bar{m}\Omega^2)/\bar{m}^2]^{1/4}$ Eq. (23) may have (i) four real roots, or (ii) one duplicated real positive root plus two real negative roots, or (iii) two negative roots plus two complex roots
	Critical Frequency ($S_+=0$) $S_+=0$ results in $\Omega=(\bar{K}/\bar{m})^{1/2}$ Eq. (23) may have (i) four different real roots (one negative, one zero, two positive, or (ii) one duplicated real positive root, one zero, plus one negative root, or (iii) two real rot plus two complex roots
	High Frequency ($S_+<0$) $S_+<0$ results in $\Omega>(\bar{K}/\bar{m})^{1/2}$ Eq. (23) has two real roots (one positive and one negative) plus two complex roots

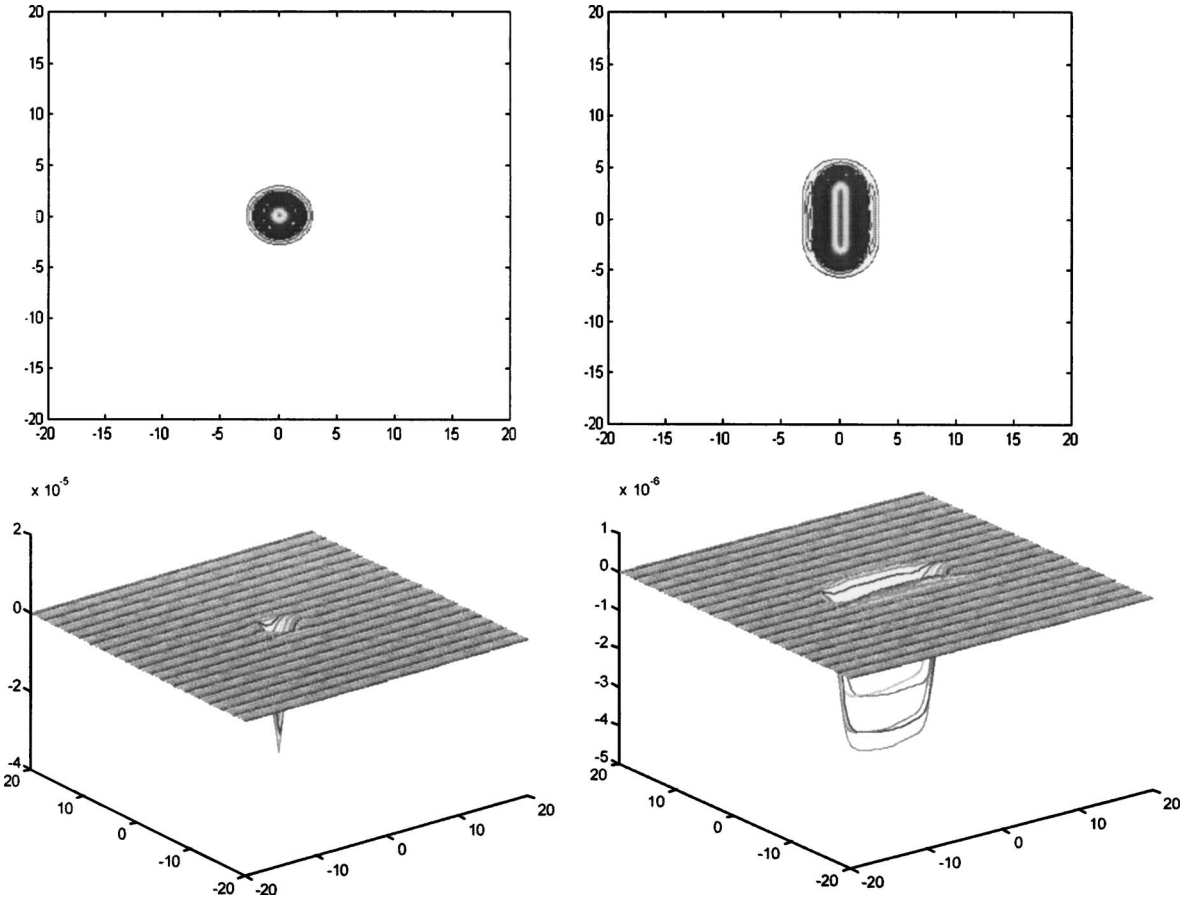


Fig. 2 Three-dimensional dynamic frequency response function (left: $r_0=1\times10^{-5}$ m, $v=30$ m/s, $\Omega=10$ Hz; right: $r_0=1$m, $v=30$ m/s, $\Omega=10$ Hz)

In summary, five poles of the integral in Eq. (21) have been identified. Similar to the derivation of Eq. (21), either numerical methods or residue theorem can be applied to evaluate Eq. (21). In the latter case, for $y\geqslant 0$ the constructed closed contour is the upper half ξ plane. One needs to specify the branch where this contour is located due to the nature of the denominator in Eq. (21). The determination of critical speed and frequencies of the moving load is based on the analysis of the denominator of an expression for the plate response. Pesterev et al. [28,29] noted that this could be misleading in some related problem, for instance, moving loads on beams.

Define Galilean transform for a follow-up coordinate system $\mathbf{x}'=(x',y')$ as $x'=x-vt$ and $y=y'$. According to Eq. (21) further define $W_l(\mathbf{x})=H_l(r_0,v,\Omega)\exp(i\Omega t)$. Numerical computation has been conducted in the follow-up coordinate system. For typical cement concrete pavement structures [27], the parameters are $\bar{P}=205.5482/\mathrm{m}$, $\bar{K}=229240.8/\mathrm{m}^4$, and $\bar{m}=0.371248\ \mathrm{kg}/(\mathrm{Nm}^4)$. Frequency response function $H_l(r_0,v,\Omega)$ is shown in Fig. 2 for different load speeds, frequencies and distributions, in which the unit for the dynamic displacement is meter.

6 Maximum Response, Static Solution, and Critical Speed

It is of great interest for both theoretical study and engineering practice to investigate the maximum response and where it occurs. Clearly the origin of this follow-up coordinate system travels with the moving load at the same speed. In order to reach the maximum, the gradient of the real part of the nominator in Eq. (17), i.e., $g(x',y')=\mathrm{Re}\{\exp[i(\xi x'+\eta y')]\}=\cos(\xi x'+\eta y')$ must become zero. In other words

$$\nabla g^T(\mathbf{x}')=(\partial g/\partial x',\partial g/\partial y')=(0,0) \tag{33}$$

Note that Eq. (33) has to be satisfied with respect to ξ and η simultaneously. By substituting $g(\mathbf{x}')=\cos(\xi x'+\eta y')$ into Eq. (33), it is found that the only nontrivial solution of these two equations is $\mathbf{x}'=(x',y')=(0,0)$, resulting in the maximum displacement response

$$W(\mathbf{x}')=(8\pi)^{-1}i\bar{P}\sum_{\mathrm{Im}(\eta_j)>0}\int_{-\infty}^{\infty}\frac{\sin\xi r_0}{\xi r_0\,\eta_j(\xi)[\eta_j^2(\xi)+\xi^2]}d\xi \tag{34}$$

The maximum occurs at the origin of the following-up coordinate system, which indicates that the displacement beneath the center of the moving line load always has the greatest magnitude and travels at the same speed as does the moving load. Because Eq. (33) is independent of the load distribution, the same conclusion applied to the situation of a moving point load. The maximum displacement response to a moving point harmonic load can be obtained by taking limit to Eq. (34) as $r_0\to 0$

$$W_P(\mathbf{x}')=(8\pi)^{-1}i\bar{P}\sum_{\mathrm{Im}(\eta_j)>0}\int_{-\infty}^{\infty}\frac{1}{\eta_j(\xi)[\eta_j^2(\xi)+\xi^2]}d\xi \tag{35}$$

Here $\lim_{r_0\to 0}\sin\xi r_0/\xi r_0=1$ has been used.

The correctness of dynamic displacement response Eq. (16) to a moving load can also be verified by comparing the degraded situ-

ation with static solution of the slab under a point load, which has been well-known or years [23]. To do so, let load velocity and frequency $v=0$, $\Omega=0$ in Eq. (16), leading to

$$W_p(\mathbf{x}) = (2\pi)^{-2}\bar{P}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp[i(\xi x+\eta y)]}{(\xi^2+\eta^2)^2+\bar{K}}d\xi d\eta \qquad (36)$$

Define 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 Eq. (36) gives

$$W_p(\mathbf{x}) = (2\pi)^{-2}\bar{P}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\cos[r\zeta\sin(\theta+\psi)]}{\zeta^4+\bar{K}}\zeta d\zeta d\psi \qquad (37)$$

It is known that the Bessel function can be expressed as $J_0(z)=(2\pi)^{-1}\int_0^{2\pi}\cos(z\cos\phi)d\phi$ [25], in which $J_0(\cdot)$ is the zeroth order of Bessel function of the first kind. Adopting this representation, Eq. (37) can be rewritten as

$$W_p(\mathbf{x}) = (2\pi)^{-1}P\int_{-\infty}^{\infty}\frac{J_0(r\zeta)}{D\zeta^4+K}\zeta d\zeta \qquad (38)$$

This is identical to the static solution given in Ref. [23].

In one-dimensional structures such as a beam, it has been experimentally observed and theoretically proved that there exists a critical speed [16,30]. In two-dimensional structures such as a plate, critical speed and critical frequency also exist, under which resonance may occur. To analyze critical velocity and critical frequency, the key lies in the characteristic Eq. (21), or more specifically in the second-order pole of the equation $\eta_j(\xi)=0$. Mathematically a duplicated root of characteristic Eq. (23) requires one of the determinants of Eqs. (26) and (27), namely, either $\Delta_1=0$ or $\Delta_2=0$ must be zero. This leads to the following equation

$$(2z-c)(2z+c)^2-4d^2=0 \qquad (39)$$

As stated earlier, z in Eq. (39) is a real root of Eq. (28) and thus a function of $\bar{K},\bar{m},v,\Omega$. Equation (39) therefore gives the condition of these parameters under which a second-order pole exists in the integrand of Eq. (21). Resonance (infinite displacement) can be observed if the combination of $\bar{K},\bar{m},v,\Omega$ satisfies Eq. (39).

The satisfaction of Eq. (39) gives precisely the condition under which the so-called critical velocity and frequency can be determined. From Table 1, however, one can develop an approximate relationship between the critical velocity and critical frequency. As it can be seen from Table 1 that in the cases of low frequency and low speed ($S_+>0$ and $S_-<0$), low frequency and high speed ($S_+>0$ and $S_-<0$) and critical frequency ($S_+=0$), a duplicated root can occur. This corresponds to a second-order pole in the complex η plane. The value of function $f_1(\xi)$ with which function $f_2(\xi)$ is tangent can be approximated by the minimum of $f_1(\xi)$. According to Eq. (24), this minimum is min $f_1(\xi)=\bar{K}-\bar{m}(\Omega^2+v^4/4)$, corresponding to $\xi=\pm(\bar{m}v^2)/2$. Because at this point function $f_2(\xi)$ approximately has the same value, taking Eq. (25) into account, we have $[\pm(\bar{m}v^2)/2](-2\bar{m}v\Omega)\approx\bar{K}-\bar{m}(\Omega^2+v^4/4)$ or equivalently

$$\bar{K}-\bar{m}(\Omega^2+v^4/4)\pm\bar{m}^2v^3\Omega\approx 0 \qquad (40)$$

Equation (40) can be thought of as an approximate equation governing the relationship between critical velocity and critical frequency. For instance, if load frequency is $\Omega=(\bar{K}/\bar{m})^{1/2}$, according to Eq. (40) the critical velocity becomes $v_{\text{critical}}\approx\bar{m}\Omega=(\bar{m}\bar{K})^{1/2}$.

7 Concluding Remarks

The displacement response of a thin plate resting on a Winkler elastic foundation subject to a moving harmonic load with line distribution is studied using integral transform method. The solution is constructed as a convolution of dynamic Green's function of the plate. It is shown that there exist critical velocity and critical frequency, which are governed by the characteristic equation of the thin plate. An analytical approximate equation is developed to describe the relationship between critical load velocity and critical frequency. It is also found that the maximum displacement response occurs at the center of the moving load and travels at the same speed with the load.

Acknowledgment

This study is supported in part by the National Science Foundation through Project No. CMS-0408390, to which the author is very grateful. The author also wishes to thank two anonymous referees for their valuable comments, which improve the presentation of this paper.

References

[1] Sousa, J. B., Lysmer, J., Chen, S. S., and Monismith, C. L., 1987, "Response of Asphalt Concrete Pavements to Dynamic Truck Loads," Report to Transportation Laboratory, University of California, Berkeley.

[2] Monismith, C. L., Sousa, J., and Lysmer, J., 1988, "Modern Pavement Design Technology Including Dynamic Load Conditions," Trans. ASAE, No. 881856.

[3] Sun, L., 1998, "Theoretical Investigations on Vehicle-Ground Dynamic Interaction, Southeast University, Nanjing, China," *Final Report Prepared for National Science Foundation of China*.

[4] Sun, L., and Deng, X., 1998, "Predicting Vertical Dynamic Loads Caused by Vehicle-Pavement Interaction," J. Transp. Eng., **124**(5), pp. 470–478.

[5] Sun, L., and Kennedy, T. W., 2002, "Spectral Analysis and Parametric Study of Stochastic Pavement Loads," J. Eng. Mech., **128**(3), pp. 318–327.

[6] Deng, X., and Sun, L., 2001, *Dynamics of Vehicle-Pavement Interaction*, People's Transport Publishing, Beijing.

[7] Sun, L., 2002, "Optimum Design of 'Road-Friendly' Vehicle Suspension Systems Subject to Rough Road Surface," Appl. Math. Model., **26**(5), pp. 635–652.

[8] Bush, A. J., 1980, "Nondestructive Testing for Light Aircraft Pavements, Phase II. Development of the Nondestructive Evaluation Methodology, Federal Aviation Administration," Final Report FAA RD-80-9.

[9] Scullion, T., Uzan, J., and Paredes, M., 1990, "A Microcomputer-Based Backcalculation System," Transp. Res. Rec., No. 1260 MODULUS.

[10] Uzan, J., and Lytton, R., 1990, "Analysis of Pressure Distribution under Falling Weight Deflectometer Loading," J. Transp. Eng., **116**(2), pp. 246–250.

[11] 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.

[12] Sun, L., and Greenberg, B., 2000, "Dynamic Response of Linear Systems to Moving Stochastic Sources," J. Sound Vib., **229**(4), pp. 957–972.

[13] Taheri, M. R., 1986, "Dynamic Response of Plates to Moving Loads," Ph.D. Thesis, Purdue University, West Lafayette, IN.

[14] 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.

[15] Zaghloul, S. M., White, T. D., Dmevich, 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, D. C.

[16] Kenney, J. T., 1954, "Steady-State Vibrations of Beam on Elastic Foundation for Moving Load," ASME J. Appl. Mech., **21**, pp. 359–364.

[17] Deshun, Z., 1999, "A Dynamic Model for Thick Elastic Plates," J. Sound Vib., **221**(2), pp. 187–203.

[18] Bo, J., 1999, "The Vertical Vibration of an Elastic Circular Plate on a Fluid-Saturated Porous Half Space," Int. J. Eng. Sci., **37**, pp. 379–393.

[19] Fryba, L., 1999, *Vibration of Solids and Structures under Moving Loads*, Thomas Telford, London.

[20] Westergaard, H. M. S., 1926, "Computation of Stresses in Concrete Roads," Highw. Res. Board, Proc. Annu. Meet. **5**(I), pp. 90–112.

[21] Yoder, E. J., and Witczak, M. W., 1975, *Principles of Pavement Design*, Wiley, New York.

[22] Hass, R., Hudson, W. R., and Zniewski, J., 1994, *Modern Pavement Management*, Krieger, Malabar, FL.

[23] Zhu, Z., Wang, B., and Guo, D., 1984, *Pavement Mechanics*, Renming Jiaotong, Beijing, China.

[24] Eringen, A. C., and Suhubi, E. S., 1975, *Elastodynamics*, Academic Press, New York, Vols. I and II.

[25] Watson, G. N., 1966, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, London.

[26] Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., 1993,

Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, Cambridge.
[27] Kim, S. M., and Roesser, J. M., 1996, "Dynamic Response of Pavement Systems to Moving Loads," Research Report, 1422-2, Center for Transportation Research, University of Texas at Austin.
[28] Pesterev, A. V., Yang, B., Bergman, L. A., and Tan, C. A., 2001, "Response of Elastic Continum Carrying Multiple Moving Oscillators," J. Eng. Mech., **127**(3), pp. 260–265.
[29] Pesterev, A. V., Tan, C. A., and Bergman, L. A., 2001, "A New Method for Calculating Bending Moment and Shear Force in Moving Load Problems," ASME J. Appl. Mech., **68**(1), pp. 252–259.
[30] Sun, L., 2001, "Closed-Form Representation of Beam Response to Moving Line Loads," ASME J. Appl. Mech., **68**(2), pp. 348–350.

Journal of Applied Mechanics

Technical Briefs

Finite Duration Impacts With External Forces

D. Dane Quinn
Department of Mechanical Engineering, The University of Akron, Akron, OH 44325-3903
e-mail: quinn@uakron.edu

This work considers the effect of external forces during finite duration collisions using an incremental model of impact. The deformation of the "rigid" body is modeled through an elastic element and the time interval over which contact occurs is of finite duration. Moreover, the work done by the external forces is nonzero during the collision. This model allows for an equivalent coefficient of restitution e to be identified. In the presence of a constant external force the coefficient of restitution depends not only on the system parameters, but the initial relative velocity at the point of impact. For external forces which tend to bring the colliding bodies together, the colliding bodies remain in contact for sufficiently small impact velocities ($e=0$) *while for larger incoming speeds, the coefficient of restitution is positive. This state dependent restitution arises from the coupling of external forces to the collision model, and is not seen in more familiar models of impact. Finally, based on the results of the experimental system and the incremental model, the standard algebraic model of restitution is modified to include these finite duration effects.* [DOI: 10.1115/1.1875552]

1 Introduction

Impact describes the contact and interaction between two masses whose dynamics, in some past time, were independent of one another. These collisions typically involve impulsive forces and moments which lead to deformation in the contacting masses. However, rigid-body dynamics allows no such material deformation and therefore the details of the impactive event are replaced by constitutive relations, or impact laws, which approximate the mechanics of the collision process. Algebraic impact laws typically invoke a *coefficient of restitution* to describe the impactive event [1–3]. However, the development and application of common algebraic descriptions of impact assume that during the collision only the contact forces do work on the system. Unconstrained pairwise impacts, impacts with a rigid, immovable surface, and collisions in the presence of kinematic constraints are three such examples.[1] During the collision interval, external, nonimpulsive forces are neglected and their contribution to the impulse acting on either of the colliding bodies is not considered. This assumption is often justified by noting that the time interval across the collision is short, and in the limit of an instantaneous collision the contribution to the impulse from external forces vanishes, provided the external forces are nonimpulsive.

1.1 Role of External Forces. This work focuses on collisions in which the contribution to the total impulse from external forces is nonzero. The resulting external impulses may significantly affect the collision dynamics if they are of the same order of magnitude as the collisional impulse. Tatara [4] experimentally showed that both the measured coefficient of restitution and impact duration depend on the magnitude of these external forces.

In such a situation, the impact can be modeled as a finite-time collision. This work considers a simple incremental model which allows for an analytic investigation of the resulting dynamics. Similar incremental models have been widely used to investigate collisions for which only the contact forces are considered, i.e., the total momentum is conserved [5–7]. To conclude, numerical results are presented based on a nonlinear contact model proposed by Hunt and Crossley [8] which qualitatively exhibits the same phenomena as seen in the simpler model.

There exist several situations in which the external impulses that act on the colliding bodies during the impact are non-negligible. For example, in impacts between one or more bodies with significant compliance, the impact duration can be of sufficient duration to lead to non-negligible external impulses. Also, if a pairwise collision is subject to external impulsive loads during the impact interval the total momentum is not expected to be conserved. Such a situation is related to simultaneous impacts [9,10] when a body is subjected to impulsive loads induced by the multiple single-point contacts and the momentum between any pair of colliding bodies is not conserved. A third example occurs as a mass comes to rest through a sequence of repeated impacts. As the impact velocity approaches zero, the impulse from the impact forces weakens. In this case although the impulse arising from the external forces may remain small, it can be significant in comparison to the collisional impulse. This final case is examined to experimentally motivate the analytical model which follows.

This work invokes a one-dimensional incremental model of a collisional event [9]. In contrast to studies that focus on the dynamics of repeated impacts [11,12], the current work concerns the dynamic behavior across a single impact. The impact is characterized by a discontinuity in the resulting dynamical system, while the loss of contact introduces a discontinuity in the second derivative of position, although the vector field at this point remains once differentiable. From this system, a kinematic coefficient of restitution is obtained that depends not only on the parameters of the system, but on the state of the system at the impact, including the initial impact velocity and the external force. This work seeks

Contributed by the Applied Mechanics Division of The American Society of Mechanical Engineers for publication in the ASME Journal of Applied Mechanics. Manuscript received by the ASME Applied Mechanics Division, March 29, 2004; final revision, September 30, 2004. Associate Editor: N. Sri Namachchivaya.

[1] Each of these examples leads to an equivalent equation of motion for the contact deformation during the collision.

 Copyright © 2005 by ASME

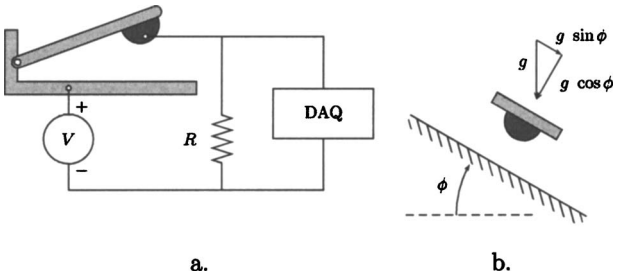


Fig. 1 Experimental setup

to develop alternative algebraic laws applicable in the above-mentioned situations from consideration of incremental models of the collision.

2 Experimental Motivation

An experimental apparatus was constructed to illustrate collision dynamics for very low speed impacts, e.g., in the final collisions of repeated impacts. The experimental system, sketched in Fig. 1, consists of a pendulum ($m=2$ kg, $l=25$ cm) oriented so as to contact a solid steel base in the horizontal position to allow for one-dimensional impact. A hemispherical impact specimen ($r=2$ cm) is attached to the steel block, roughly at the assembly's center of percussion (this location will vary slightly with the mass of the specimen). Finally, the experimental apparatus can be inclined through an angle ϕ to vary the gravitational force in the direction normal to the contact surface.

To develop an *a posteriori* estimation of the coefficient of restitution, the incoming and outgoing velocities of the specimen are required, which can be estimated from the time interval between contacts. For periods when the mass is in flight, the equation of motion, assuming small displacements, reduces to $\ddot{x}=-g\cos\phi$, where x is the displacement of the impact specimen above the impact surface and $g\cos\phi$ is the gravitational acceleration normal to the collision plane. Thus, if the velocity of impact specimen at the k-th liftoff is v_k, the duration of the k-th period of flight, denoted by Δt_k, is simply

$$\Delta t_k \sim \frac{2}{g\cos\phi} v_k.$$

From this, an experimental coefficient of restitution for the k-th impact can be determined from the duration intervals as

$$e_k \equiv -\frac{v_{k+1}}{v_k} \sim \frac{\Delta t_{k+1}}{\Delta t_k}. \tag{1}$$

Therefore, the transition points of the process must be determined, that is, the times at which collisions initiate and terminate, without regard for the contact forces that occur during the collision. The time intervals are obtained by applying a positive voltage to the apparatus with a specimen electrically isolated from the base. A metallic foil is bonded to the specimen and the system is connected to the data acquisition system as represented in Fig. 1(*a*). During each individual collision, when the specimen is in contact with the base, the electrical circuit is closed and a positive voltage is measured. In contrast, during the noncontact intervals, the circuit is open and zero voltage is recorded. Impact data was sampled at a rate of 2 kHz, with the data acquisition performed by LABVIEW. Finally, the specimens were released from a height of approximately 25 mm. Sample voltage measurements are shown in Fig. 2 for two different inclinations of the contact plane.

Finally, these identified time intervals are used to determine the experimentally measured coefficient of restitution as shown in Fig. 3, versus impact velocity (e_k versus v_k) according to Eq. (1). These experimental tests clearly show that, for very low velocity impacts, the measured coefficient of restitution depends on both the impact velocity and the orientation of the contact plane with respect to gravity. Motivated by these observations, a model for such impacts is formulated below.

3 Incremental Model

Consider a body impacting a rigid surface. It is assumed that the deformation due to the collision is localized in a small region,

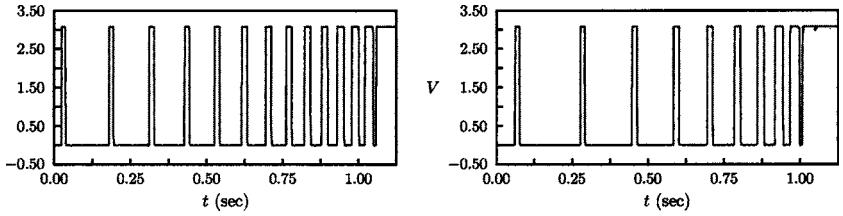


Fig. 2 Representative impact data (voltage versus time)

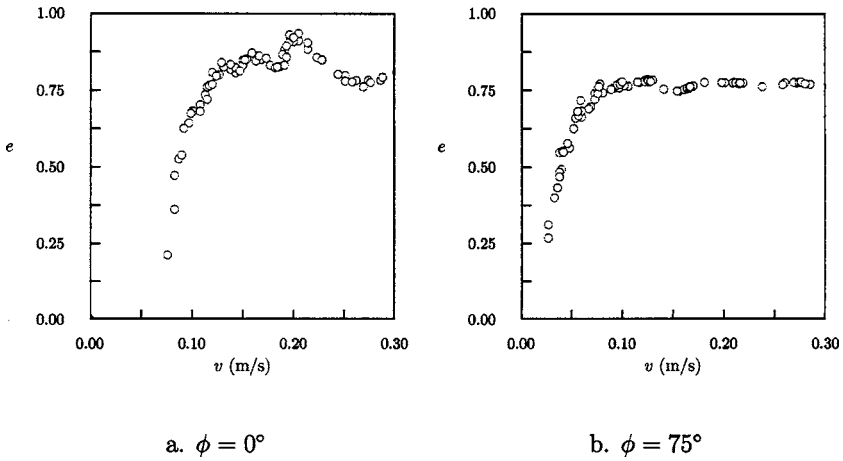


Fig. 3 Experimental coefficient of restitution. Each panel is the compilation of ten individual tests

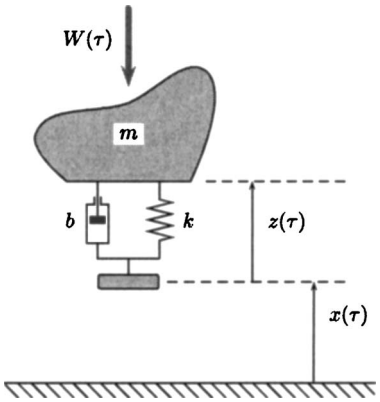


Fig. 4 Incremental model

while the remainder of the body undergoes rigid-body motion. Therefore the body is decomposed into a rigid body of mass m and a massless base, with the contact stiffness and dissipation represented by elastic and damping elements. Moreover, m is subject to an applied force $W(\tau)$ (see Fig. 4).

As shown in Fig. 4, $x(\tau)$ measures the position of the massless contact point, while $z(\tau)$ represents the deformation across the contact zone. Note that if $x(\tau)=0$, the mass is in contact with the surface. The resulting model is governed by the following system of equations:

$$m(z''+x'') = -bz' - kz - W(\tau),$$

$$0 = bz' + kz + \tilde{N},$$

where $d/d\tau \equiv (\,)'$ and $\tilde{N}$ represents the contact force.

In the Appendix, Sec. I, these equations are nondimensionalized to yield

$$\begin{aligned}(\ddot{z}+\ddot{x}) + 2\zeta\dot{z} + z &= -F(t),\\ 0 &= 2\zeta\dot{z} + z + N,\end{aligned} \tag{2}$$

with $d/dt \equiv (\,)$. Finally, the contact force $N(t)$ obeys the inequality constraint $N \geqslant 0$. When the body is not in contact with the surface, $N(t)=0$ and the equations governing free motion become

$$\ddot{z} + \ddot{x} = -F(t),$$

$$0 = 2\zeta\dot{z} + z.$$

Contact initiates when, during a period of free motion, $x(t)=0$.

During contact, the kinematic constraint implies that $x(t)\equiv 0$, so that the equation governing the dynamical behavior of the rigid body is described by a single-degree-of-freedom oscillator

$$\ddot{z} + 2\zeta\dot{z} + z = -F(t),$$

and the contact force becomes

$$N(t) = -(2\zeta\dot{z} + z).$$

$N(t)$ is required to be non-negative, so the termination condition for the collision can be expressed as $N(t_\ell)=0$, or

$$0 = (2\zeta)\dot{z}(t_\ell) + z(t_\ell),$$

where $t=t_\ell$ is the time at which separation occurs and the collision terminates [6]. With this description, the relative displacement at termination $z(t_\ell)$ is not necessarily zero.

While a closed-form solution for $z(t)$ exists while the mass is in contact with the surface, the time at which the collision terminates is determined by a transcendental equation. To determine the dynamical behavior these closed form solutions are used, while numerically solving the transitional conditions which determine the transition between these free and constrained motion using a Newton-Raphson algorithm. This procedure is analogous to that used in Shaw and Shaw [13].

4 Coefficient of Restitution

A coefficient of restitution can be identified when $F(t)=F_0 \equiv$ constant, that is, the external term is constant during the duration of the collision. Recall that from the Appendix, the time scale of the nondimensional equations is consistent with the impact duration. Thus, provided that the external force does not vary on this time scale, considering $F(t)$ to be constant during the collision is consistent.

Therefore, analogous to Newton's coefficient of restitution

$$e = -\frac{\dot{z}(t_\ell)}{\dot{z}(t_i)}. \tag{3}$$

However, the impact occurs over a finite time interval $\Delta t = t_\ell - t_i$ which must be determined from the model.

To calculate e, the equations of motion governing contact

$$\ddot{z} + (2\zeta)\dot{z} + z = -F_0,$$

are solved subject to the initial conditions

$$z(t_i) = 0, \quad \dot{z}(t_i) = -\dot{z}_0.$$

Without loss of generality $t_i=0$ can be assumed. During contact, these equations can be solved in closed-form as

$$\begin{aligned}z(t) = -F_0 &- \frac{\dot{z}_0}{2\sqrt{\zeta^2-1}}\left[e^{(-\zeta+\sqrt{\zeta^2-1})t} - e^{(-\zeta-\sqrt{\zeta^2-1})t}\right]\\ &- \frac{F_0}{2\sqrt{\zeta^2-1}}\left[\frac{e^{(-\zeta+\sqrt{\zeta^2-1})t}}{(-\zeta+\sqrt{\zeta^2-1})} - \frac{e^{(-\zeta-\sqrt{\zeta^2-1})t}}{(-\zeta-\sqrt{\zeta^2-1})}\right].\end{aligned}$$

These equations can then be used to determine the duration of the impact t_ℓ. This time can be used to determine the velocity at liftoff and hence the equivalent coefficient of restitution obtained from this incremental model. However, t_ℓ is determined from a transcendental equation.

Zero External Force. When the applied load F_0 vanishes, the coefficient of restitution depends only on the damping ratio ζ, and can be solved in closed form to yield

$$e_\infty = (\zeta - \sqrt{\zeta^2-1})^{(2\zeta)/(\sqrt{\zeta^2-1})}, \tag{4}$$

while the duration of impact reduces to

$$t_{\ell,\infty} = \ln[(\zeta + \sqrt{\zeta^2-1})^{2/(\sqrt{\zeta^2-1})}]. \tag{5}$$

For $\zeta \in [0,1)$, $t_{\ell,\infty}$ and e_∞ reduce to

$$t_{\ell,\infty} = \frac{2}{\sqrt{1-\zeta^2}}\phi, \quad e_\infty = \exp\left(-\frac{2\zeta\phi}{\sqrt{1-\zeta^2}}\right), \quad \tan\phi = \frac{\sqrt{1-\zeta^2}}{\zeta},$$

As seen in Fig. 5, the value of e_∞ decreases monotonically as ζ, the damping ratio, increases. Thus in this limit e_∞ can be identified with an equivalent damping ratio. For example, with $\zeta=0.10$, $e_\infty = 0.744$. Moreover, in the absence of external influences, the equivalent coefficient of restitution is independent of the impact velocity for this model.

Nonzero External Force. Unlike the unforced response, for $F_0 \neq 0$, the coefficient of restitution depends not only on the damping ratio ζ, but on the quantity ν, the ratio of the nondimensional impact velocity to the nondimensional applied load directed normal to the contact surface, that is

$$\nu \equiv \frac{\dot{z}_0}{F_0}.$$

Unfortunately, the final collision velocity, and therefore the *a posteriori* coefficient of restitution e, can no longer be determined in

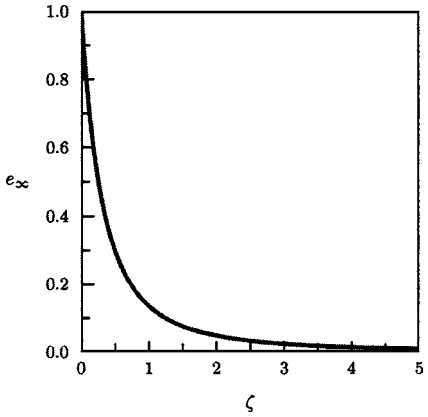


Fig. 5 Equivalent Newtonian restitution ($F_0=0$)

closed form. Rather, this quantity is evaluated numerically. In Fig. 6, e is calculated for three different values of ζ as ν is varied. In each figure for sufficiently small impact velocities the mass does not leave the surface—the equivalent coefficient of restitution is zero. For a fixed F_0, as the impact velocity increases, there exists a point for which the mass leaves the surface following the collision, and the coefficient of restitution becomes nonzero. As the impact velocity is further increased, or equivalently the applied load is decreased, the coefficient of restitution approaches the limiting value e_∞ as described above (the dashed line in each panel). When the external force vanishes, the coefficient of restitution remains constant as $\dot{z}_0$ varies, at a value e_∞ as defined in Eq. (4) (see Fig. 5). Only in this case can the coefficient of restitution be considered as a "material parameter," a quantity independent of the initial impact velocity $\dot{z}_0$.

Thus with the introduction of external forces into the incremental model and nonzero energy dissipation ($\zeta>0$), the equivalent coefficient of restitution vanishes for sufficiently small values of ν. The critical point at which the equivalent restitution vanishes is defined by the conditions

$$N(t_{\ell,\mathrm{cr}})=0, \quad \dot{N}(t_{\ell,\mathrm{cr}})=0,$$

which, taken together, imply the nonzero conditions for the position and velocity

$$z(t_{\ell,\mathrm{cr}})=(-4\zeta^2)F_0, \quad \dot{z}(t_{\ell,\mathrm{cr}})=(2\zeta)F_0.$$

These conditions lead to an implicitly defined relationship between ζ, the nondimensional damping ratio, and ν_{cr}, the critical ratio between the impact velocity and the external load of the form

$$2\sqrt{\zeta^2-1}=\{(\zeta-\sqrt{\zeta^2-1})^3 e^{(\zeta+\sqrt{\zeta^2-1})t_{\ell,\mathrm{cr}}}-(\zeta+\sqrt{\zeta^2-1})^3 e^{(\zeta-\sqrt{\zeta^2-1})t_{\ell,\mathrm{cr}}}\},$$

$$2\nu_{\mathrm{cr}}\sqrt{\zeta^2-1}=\{(\zeta-\sqrt{\zeta^2-1})^2 e^{(\zeta+\sqrt{\zeta^2-1})t_{\ell,\mathrm{cr}}}-(\zeta+\sqrt{\zeta^2-1})^2 e^{(\zeta-\sqrt{\zeta^2-1})t_{\ell,\mathrm{cr}}}\},$$

where $t_{\ell,\mathrm{cr}}$ is the unknown time interval of the impact as the equivalent restitution vanishes. The resulting relationship between ζ and ν_{cr}, obtained by numerically evaluating these conditions, is shown in Fig. 7. For small ζ, ν_{cr} is numerically observed to scale

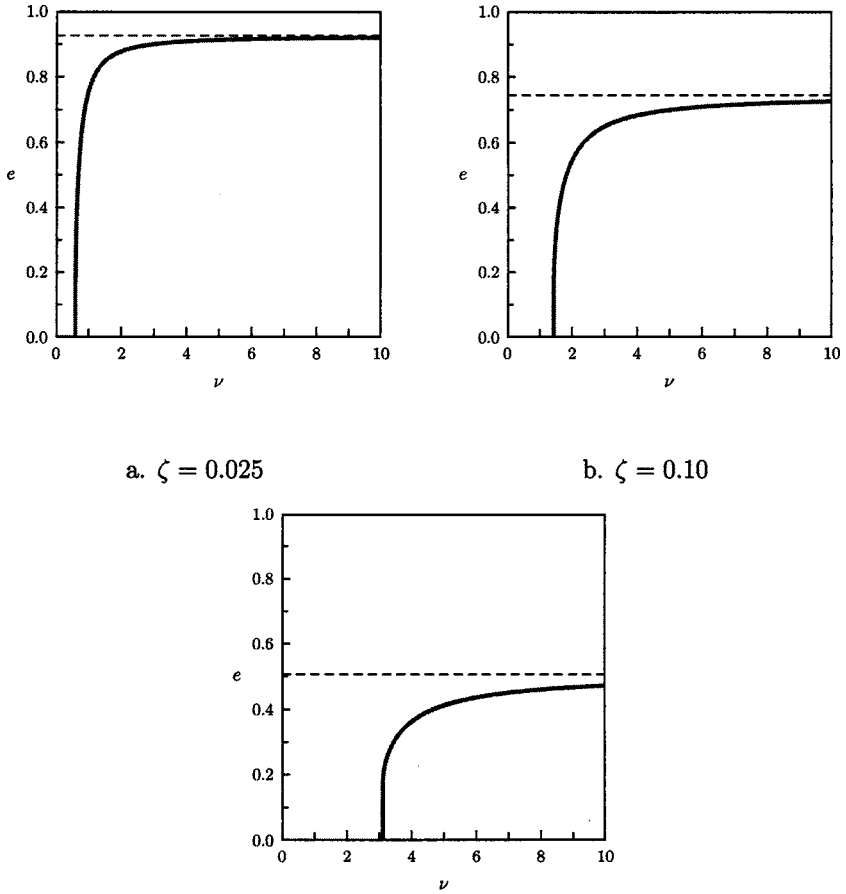


Fig. 6 Equivalent coefficient of restitution. The value of e_∞ is shown by the dashed line.

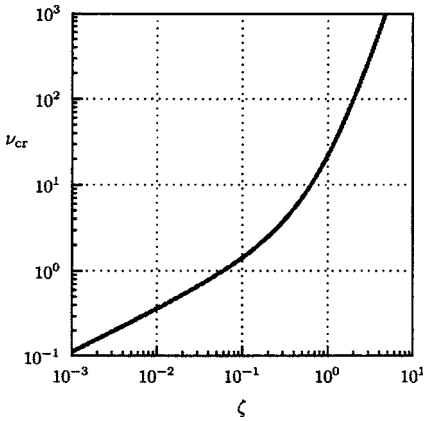


Fig. 7 Critical ν below which the coefficient of restitution vanishes

like $\sim\sqrt{\zeta}$, while for ζ large the curve increases as $\sim\zeta^3$.

For a specified value of ζ, the corresponding value of ν_{cr} characterizes the point at which the external force plays a significant role in the relative velocity after the collision. Because ζ and e_∞ are one-to-one (see Fig. 5), ν_{cr} can be identified with e_∞, the coefficient of restitution in the absence of external forces, as shown in Fig. 8. The value of ν_{cr} increases as e_∞ decreases.

As $\nu \to \nu_{cr}$, although the equivalent coefficient of restitution vanishes, the nondimensional duration of the impact remains finite. In Fig. 9, t_ℓ is shown as ζ varies. The solid curve represents $t_{\ell,cr}$, the collision duration for $\nu=\nu_{cr}$. For comparison the dashed line shows the value of $t_{\ell,\infty}$, the collision duration in the absence of external forces ($F_0=0$). For $\zeta=0$, $t_{\ell,cr}=2\pi$ and $t_{\ell,\infty}=\pi$. As ζ increases, $t_{\ell,cr}$ and $t_{\ell,\infty}$ can be approximated as

$$t_{\ell,cr} \sim \ln((2\zeta)^{3/\zeta}),$$
$$t_{\ell,\infty} \sim \ln((2\zeta)^{2/\zeta}), \quad \text{as } \zeta \to \infty.$$

As ζ increases the ratio of these two times rapidly converges to 3/2 and both impact times approach zero.

This incremental model contains two nondimensional parameters which describe the post-collisional behavior—the damping ratio ζ and the quantity $\nu \equiv \dot{z}/F$. In terms of dimensional quantities these can be expressed as

$$\zeta = \frac{b}{2\sqrt{km}}, \quad \nu = \sqrt{km}\frac{z'}{W}, \tag{6}$$

where k represents a "contact stiffness," and b describes the "contact damping." From the latter of these relations, the value of z'_{cr}, the dimensional impact velocity at which the external force becomes significant may be determined to be

$$z'_{cr} = \frac{W}{\sqrt{km}}\nu_{cr}.$$

Typically this implies that the external forces cannot be neglected when the impact velocity is small, as occurs when a body comes to rest in a series of repeated impacts. Alternatively, as described in the Introduction, the value of z'_{cr} increases with increasing W or decreasing k (equivalently increasing compliance). Finally, as seen in Fig. 8, for bodies with low values of e_∞, the value of ν_{cr} increases so that z'_{cr} increases as well.

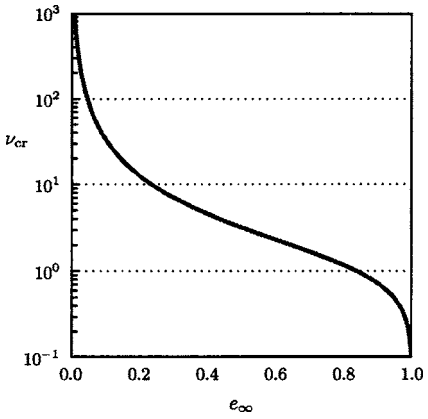


Fig. 8 Critical ν versus e_∞

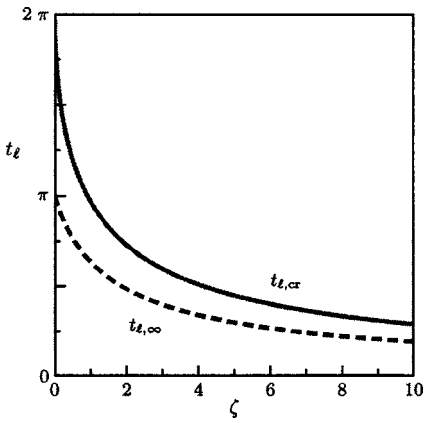


Fig. 9 Duration of the collision t_ℓ

5 Nonlinear Contact Model

The above ideas can be extended to other contact models. Hunt and Crossley [8] develop a nonlinear incremental contact law based on Hertzian contact and experimentally observed scalings. This model can be nondimensionalized and placed in the form (see the Appendix, Sec. II)

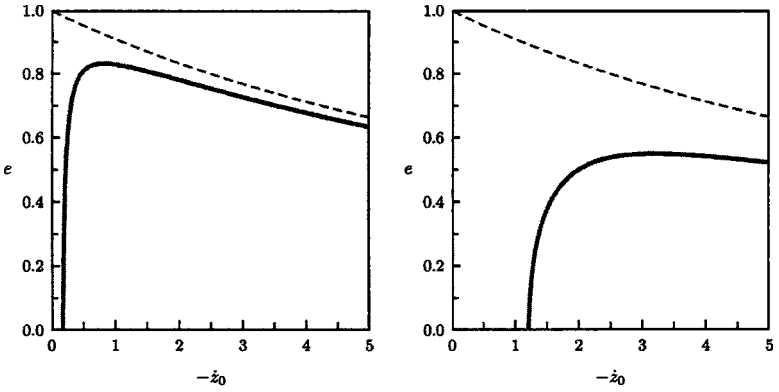
$$\ddot{z} + \sqrt{|z|}\left(\frac{3}{2}\delta|z|\dot{z} + z\right) = -F_0, \tag{7}$$

subject to the initial conditions

$$z(t_i) = 0, \quad \dot{z}(t_i) = -\dot{z}_0.$$

The quantity F_0 has been added to represent an external force applied to the body. For this nonlinear contact model, an *a posteriori* kinematic coefficient of restitution can be defined as in Eq. (3). However, unlike the previous analysis for the viscoelastic contact law, closed form solutions to Eq. (7) do not exist in terms of elementary functions. This equation must be instead solved numerically to determine the post-collisional velocity and the time duration of the impact.

For this nonlinear contact model, the collision terminates when the contact force vanishes and the resulting identified coefficient of restitution is shown in Fig. 10 as the initial impact velocity is varied. In the absence of an external load ($F_0=0$), the coefficient of restitution decreases with increasing impact velocity (shown as the dashed line in each panel). However, for $F_0>0$, there exists an interval of sufficiently small impact velocities for which the resulting kinematic coefficient of restitution is identically zero. For small impact velocities e vanishes, and then rises quickly beyond some critical velocity. As the impact velocity increases further, the equivalent coefficient of restitution decreases, mimicking the results for $F_0=0$. The behavior of the collision subject to this nonlinear contact law is qualitatively similar to the behavior predicted with the linear viscoelastic contact law.



a. $F_0 = 0.25$ b. $F_0 = 1.00$

Fig. 10 Equivalent Newtonian restitution for the nonlinear contact model of Hunt and Crossley (δ=0.10). The measured coefficient for F_0=0 is shown for reference as the dashed line.

6 Modified Algebraic Model

The dynamics associated with repeated impacts has been shown to lead to "chatter" and is difficult to numerically simulate because the time duration between impacts becomes vanishingly small as a specimen comes to rest. From the model considered within this work, if the quantity $\dot{z}/F_0$ is large, the coefficient of restitution identified by the above model is well approximated by Newtonian restitution. However, based on the motivating experiment (see Fig. 3) as well as the analytical results illustrated in Figs. 6 and 10, the present study implies that for low impact velocities, one may introduce an appropriately chosen cutoff in the impact velocity, below which the coefficient of restitution vanishes.

Based on the linear viscoelastic model, given the coefficient of restitution one can estimate the equivalent linear damping coefficient ζ (see Fig. 5). From the results in Fig. 7, one can then calculate the value of ν at which the restitution vanishes. Therefore, the dimensional cut-off velocity corresponding to nonzero restitution scales as

$$z'_{\mathrm{cr}} = \frac{W}{\sqrt{m}}\frac{\nu_{\mathrm{cr}}}{\sqrt{k}} = \left(\frac{W}{m}\right)\frac{\nu_{\mathrm{cr}}}{\sqrt{\frac{k}{m}}}.$$

This critical velocity can be used identify a second parameter σ, related to a characteristic time scale of the collision as

$$\sigma = \frac{\nu_{\mathrm{cr}}}{\sqrt{\frac{k}{m}}}.$$

More generally, one can define the dimensional quantity σ as a second material parameter independent of any specific contact law (analogous to the identification of the coefficient of restitution as a material parameter). Thus a suitable algebraic collision model may take the form

$$e = \begin{cases} e^*, & z' \geq \left(\frac{W}{m}\right)\sigma, \\ 0, & 0 \leq z' < \left(\frac{W}{m}\right)\sigma, \end{cases} \tag{8}$$

where e^* is the coefficient of restitution which would exist in the absence of external impulses. Such a model introduces a cut-off velocity which depends on the colliding bodies (through the quantity σ) as well as the external forces W.

7 Discussion and Conclusions

The modified algebraic law can be evaluated against the experimental data described in Sec. 2. The varying inclination of the contact plane implies that the external force is $W=mg\cos\phi$, so that the dimensional cut-off velocity from Eq. (8) becomes

$$z' = (g\sigma)\cos\phi. \tag{9}$$

In Fig. 11 the experimentally measured cut-off velocity is shown versus ϕ. Superimposed upon the data is the predicted velocity with $\sigma=7.83\times10^{-3}$ s, as determined by a least-squares fit to the data. The agreement between the experimental data and the predicted dependence on ϕ is excellent.

A simple one-and-a-half degrees-of-freedom system has been considered that serves as an incremental model of collisions, subject to appropriate constitutive assumptions for the contact force. For a viscoelastic contact model, the contact dynamics can be solved in closed form to relate the state of the dynamical system before and after the collision. Further, an equivalent algebraic collision law is developed from this model. This reduction removes the finite-time aspects of the collision process from the model at the cost of an additional parameter that accounts for the effect of external forces. These effects are important when the impulse arising from external forces is non-negligible when compared to that of the impulse arising from the contact force. In addition, a nonlinear contact model due to Hunt and Crossley, based on Hertzian contact, was also used and showed qualitatively similar behavior for sufficiently low impact velocities.

The model due to Hunt and Crossley is in part based on work by Goldsmith [14], who studied the effect of impact velocities on

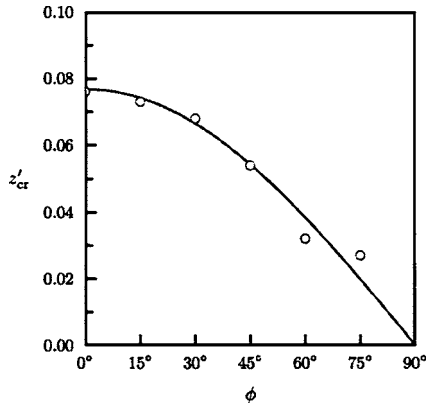


Fig. 11 Experimental critical cutoff velocity. The solid curve is Eq. (9) with $\sigma=7.83\times10^{-3}$ s.

the coefficient of restitution. He found that experimentally the coefficient of restitution increases as the contact velocity approaches zero, possibly due to decreasing plastic deformation in the contact zone. These results stand in apparent contrast to this work. However, the experimental configuration described by Goldsmith eliminates the external compressive force, which is necessary for the behavior described in the current work.

This work illustrates the dynamics of collisions which include the role of non-negligible impulses arising from external forces. In particular, as the impact velocity between colliding bodies decreases, the external impulses give rise to a cut-off velocity below which the equivalent coefficient of restitution vanishes. Above the critical velocity, the equivalent coefficient of restitution rises rapidly to the value in the absence of external forces. This qualitative behavior was exhibited in a motivating experimental study and then shown to exist to two analytical systems, one based on a linear viscoelastic contact model and the second based on a nonlinear contact model develop by Hunt and Crossley. Finally based on the analytical results, these finite-duration effects can be incorporated into existing algebraic models of restitution through the introduction of a cut-off velocity which scales with the external forces and also depends on a material parameter related to the contact stiffness.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. CMS–0084162, and a Firestone Research Grant through the University of Akron. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank Abdelghani Zniber and David Koci for their assistance in obtaining the experimental results. In addition, Dr. Anindya Chatterjee provided many valuable comments on a early draft and suggested the experimental demonstration.

Appendix: Nondimensionalization

Viscoelastic Contact Model. The equations of motion are nondimensionalized through the following transformation to time

$$t = \sqrt{\frac{k}{m}}\tau,$$

while the applied load and contact force are scaled as

$$\frac{W(\tau)}{k} = F(t), \quad \frac{\tilde{N}}{k} = N.$$

Finally, defining the nondimensional damping ratio

$$2\zeta = \frac{b}{\sqrt{km}},$$

yields Eqs. (2). The nondimensional time scale is associated with the relative motion across the compliance. With this scaling, the nondimensional natural frequency for the mass is unity when in contact with the surface. In the absence of external forces, this model, in similar forms, has been considered previously [6,15,16].

Nonlinear Contact Model. The incremental model proposed by Hunt and Crossley [8] can be written as

$$mz'' + k\sqrt{|z|}\left(\frac{3}{2}\alpha|z|z' + z\right) = -W(\tau),$$

where W_0 has been added to represent a constant external force applied to the body. The stiffness k arises from a Hertzian contact assumption, while the nonlinear damping, proportional to α is chosen to reflect the experimentally observed linear decrease in coefficient of restitution with increasing impact velocity. Note that k has units of force/length$^{3/2}$ while α has units of velocity^{-1}.

To nondimensionalize, scale time by ω and the displacement by a characteristic length ξ, to yield

$$\ddot{z} + \sqrt{|z|}\left(\frac{3}{2}\delta|z|\dot{z} + z\right) = -F_0,$$

where $\omega^2 = m/(k\sqrt{\xi})$, and

$$\delta = \alpha\sqrt{\frac{k}{m}}\xi^{5/4}, \quad F_0 = \frac{W_0}{k\xi^{3/2}}.$$

Finally, the velocity scales as

$$z' = \frac{\xi}{\omega}\dot{z} = \left(\sqrt{\frac{k}{m}}\xi^{5/4}\right)\dot{z}.$$

The characteristic length ξ is choosen to be the maximum compression during a collision which arises in the system with zero damping ($\alpha=0$) and impact velocity $z'(0)=z_0'$, found to be

$$\xi = \left(\sqrt{\frac{m}{k}}\frac{\sqrt{5}}{2}z_0'\right)^{4/5}.$$

Since α has units of velocity^{-1}, z_0' is choosen to have magnitude $2/\sqrt{5}$, with the same units as α^{-1}. Therefore the nondimensional values of δ and the velocity have the same magnitude as their dimensional counterparts, while the external forcing can be written as

$$F_0 = \frac{W_0}{m^{3/5}k^{2/5}}.$$

References

[1] Brach, R. M., 1989, "Rigid Body Collisions," ASME J. Appl. Mech., **56**, pp. 133–138.
[2] Wang, Y., and Mason, M. T., 1992, "Two-dimensional Rigid-Body Collisions with Friction," ASME J. Appl. Mech., **59**, pp. 635–642.
[3] Stronge, W. J., 1990, "Rigid Body Collisions With Friction," Proc. R. Soc. London, Ser. A, **A431**, pp. 169–181.
[4] Tatara, Y., 1977, "Effects of External Force on Contacting Times and Coefficient of Restitution in a Periodic Collision," ASME J. Appl. Mech., **44**(4), pp. 773–774.
[5] Stronge, W. J., 2000, *Impact Mechanics*, Cambridge University Press, Cambridge, UK.
[6] Butcher, E. A., and Segalman, D. J., 1999, "Characterizing Damping and Restitution in Compliant Impacts via Linear Viscoelastic Models," 1999 ASME Design Engineering Technical Conferences, Las Vegas, Nevada, September 12—15, DETC99/VIB-8335, ASME, New York.
[7] Fandrich, M., 1998, "Modelling Nondestructive Impacts Macroscopically," Math. Comput. Modell., **28**(4–8), pp. 205–224.
[8] Hunt, K. H., and Crossley, F. R. E., 1975, "Coefficient of Restitution Interpretated as Damping in Vibroimpact," ASME J. Appl. Mech. , pp. 440–445.
[9] Brogliato, B., 1999, *Nonsmooth Mechanics: Models, Dynamics and Control*, 2nd ed., Springer-Verlag, London.
[10] Hurmuzlu, Y., and Ceanga, V., 2000, "Impulse Correlation Ratio in Solving Multiple Impact Problems," *Impacts in Mechanical Systems*, B. Brogliato, ed., (Vol. 551 of Lecture Notes in Physics), Springer-Verlag, Berlin, pp. 235-273.
[11] Holmes, P. J., and Marsden, J., 1982, "Melnikov's Method and Arnold Diffusion for Perturbations of Integrable Hamiltonian Systems," J. Math. Phys., **23**, pp. 669–675.
[12] Koditschek, D. E., and Bühler, M., 1991, "Analysis of a Simplified Hopping Robot," Int. J. Robot. Res., **10**(6), pp. 587–605.
[13] Shaw, J., and Shaw, S. W., 1989, "The Onset of Chaos in a Two-degree-of-freedom Impacting System," ASME J. Appl. Mech., **56**, pp. 168–174.
[14] Goldsmith, W., 1960, *Impact: The Theory and Physical Behaviour of Colliding Solids*, Edward Arnold Ltd., London.
[15] Brach, R. M., 1991, *Mechanical Impact Dynamics: Rigid Body Collisions*, Wiley, New York.
[16] Chatterjee, A., 1997, "Rigid Body Collisions: Some General Considerations, New Collisison Laws, and Some Experimental Data," Ph.D. thesis, Cornell University, Ithaca, New York.

Which Formulation Allows Using a Constant Shear Modulus for Small-Strain Buckling of Soft-Core Sandwich Structures?

Zdeněk P. Bažant
McCormick School Professor and W. P. Murphy Professor of Civil Engineering and Materials Science
Fellow ASME
Northwestern University, 2145 Sheridan Road, CEE, Evanston, IL 60208
e-mail: z-bazant@northwestern.edu

Alessandro Beghini
Graduate Research Assistant
Northwestern University, Evanston, IL 60208
e-mail: a-beghini@northwestern.edu

Although the stability theories energetically associated with different finite strain measures are mutually equivalent if the tangential moduli are properly transformed as a function of stress, only one theory can allow the use of a constant shear modulus G if the strains are small and the material deforms in the linear elastic range. Recently it was shown that, in the case of heterogeneous orthotropic structures very soft in shear, the choice of theory to use is related to the problem of proper homogenization and depends on the type of structure. An example is the difference between Engesser's and Haringx's formulas for critical load of columns with shear, which were shown to be energetically associated with Green's and Almansi's Lagrangian finite strain tensors. In a previous brief paper of the authors in a conference special issue, it was concluded on the basis of energy arguments that, for constant G, Engesser's formula is correct for sandwich columns and Haringx's formula for elastomeric bearings, but no supporting experimental results were presented. To present them, is the main purpose of this technical brief. [DOI: 10.1115/1.1979516]

Introduction

The question of a correct formula for critical load, P_{cr}, of elastic structures such as sandwich columns, composite columns, lattice columns, helical springs, and elastomeric bearings, in which shear deformations dominate, has been subject of polemics for decades; see [[1], and references therein]. The best known examples are the formulas of Engesser and Haringx. The polemics were settled in 1971 [2] by the demonstration that these two formulas are equivalent if the shear modulus G is properly transformed as a function of the axial stress; see also [3]. This showed that, if the material behavior is nonlinear and the stress-strain curve is defined in terms of the finite strain tensor as that used to interpret test results, it does not matter which formula is chosen because the difference is only in the form of stress dependence of tangent shear modulus G. This conclusion, however, leads to apparently paradoxical properties when the strains are so small that the material is in the linear range of response.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, April 6, 2004; final revision, December 30, 2004. Associate Editor: N. Triantafyllides.

In [3], it was shown that such a situation may arise for built-up (lattice) columns with very weak shear bracing, and that only the Engesser formula is correct for such columns. In [1], it was shown that such a situation also arises for sandwich columns very soft in shear.

For such columns, the equivalence of Engesser's and Haringx's formulas requires that the tangent shear modulus G cannot be constant in both of these two formulas even if the strains are so small that the core is in the linear range of response.

Because all of the axial load in a sandwich is carried by the skins and all of the shear force is resisted by the core, the stress dependence of G of the core means that the properties of the core depend on the axial normal stress in the skins (or facings). This may seem to be paradoxical. But the variational energy analysis in [1] showed that this is simply a consequence of the fact that the cross sections of the core may be assumed to remain plane.

In [4], variational energy analysis of the same kind as in [1] was used to prove that, if all the strains at critical load, P_{cr}, are small, and if a constant G (as measured in small-strain torsion) is to be used, then, for sandwich beams, only Engesser's formula is correct, and, for elastomeric bearings, only Haringx's formula is correct. However, experimental results supporting this conclusion have not been presented. To present them is the main objective of this Technical Note. A secondary objective is to briefly outline an extension of this analysis to arbitrary soft-in-shear structures under multiaxial stress, such as layered or highly orthotropic bodies. In full detail, this extension is found in [5].

Background on Variational Analysis of Homogeneous Orthotropic Columns With Shear

The variational energy analysis presented in [2] showed that the differences between various stability theories for buckling with shear arise from different choices of the finite strain measure. All the finite strain tensors used in stability theories so far belong to the class of Doyle-Ericksen finite strain tensors $\boldsymbol{\epsilon}^{(m)}=(\boldsymbol{U}^m-\boldsymbol{I})/m$ where m=real parameter, $\boldsymbol{I}$=unit tensor, and $\boldsymbol{U}$=right-stretch tensor. In particular, $m=2$ gives Green's Lagrangian strain tensor, and $m=-2$ Almansi's Lagrangian strain tensor. For calculating the critical load, only the second-order approximation of these tensors matters; it reads [2]:

$$\epsilon_{ij}^{(m)} = e_{ij} + \tfrac{1}{2}u_{k,i}u_{k,j} - \alpha e_{ki}e_{kj}, \quad e_{ij} = \tfrac{1}{2}(u_{i,j} + u_{j,i}), \quad \alpha = 1 - \tfrac{1}{2}m \tag{1}$$

where e_{ij}=small-strain tensor (linearized), u_i=displacement components, and the subscripts refer to Cartesian coordinates x_i ($i=1,2,3$). The finite strain tensors $\boldsymbol{\epsilon}^{(n)}=(\boldsymbol{U}^n-\boldsymbol{U}^{-n})/2n$ proposed in [6] do not fit the class of Doyle-Ericksen tensors but the second-order approximation of $\boldsymbol{\epsilon}^{(n)}$ for any n coincides with (1) for $m \to 0$ (Hencky strain tensor). The stability criteria obtained from any of these strain measures have been shown in [2] to be mutually equivalent if the tangential moduli $C_{ijkl}^{(m)}$ associated with different m-values satisfy the relation:

$$C_{ijkl}^{(m)} = C_{ijkl}^{(2)} + \frac{2-m}{4}(S_{ik}\delta_{jl} + S_{jk}\delta_{il} + S_{il}\delta_{jk} + S_{jl}\delta_{ik}) \tag{2}$$

(see also [3]) where $C_{ijkl}^{(2)}$=components of tangential moduli tensor $\boldsymbol{C}^{(m)}$ associated with Green's Lagrangian strain ($m=2$) and S_{ij}=components of current stress tensor $\boldsymbol{S}$ (Cauchy stress).

According to the standard simplifying hypothesis that the cross sections of the core remain plane, which makes the problem one-dimensional, the second-order accurate expression for the incremental potential energy of a column that is initially in equilibrium in an undeflected position is

 Copyright © 2005 by ASME

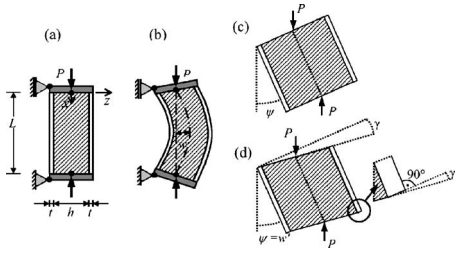


Fig. 1 Column in (*a*) initial state, (*b*) deflected state; contribution to the shear deformation of a sandwich beam element: (*c*) bending, and (*d*) shear

$$\delta^2\mathcal{W}=\int_0^L\int_A\left[S^0(y,z)(\epsilon_{11}^{(m)}-e_{11})+\tfrac{1}{2}E^{(m)}(y,z)e_{11}^2+\tfrac{1}{2}G^{(m)}(y,z)\gamma^2\right]\mathrm{d}A\,\mathrm{d}x \tag{3}$$

where A=cross-section area; $S^0(y,z)=-P/A$=initial axial normal stress; P=axial compressive load; $E^{(m)}$, $G^{(m)}$=tangent elastic moduli in the axial direction and shear, associated with finite strain tensor $\epsilon^{(m)}$; $\gamma=u_{1,3}+u_{3,1}$=shear angle; $\psi(x)$=rotation angle of cross section; $x_1=x$=axial coordinate, $x_3=z$=transverse coordinate in the plane of buckling, $u_1=u(x)$=axial displacement (in x-direction), and $u_3=w(x)$=transverse deflection (see Figs. 1(*a*) and 1(*b*)). Applying the Trefftz condition of critical state to Eq. (3), one can derive the critical load formulas associated with finite strain tensor of any parameter m [1]. In particular for $m=2$ and $m=-2$, one obtains the well-known formulas:

$$\text{for } m=2: \quad P_{\mathrm{cr}}=\frac{P_E}{1+(P_E/G^{(2)}A)} \quad \text{(Engesser)} \tag{4}$$

$$\text{for } m=-2: \quad P_{\mathrm{cr}}=\frac{G^{(-2)}A}{2}\left[\sqrt{1+\frac{4P_E}{G^{(-2)}A}}-1\right] \quad \text{(Haringx)} \tag{5}$$

where $P_E=\pi^2EI/l^2$=Euler's critical load of the column. It can be checked that (5) is obtained from (4) if the following substitution (a special case of (2) [2,3]) is made:

$$G^{(2)}=G^{(-2)}+P/A \tag{6}$$

Correct Critical Load Formula for Sandwich Columns and Experimental Evidence

If the tangent modulus G in the presence of normal stress happens to be constant with respect to Green's Lagrangian strain measure (corresponding to Engesser's theory, $m=2$), it cannot be constant with respect to the Almansi's Lagrangian strain measure (corresponding to Haringx's theory, $m=-2$), and vice versa. Thus the main question is how to use the constant shear modulus G that is measured by small-strain shear tests, for example by torsion of a hollow circular tube. Should this constant modulus be used in Engesser's formula, or in Haringx's formula, or in a formula for some other m-value? The correct answer for a sandwich is Engesser's formula. This answer was not reached in [1] because the physical interpretation of the variational analysis for a sandwich was incomplete. The correct interpretation, leading to the correct conclusion, was presented in [4] (and is also reviewed in [5]). It may be summarized as follows.

We consider a sandwich beam element to undergo a rigid-body rotation through a small angle w' followed by homogeneous pure shear deformation γ, as shown in Figs. 1(*c*) and 1(*d*). Using the notation shown in Figs. 1(*a*) and 1(*b*), we can describe the displacement and deformation fields as follows:

$$u_1=u_{1,1}=u_{1,3}=e_{11}=0, \quad u_{3,1}=\gamma, \quad e_{13}=e_{31}=\gamma/2 \tag{7}$$

According to the assumption of negligible skin thickness, the shear deformation within each skin vanishes (as shown in the zoomed region in Fig. 1(*d*)), although the skins as a pair, of course, do exhibit shear deformation (Fig. 1(*d*)). First we calculate the work as a special case of Eq. (3) for a sandwich column. After substitution of the homogeneous strain field (7) into (3), the flexural terms vanish and one obtains (for a cross section of width b in the y direction):

$$\delta^2\mathcal{W}=\left(bhG+\frac{2-m}{4}P\right)\frac{\gamma^2}{2}-\frac{Pw'^2}{2} \tag{8}$$

Superscript m is here omitted because the core is in small strain, in which case the shear modulus G is independent of the specific choice of strain measure.

Second, we figure out the work on the sandwich beam element by a direct elementary reasoning. During the rotation of the beam element (Fig. 1(*c*)), the second-order approximation to the vertical shortening is $w'^2/2$, and the work of the axial force due to this shortening is $-Pw'^2/2$ (note that the work of loads must be taken as negative because the potential energy of internal stresses is taken as positive). During the subsequent shear deformation of the core (Fig. 1(*d*)), the work of shear stresses obviously is $bhG\gamma^2/2$. So the total work is

$$\delta^2\mathcal{W}=bhG\frac{\gamma^2}{2}-\frac{Pw'^2}{2} \tag{9}$$

Expressions (8) and (9) must coincide. Their comparison immediately reveals that this happens if and only if $m=2$. So, this is the value of m that allows the use of a constant shear modulus in the small-strain analysis of a sandwich. For this, and only this, value of m, the variational result in (8) coincides with the second-order work obtained by the elementary reasoning based on (9). Therefore, it must be concluded that whenever the strains are small and the shear modulus is assumed to be constant, sandwich structures must be analyzed on the basis of Engesser-type theory.

This result does not come as a surprise. Already in 1966, Plantema [7] justified Engesser's theory on the basis of several experimental data. The Engesser-type theory has also been adopted by Zenkert [8].

The most extensive support for Engesser's theory is provided by the recent tests of Fleck and Sridhar [9] shown in Figs. 2(*a*)–2(*c*). These tests also validate the correctness of the energetic argument in [5] summarized here. In these tests, the sandwich columns were fixed at both ends. The skins were made of the same material and had the same width. The cores had the same in-plane thickness and were made of Divinycell PVC foam. Three different densities of the foam, with designations H30, H100, and H200, and several column lengths L, giving different slenderness $l/(h+2t)$, were used. The diagrams in Fig. 2 are plotted in logarithmic scales and dimensionless coordinates, as P_{cr}/P_E versus $l/(h+2t)$. The agreement with Engesser's formula is good (except for one anomalous point), while the formula of Haringx is seen to deviate significantly from the test results if a constant G is used.

Correct Critical Load Formula for Homogenized Elastomeric Bearings and General Layered or Orthotropic Structures

For elastomeric bearings and helical springs, discussed in detail in [5], a large number of studies and all experiments clearly favor Haringx-type theory. The mathematical argument is analogous to that pursued for sandwich columns. If one calculates the work as a special case of Eq. (3) for a pure shear deformation and compares it to the expression obtained by direct elementary reasoning (based on the sketch in Fig. 3), one concludes that the only m value for which the two expressions are the same is $m=-2$, which corresponds to Haringx's formula [5].

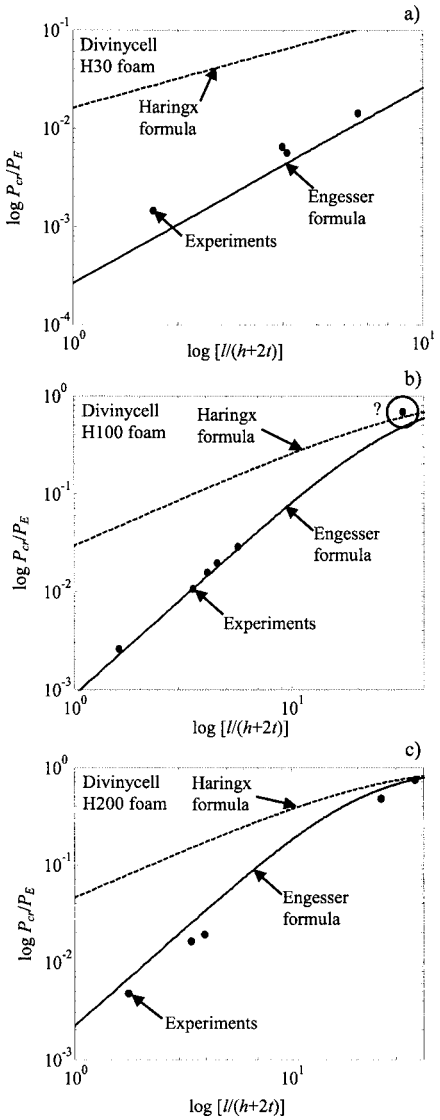


Fig. 2 Comparison of buckling formulas and experimental results for various density of Divinycell foam

The same type of argument may be applied to general homogenized soft-in-shear structures, e.g., layered bodies and unidirectional fiber reinforced composites. The Engesser-type or Haringx-type theory is found to apply if the compressive force is applied in the direction parallel or normal to the layers or fibers, respectively.

For the general case of a soft-in-shear orthotropic structure under biaxial stress, it is found [5] that a theory corresponding to a general m-value must be used. It has been shown [5] that m is given by the following equation:

$$m = \frac{2(S_3^0/S_1^0) - 2}{(S_3^0/S_1^0) + 1} \tag{10}$$

where S_1^0 and S_3^0 are the initial normal stresses in the directions of orthotropy.

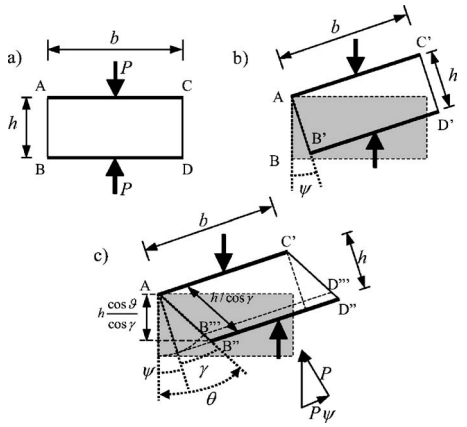


Fig. 3 Contribution to the shear deformation of an elastomeric bearing element: (*a*) initial state, (*b*) bending, and (*c*) shear

To permit applying finite element discretization to homogenized soft-in-shear structures of this kind, it would be useful to generalize the current commercial finite element programs for an arbitrary value of $m \neq 2$ [4,5]. At present, all of them use an updated Lagrangian algorithm energetically associated with Green's Lagragian finite strain tensor, i.e., with $m=2$, and thus, in general, cannot give correct results for homogenized soft-in-shear structures [5].

Conclusion

The recent conclusion that for soft-in-shear sandwich columns in small-strain and in the linear elastic range the critical load is predicted correctly by Engesser's (but not Haringx's) formula with a constant shear modulus of core is shown to be in good agreement with recent extensive test results.

Acknowledgment

Financial support under Grant No. N00014-02-I-0622 from the Office of Naval Research to Northwestern University (monitored by Dr. Yapa D. S. Rajapakse) is gratefully acknowledged.

References

[1] Bažant, Z. P., 2003, "Shear Buckling of Sandwich, Fiber-Composite and Lattice Columns, Bearings and Helical Springs: Paradox Resolved," ASME J. Appl. Mech., **70**, pp. 75–83.
[2] Bažant, Z. P., 1971, "A Correlation Study of Incremental Deformations and Stability of Continuous Bodies," ASME J. Appl. Mech., **38**, pp. 919–928.
[3] Bažant, Z. P., and Cedolin, L., 1991, *Stability of Structures: Elastic, Inelastic, Fracture and Damage Theories*, Oxford University Press, New York; and republication with updates, Dover, New York, 2003.
[4] Bažant, Z. P., and Beghini, A., 2004, "Sandwich Buckling Formulas and Applicability of Standard Computational Algorithm for Finite Strain," Composites, Part B, **35**, pp. 573–581.
[5] Bažant, Z. P., and Beghini, A., 2004, "Stability and Finite Strain of Homogenized Structures Soft in Shear: Sandwich or Fiber Composites, and Layered Bodies," Theoretical and Applied Mechanics Report No. 04-03/C441s, Northwestern University, Evanston, IL; also Int. J. Solids Struct., (in press).
[6] Bažant, Z. P., 1998, "Easy-to-Compute Tensors with Symmetric Inverse Approximating Hencky Finite Strain and its Rate," ASME J. Eng. Mater. Technol., **120**, pp. 131–136.
[7] Plantema, F. J., 1966, *Sandwich Construction: The Bending and Buckling of Sandwich Beams, Plates and Shells*, Wiley, New York, p. 9.
[8] Zenkert, D., 1997, *The Handbook of Sandwich Construction*, EMAS, UK.
[9] Fleck, N. A., and Sridhar, I., 2002, "End Compression of Sandwich Column," Composites, Part A, **33**, pp. 353–359.

Three-Dimensional Flexure of Rectangular Plates Made of Functionally Graded Materials

Isaac Elishakoff
Department of Mechanical Engineering, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431-0091
e-mail: elishako@fau.edu

Cristina Gentilini
Dipartimento di Ingegneria delle Strutture, dei Trasporti, delle Acque, del Rilevamento, del Territorio, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy
e-mail: cristina.gentilini@mail.ing.unibo.it

A three-dimensional solution for the problem of transversely loaded, all-round clamped rectangular plates of arbitrary thickness is presented within the linear, small deformation theory of elasticity. The Ritz minimum energy principle is employed to derive the governing equation of the plate made of functionally graded materials. In theory, if we employ an infinite number of terms in the displacement series, the exact solution can be determined. However, a practical limit always exists due to numerical implementation. The solution has a validity comparable to some higher order theories. A power-law distribution for the mechanical characteristics is adopted to model the continuous variation of properties from those of one component to those of the other. The displacements and stresses of the plate for different values of the power-law exponent are investigated. [DOI: 10.1115/1.1985429]

1 Introduction

The concept of functionally graded materials (FGMs) has been proposed in the beginning of the 90's by Japanese researchers [1,2]. The idea came from the awareness that the abrupt transition in material composition and properties may result in sharp local concentrations of stress. In FGMs, the microstructure changes continuously, so that material properties in the thickness direction vary from, say, E_c, to the value E_m, where E_c and E_m are the moduli of elasticity of two different materials. In the last five years researchers have shown a huge interest in the statics, dynamics, and stability of structures made of FGMs (see, for example, Suresh and Mortensen [3] and Reddy [4]). Several investigators devoted themselves to the study of simply supported FG plates by means of different theories. Della Croce and Venini [5] developed a hierarchic family of finite elements according to the Reissner-Mindlin theory, while Reddy [6] and Yang and Shen [7] dealt with higher-order shear deformation theories. Three-dimensional analyses were carried out, among the others, by Reddy and Cheng [8] and Vel and Batra [9]. Studies about clamped FG plates have been proposed by Cheng and Batra [10] for thermoelastic elliptic plates and by Qian and Batra [11] for rectangular plates using a higher order shear and normal deformable plate theory together with a meshless local Petrov-Galerkin method. Since FG plates may have substantial thickness, previous studies [5–11] based upon higher order theories are extremely useful in providing realistic results.

Herein we employ the Ritz method for the three-dimensional static analysis of all-round clamped FG plates subjected to a uniformly distributed normal load on the top surface. In this study an aluminum/silicon carbide graded plate with a power-law variation of the volume fraction of the constituents is considered. Both Young's modulus and Poisson's ratio are assumed to vary in the thickness direction. According to the Ritz method, the spatial displacement components in the three coordinate directions are represented by sets of one-dimensional Chebyshev polynomials multiplied by corresponding boundary functions, such that the essential geometric conditions are identically satisfied at the outset. The effects of variation of the volume fraction of the constituent materials on the through the thickness deflections, in-plane displacements and axial stress distributions are studied.

2 Formulation of the Problem

A FG rectangular plate of length a, width b, and uniform thickness h is considered. The volume and the top surface are indicated with V and Ω, respectively. The origin of the Cartesian coordinate system (x,y,z) is the center of the plate and the axes are parallel to the edges of the plate. It is assumed that the FGM is made of a mixture of a ceramic and a metallic component. Young's modulus E and Poisson's ratio ν vary across the thickness according to the following equations:

$$E(z) = (E_c - E_m)V_f + E_m, \quad \nu(z) = (\nu_c - \nu_m)V_f + \nu_m,$$

$$V_f = (z/h + 1/2)^N, \tag{1}$$

where V_f is the ceramic volume fraction and N is the volume fraction exponent which takes positive real values. The value of N equal to zero represents a fully ceramic plate, when N is approaching the case of the fully metallic plate is obtained. According to this distribution, the bottom surface, $z=-h/2$, of the functionally graded plate is pure metal and the top surface, $z=h/2$, is pure ceramic. Subscripts c and m stand for the corresponding properties of the pure ceramic and pure metal. The above power-law assumption reflects a simple rule of mixture used to obtain the effective properties of the FG plate [6].

3 A Three-Dimensional Analysis

The generic configuration of the plate is described by the displacement components $u(x,y,z)$, $v(x,y,z)$, and $w(x,y,z)$ in the x, y, and z directions, respectively. In the three-dimensional setting, the stress-displacement relations are:

$$\sigma_{xx} = \lambda(u_{,x} + v_{,y} + w_{,z}) + 2\mu(u_{,x}), \tag{2}$$

$$\sigma_{yy} = \lambda(u_{,x} + v_{,y} + w_{,z}) + 2\mu(v_{,y}), \tag{3}$$

$$\sigma_{zz} = \lambda(u_{,x} + v_{,y} + w_{,z}) + 2\mu(w_{,z}), \tag{4}$$

$$\sigma_{xy} = \mu(u_{,y} + v_{,x}), \tag{5}$$

$$\sigma_{xz} = \mu(u_{,z} + w_{,x}), \tag{6}$$

$$\sigma_{yz} = \mu(v_{,z} + w_{,y}), \tag{7}$$

where λ and μ are the Lamè coefficients:

$$\lambda(z) = \nu E/[(1+\nu)(1-2\nu)], \quad \mu(z) = E/[2(1+\nu)], \tag{8}$$

functions of z and a comma followed by a variable denotes differentiation with respect to that variable. The strain energy of a three-dimensional solid can be written as:

Contributed by the Applied Mechanics Division of The American Society of Mechanical Engineers for publication in the ASME Journal of Applied Mechanics. Manuscript received by the ASME Applied Mechanics Division, May 31, 2004; final revision, January 18, 2005. Associate Editor: S. Mukherjee.

 Copyright © 2005 by ASME

$$\Phi = \frac{1}{2}\int_V \{(\lambda+2\mu)[(u_{,x})^2+(v_{,y})^2+(w_{,z})^2] + 2\lambda(u_{,x}v_{,y}+u_{,x}w_{,z} + v_{,y}w_{,z}) + \mu[(u_{,y}+v_{,x})^2+(u_{,z}+w_{,x})^2+(v_{,z}+w_{,y})^2]\}\mathrm{d}V. \tag{9}$$

The work done by a transverse loading $q(x,y)$ acting over the surface Ω of the plate is given by:

$$W = \int_\Omega q(x,y)w(x,y,h/2)\mathrm{d}\Omega. \tag{10}$$

For simplicity, the (ξ,η,ζ) nondimensional coordinate system is introduced, where:

$$\xi = 2x/a, \quad \eta = 2y/b, \quad \zeta = 2z/h. \tag{11}$$

To apply the Ritz method, the displacement components in the nondimensional coordinate system are approximated by expressions of the type:

$$u(\xi,\eta,\zeta) = F_u(\xi,\eta)\sum_i^I\sum_j^J\sum_k^K C^u_{ijk}P_i(\xi)P_j(\eta)P_k(\zeta), \tag{12}$$

$$v(\xi,\eta,\zeta) = F_v(\xi,\eta)\sum_\ell^L\sum_m^M\sum_n^N C^v_{\ell mn}P_l(\xi)P_m(\eta)P_n(\zeta), \tag{13}$$

$$w(\xi,\eta,\zeta) = F_w(\xi,\eta)\sum_p^P\sum_q^Q\sum_r^R C^w_{pqr}P_p(\xi)P_q(\eta)P_r(\zeta), \tag{14}$$

where C^u_{ijk}, $C^v_{\ell mn}$, and C^w_{pqr} denote undetermined components, F_u, F_v, and F_w are appropriate boundary functions, while $P_s(\chi)$ is the so-called coordinate function, with s representing the serial index and $\chi=\xi,\ \eta,\ \zeta$. In this work, Chebyshev polynomials are considered as coordinate functions. Thus, in Eqs. (12)–(14), $P_s(\chi)$ is the one-dimensional sth Chebyshev polynomial which can be written via a recurrence relation [12]:

$$P_{s+1}(\chi) = 2\chi P_s(\chi) - P_{s-1}(\chi) \text{ with } s=2,3,\ldots \text{ and } \chi=\xi,\eta,\zeta. \tag{15}$$

For $s=1$, $P_1(\chi)$ is equal to 1. It is known that Chebyshev polynomials constitute an orthogonal polynomial sequence with respect to the weighting function $\varpi(\chi)=(1-\chi^2)^{-1/2}$ on the interval $[-1,1]$.

The load $q(\xi,\eta)$ is assumed to be applied downward on the top surface and is represented in the general form of a double series:

$$q(\xi,\eta) = \sum_\alpha\sum_\beta Q_{\alpha\beta}P_\alpha(\xi)P_\beta(\eta), \tag{16}$$

where $P_\alpha(\xi)$ and $P_\beta(\eta)$ are Chebyshev polynomials.

4 Solution Procedure

For equilibrium, the total energy functional Π, defined as the difference between the strain energy Φ, Eq. (9), and the potential energy of the load W, Eq. (10), $\Pi=\Phi-W$, of the system must be a minimum. This is accomplished by minimizing the total energy Π with respect to the components C^u_{ijk}, $C^v_{\ell mn}$, and C^w_{pqr}:

$$\partial\Pi/\partial C^u_{ijk}=0, \quad \partial\Pi/\partial C^v_{\ell mn}=0, \quad \partial\Pi/\partial C^w_{pqr}=0. \tag{17}$$

From Eq. (17), the following governing equation is obtained:

$$\mathbf{KC}-\mathbf{Q}=\mathbf{0} \tag{18}$$

in which $\mathbf{K}$, $\mathbf{Q}$, and $\mathbf{C}$ are the stiffness matrix, the load vector and the vector of the displacement coefficients, respectively, in the form:

$$\mathbf{K}=\begin{bmatrix}\mathbf{K}_{uu} & \mathbf{K}_{uv} & \mathbf{K}_{uw}\\ & \mathbf{K}_{vv} & \mathbf{K}_{vw}\\ Sym & & \mathbf{K}_{ww}\end{bmatrix}, \quad \mathbf{Q}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{Q}^z\end{bmatrix}, \quad \mathbf{C}=\begin{bmatrix}\mathbf{C}^u\\ \mathbf{C}^v\\ \mathbf{C}^w\end{bmatrix}. \tag{19}$$

The explicit form of the elements in the stiffness sub-matrices $\mathbf{K}_{ij}$ with $i,j=u,v,w$ in Eq. (19), are given by:

$$\mathbf{K}_{uu} = D^{11}_{uiu\bar{i}}E^{00}_{uju\bar{j}}F^{00}_{k\bar{k}} + \left(\frac{a}{b}\right)^2 D^{00}_{uiu\bar{i}}E^{11}_{uju\bar{j}}T^{00}_{k\bar{k}} + \left(\frac{a}{h}\right)^2 D^{00}_{uiu\bar{i}}E^{00}_{uju\bar{j}}T^{11}_{k\bar{k}},$$

$$\mathbf{K}_{uv} = \frac{a}{b}(D^{10}_{uivl}E^{01}_{ujvm}S^{00}_{kn} + D^{01}_{uivl}E^{10}_{ujvm}T^{00}_{kn}),$$

$$\mathbf{K}_{uw} = \frac{a}{h}(D^{10}_{uiwp}E^{00}_{ujwq}S^{01}_{kr} + D^{01}_{uiwp}E^{00}_{ujwq}T^{10}_{kr}),$$

$$\mathbf{K}_{vv} = \left(\frac{a}{b}\right)^2 D^{00}_{vlv\bar{l}}E^{11}_{vmv\bar{m}}F^{00}_{n\bar{n}} + \left(\frac{a}{h}\right)^2 D^{00}_{vlv\bar{l}}E^{00}_{vmv\bar{m}}T^{11}_{n\bar{n}} + D^{11}_{vlv\bar{l}}E^{00}_{vmv\bar{m}}T^{00}_{n\bar{n}},$$

$$\mathbf{K}_{vw} = \frac{a^2}{hb}(D^{00}_{vlwp}E^{10}_{vmwq}S^{01}_{nr} + D^{00}_{vlwp}E^{01}_{wmwq}T^{10}_{nr}),$$

$$\mathbf{K}_{ww} = \left(\frac{a}{h}\right)^2 D^{00}_{wpw\bar{p}}E^{00}_{wqw\bar{q}}F^{11}_{r\bar{r}} + \left(\frac{a}{b}\right)^2 D^{00}_{wpw\bar{p}}E^{11}_{wqw\bar{q}}T^{00}_{r\bar{r}} + D^{11}_{wpw\bar{p}}E^{00}_{wqw\bar{q}}T^{00}_{r\bar{r}}, \tag{20}$$

in which

$$D^{rs}_{\alpha\sigma\beta\theta} = \int_{-1}^{1}\frac{\mathrm{d}^r[f^1_\alpha(\xi)P_\sigma(\xi)]}{\mathrm{d}\xi^r}\frac{\mathrm{d}^s[f^1_\beta(\xi)P_\theta(\xi)]}{\mathrm{d}\xi^s}\mathrm{d}\xi,$$

$$E^{rs}_{\alpha\sigma\beta\theta} = \int_{-1}^{1}\frac{\mathrm{d}^r[f^2_\alpha(\eta)P_\sigma(\eta)]}{\mathrm{d}\eta^r}\frac{\mathrm{d}^s[f^2_\beta(\eta)P_\theta(\eta)]}{\mathrm{d}\eta^s}\mathrm{d}\eta,$$

$$F^{rs}_{\sigma\theta} = \int_{-1}^{1}\frac{E(\zeta)(1-\nu(\zeta))}{(1+\nu(\zeta))(1-2\nu(\zeta))}\frac{\mathrm{d}^rP_\sigma(\zeta)}{\mathrm{d}\zeta^r}\frac{\mathrm{d}^sP_\theta(\zeta)}{\mathrm{d}\zeta^s}\mathrm{d}\zeta,$$

$$S^{rs}_{\sigma\theta} = \int_{-1}^{1}\frac{E(\zeta)\nu(\zeta)}{(1+\nu(\zeta))(1-2\nu(\zeta))}\frac{\mathrm{d}^rP_\sigma(\zeta)}{\mathrm{d}\zeta^r}\frac{\mathrm{d}^sP_\theta(\zeta)}{\mathrm{d}\zeta^s}\mathrm{d}\zeta,$$

$$T^{rs}_{\sigma\theta} = \int_{-1}^{1}\frac{E(\zeta)}{2(1+\nu(\zeta))}\frac{\mathrm{d}^rP_\sigma(\zeta)}{\mathrm{d}\zeta^r}\frac{\mathrm{d}^sP_\theta(\zeta)}{\mathrm{d}\zeta^s}\mathrm{d}\zeta, \tag{21}$$

$$(r,s=0,1;\ \alpha,\beta=u,v,w;\ \sigma=i,j,k,l,m,n,p,q,r;$$

$$\theta=\bar{i},\bar{j},\bar{k},\bar{l},\bar{m},\bar{n},\bar{p},\bar{q},\bar{r})$$

where Young's modulus $E(\zeta)$ and Poisson's ratio $\nu(\zeta)$ have the expressions of Eq. (1). The functions $f^1_\alpha(\xi)$, $f^1_\beta(\xi)$, $f^2_\alpha(\eta)$, and $f^2_\beta(\eta)$ are employed to satisfy boundary conditions. In the Ritz method, the satisfaction of geometric boundary conditions must be ensured. In this study, a fully clamped plate is considered, thus the edge conditions can be specified as:

$$u=0, \quad v=0, \quad w=0 \text{ at } \xi=\pm1, \quad \eta=\pm1. \tag{22}$$

The boundary functions $F_u(\xi,\eta)$, $F_v(\xi,\eta)$, and $F_w(\xi,\eta)$ in Eqs. (12)–(14) ensure that the displacement components satisfy the essential geometric boundary conditions for the plate in a pointwise manner. They can be written in the following way:

$$F_\delta(\xi,\eta) = f^1_\delta(\xi)f^2_\delta(\eta) \text{ with } \delta=u,v,w \tag{23}$$

where, in the case of a clamped plate, $f^1_\delta(\xi)$ and $f^2_\delta(\eta)$ are given by $1-\xi^2$ and $1-\eta^2$, respectively.

The element of the load vector $\mathbf{Q}$ in Eq. (19) reads:

$$\mathbf{Q}^z = \frac{a}{2h}\sum_{\alpha}\sum_{\beta} Q_{\alpha\beta} \times \int_{-1}^{1} P_{\alpha}(\xi)P_{p}(\xi)f_{w}^{1}(\xi)\mathrm{d}\xi \times \int_{-1}^{1} P_{\beta}(\eta)P_{q}(\eta)f_{w}^{2}(\eta)\mathrm{d}\eta P_{r}(1). \quad (24)$$

Substituting the expressions of the displacement components, Eqs. (12)–(14), in the expressions of the strain energy Φ and of the potential energy of the load W, Eqs. (9) and (10), the governing Eq. (18) can be obtained and from that equation the coefficient vector $\mathbf{C}$ is determined.

5 Results and Discussion

In this section, an all-round FG square clamped plate ($a/b=1$) of thickness-side ratio $h/a=0.2$ is considered. The constituent materials of the FG plate are aluminum (Al) and silicon carbide (SiC) with mechanical properties: $E_m=70$ GPa, $\nu_m=0.30$, $E_c=427$ GPa, and $\nu_c=0.17$ [13].

The case of a load q_0 uniformly distributed over the upper surface of the plate is considered. To calculate any particular coefficient $Q_{\alpha\beta}$ of the series of Eq. (16), both sides of Eq. (16) are multiplied by $P_{\alpha'}(\xi)P_{\beta'}(\eta)\varpi_{\xi}(\xi)\varpi_{\eta}(\eta)$, where $\varpi_{\xi}(\xi)$, $\varpi_{\eta}(\eta)$ are weighting functions, and integrated twice from -1 to 1. From the orthogonality property of Chebyshev polynomials, the resulting integrals are different from zero only when $\alpha'=\beta'=1$, that means $P_{\alpha'}(\xi)=P_{\beta'}(\eta)=1$. As a consequence, also α and β have to be equal to 1, that means $P_{\alpha}(\xi)=P_{\beta}(\eta)=1$. From these considerations, the only coefficient different from zero is Q_{11} and is equal to q_0. In the calculations, the value of the uniformly distributed load q_0 is taken to be 10^6 N/m^2.

Numerical results are presented in terms of the following nondimensional displacement and stress parameters: $\bar{u}=u/h$, $\bar{w}=w/h$, and $\bar{\sigma}_{xx}=h^2\sigma_{xx}/(a^2|q_0|)$.

A convergence study on the central deflection $\bar{w}$ shows that the number of terms in the series must be chosen equal to 8, 8, and 6 in the x, y, and z direction, respectively, as represented in Fig. 1.

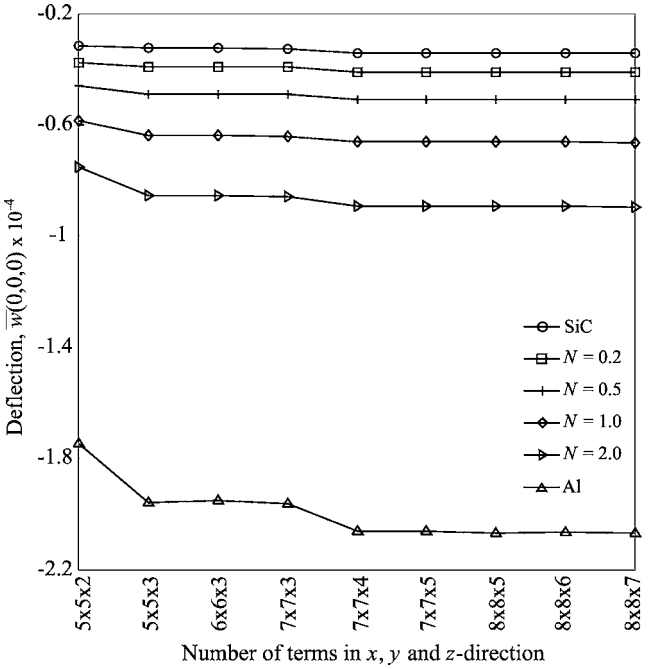


Fig. 1 Convergence study on the nondimensional central deflection $\bar{w}$

The analyses are performed for different values of the volume fraction exponent $N=0.2$, 0.5, 1, and 2.

In Fig. 2 the nondimensional central deflection $\bar{w}$ and the in-plane displacement $\bar{u}$, evaluated at $\xi=1/2$ and $\eta=1/2$, versus the thickness direction are depicted. The nondimensional central deflection $\bar{w}$ of the purely metallic plate was found to be of the largest magnitude and that of the purely ceramic plate, of the smallest magnitude. All the plates with intermediate properties undergo corresponding intermediate values. The purely aluminum plate exhibits the largest in-plane displacements in absolute value at the top and at the bottom surfaces, while the plate made only by silicon carbide has the smallest ones. This is expected because the

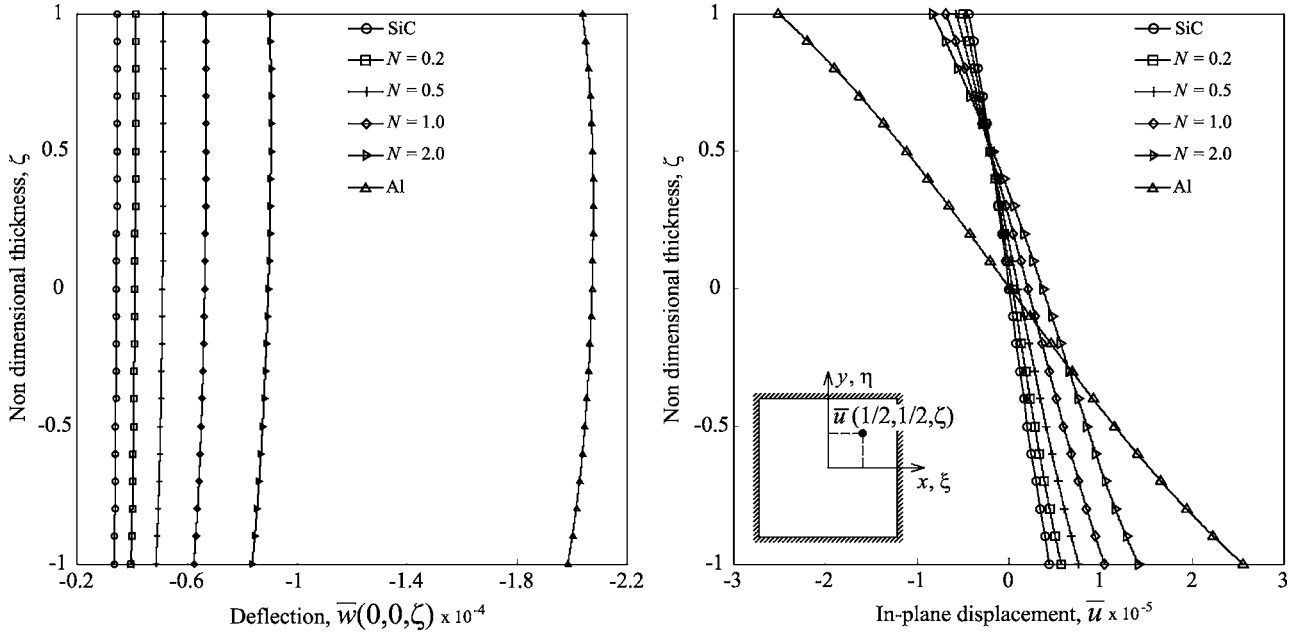


Fig. 2 Nondimensional deflection $\bar{w}$ and in-plane displacement $\bar{u}$ through the thickness

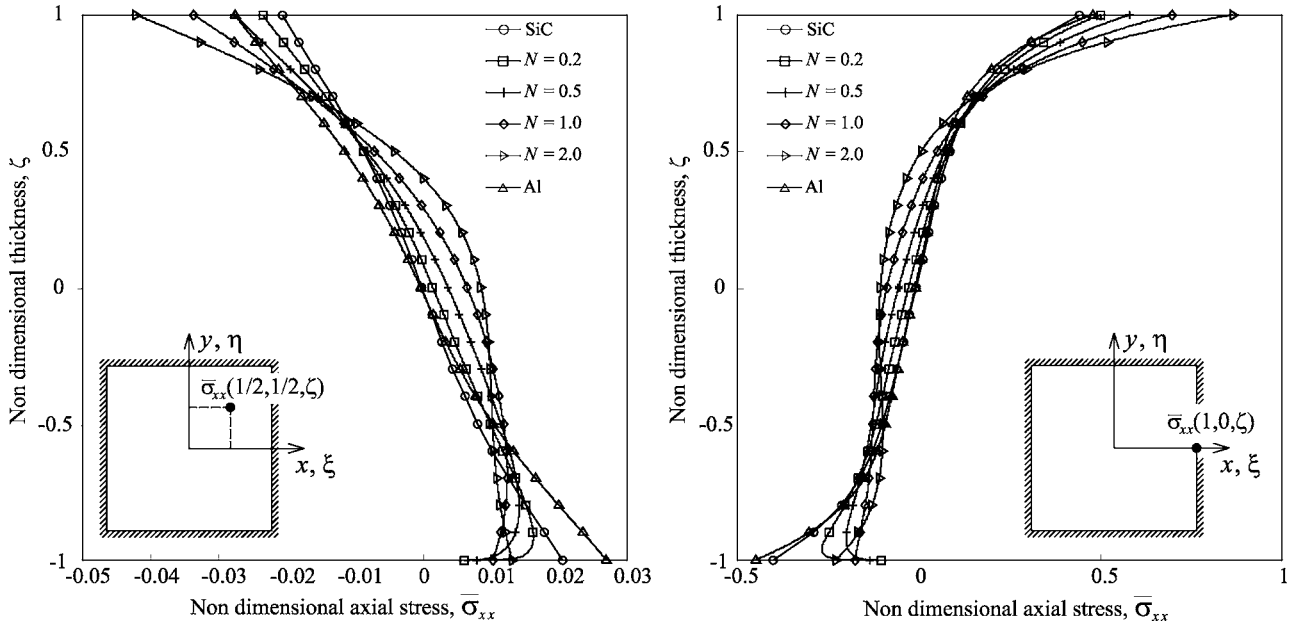


Fig. 3 Nondimensional axial stresses $\bar{\sigma}_{xx}$ at $\xi=1/2$, $\eta=1/2$ and in the clamped edge through the thickness

metallic plate has a lower stiffness than the ceramic plate. As it can be seen, at the bottom surface, $\zeta=-1$ and its vicinity, the in-plane displacements $\bar{u}$ of the FG plates are in between those with homogeneous constituents. The same takes place at the upper surface $\zeta=1$ and its vicinity. Remarkably, across the thickness, there is a central region where the FG plates with different values of the exponent N exhibit a response that is not intermediate to the fully metallic and fully ceramic plates. Besides, there is a point at $\zeta \simeq 0.5$, where the plates for different values of N share the same in-plane displacement. It should be remarked that lack of symmetry in the displacement profiles is attributable to the fact that there is no symmetry in the load itself with respect to the mid-plane of the plate; in addition the mechanical properties vary through the thickness of the plate in accordance with nonsymmetric power-laws, Eq. (1).

Figure 3 represents the plots of the nondimensional axial stresses $\bar{\sigma}_{xx}$ evaluated at $\xi=1/2$, $\eta=1/2$, and in the clamped edge ($\xi=1$, $\eta=0$) through the thickness. The axial stress $\bar{\sigma}_{xx}$ at $\xi=1/2$, $\eta=1/2$, due to the application of the pressure loading, is compressive at the top surface and tensile at the bottom surface. For the several volume fraction exponents chosen, the plate corresponding to $N=2$ yields the maximum compressive stress at the top surface, $\zeta=1$, while the purely metallic plate experiences the maximum tensile stress at the bottom surface, $\zeta=-1$. While in the clamped edge the stress profiles $\bar{\sigma}_{xx}$ is of tension at the top surface and of compression at the bottom. It is worth noting that the stress profiles $\bar{\sigma}_{xx}$ at $\xi=1/2$, $\eta=1/2$, and $\bar{\sigma}_{xx}$ at $\xi=1$, $\eta=0$ for the fully ceramic and fully metallic plates do not coincide. This is due to the fact that there is a strong dependence of the stresses on the Poisson's ratios, that for the aluminum and silicon carbide components are different.

6 Conclusion

In this paper, the Ritz energy method is developed on the basis of the three-dimensional elasticity theory to predict displacements and stresses of all-round clamped functionally graded plates subjected to a uniform load on the top surface. A two-component graded plate with a power-law variation of the volume fraction of the constituents through the thickness has been considered. Chebyshev polynomials are assumed as coordinate functions multiplied by appropriate boundary functions in order to satisfy the geometric boundary conditions. The effect of varying the volume fraction on the nondimensional displacements and stresses has been presented. From the reported results, it can be concluded that the displacements and the axial stresses in an arbitrary point of a functionally graded plate do not necessarily lie between those of the ceramic and metal. Thus, the gradient in material properties plays an important role in determining the response of the FG plates. It should be remarked that the concepts outlined in this study are not confined to the all-round clamped FG plates. They can be directly extended also to other boundary conditions, power-law distributions and thicknesses. The presented three-dimensional numerical results are important in complementing solutions derived from other plate theories like first and higher order plate theories.

References

[1] Fukui, Y., 1991, "Fundamental Investigation of Functionally Gradient Material Manufacturing System Using Centrifugal Force," JSME Int. J., Ser. III **34**, pp. 144–148.

[2] Koizumi, M., 1997, "FGM Activities in Japan," Composites, Part B **28**, pp. 1–4.

[3] Suresh, S., and Mortensen, A., 1998, *Fundamentals of Functionally Graded Materials*, The University Press, Cambridge, UK.

[4] Reddy, J. N., 2004, *Mechanics of Laminated Composite Plates and Shells: Theory and Analysis*, CRC Press, Boca Raton, FL.

[5] Della Croce, L., and Venini, P., 2004, "Finite Elements for Functionally Graded Reissner-Mindlin Plates," Comput. Methods Appl. Mech. Eng. **193**(9-11), pp. 705–725.

[6] Reddy, J. N., 2000, "Analysis of Functionally Graded Plates," Int. J. Numer. Methods Eng. **47**, pp. 663–684.

[7] Yang, J., and Shen, H.-S., 2003, "Nonlinear Bending Analysis of Shear Deformable Functionally Graded Plates Subjected to Thermo-Mechanical Loads Under Various Boundary Conditions," Composites, Part B **34**, pp. 103–115.

[8] Reddy, J. N., and Cheng, Z.-Q., 2001, "Three-Dimensional Thermomechanical Deformations of Functionally Graded Rectangular Plates," Eur. J. Mech. A/Solids **20**(5), pp. 841–855.

[9] Vel, S. S., and Batra, R. C., 2003, "Three-Dimensional Analysis of Transient Thermal Stresses in Functionally Graded Plates," Int. J. Solids Struct. **40**(25), pp. 7181–7196.

[10] Cheng, Z.-Q., and Batra, R. C., 2000, "Three-Dimensional Thermoelastic Deformations of a Functionally Graded Elliptic Plate," Composites, Part B **31**, pp. 97–106.

[11] Qian, L. F., and Batra, R. C., 2004, "Transient Thermoelastic Deformations of a Thick Functionally Graded Plate," J. Therm. Stresses **27**, pp. 705–740.

[12] Snyder, M. A., 1966, *Chebyshev Methods in Numerical Approximation*, Prentice-Hall, Englewood Cliffs, NJ.

[13] Qian, L. F., Batra, R. C., and Chen, L. M., 2004, "Analysis of Cylindrical Bending Thermoelastic Deformations of Functionally Graded Plates by a Meshless Local Petrov-Galerkin Method," Comput. Mech. **33**, pp. 263–273.

An Optical Interferometric Band as an Indicator of Plastic Deformation Front

Sanichiro Yoshida
e-mail: syoshida@selu.edu
Department of Chemistry and Physics, Southeastern Louisiana University, SLU 10878, Hammond, LA 70402

Hideyuki Ishii

Kensuke Ichinose
e-mail: ichiken@cck.dendai.ac.jp

Kenji Gomi

Kiyoshi Taniuchi

Department of Mechanical Engineering, Tokyo Denki University, 2-2 Kanda Nishiki-cho, Chiyoda, Tokyo 101-8457, Japan

The Lüders' front and a previously discovered optical interferometric band structure were observed simultaneously in steel specimens under tensile loading. The observed Lüders' front and optical band structure show the same propagation characteristics, confirming our previous interpretation that the optical band structure represents the plastic deformation front. Analysis shows that the stress at which the optical band structure begins to appear is approximately 10% lower than the corresponding Lüders' front, indicating that the optical band structure reveals the plastic deformation front with higher sensitivity than the Lüders' front. [DOI: 10.1115/1.1985431]

1 Introduction

Previously, Yoshida et al. [1] discovered an optical interferometric band-structure whose appearance is quite similar to that of the plastic deformation front. This band structure, called the white band (WB) hereafter [2], can be observed in an interferometric fringe pattern formed by a technique known as the in-plane sensitive, electronic spackle-pattern interferometry (ESPI) [3]. Near the yield point, the WB propagates in a way similar to the plastic deformation front known as the Lüders' front [4], and in later stages, it resembles a more developed deformation front which eventually stays at the location of fracture [1]. These observations indicate that the WB can be used as an indicator of the plastic deformation front, and will be useful for various studies including quantitative analysis of plastic deformation. In fact, a previous study [5] indicates that the stress is getting concentrated at the banded region where the WB is about to form. To further continue this line of research, it is important to confirm that the WB indeed represents the plastic deformation front.

In this study, we simultaneously monitored Lüders' fronts and WBs on the front and rear surfaces, respectively, of the same specimen under tensile loading. Consequently, pairs of Lüders' fronts and WBs were observed to appear concurrently and propagate synchronously at locations close to each other, indicating that they certainly represents the same phenomenon. Also of interest is that the WB begins to appear at about 10% lower stress than the Lüders' front, showing that it indicates the plastic deformation front with higher sensitivity than the Lüders' front.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, October 12, 2003; final revision, February 7, 2005. Associate Editor: K. Ravi-Chandar.

2 Experimental Setup

The material used in this study was commercially available structural steel SS400. Prior to the machining, the material was annealed at 900 °C, and its grain size was measured to be 15 μm. The material was then machined to be two types of specimens; the first type was a bone shape with a parallel part of 50 mm long, 10 mm wide, and 5 mm thick, and the second type was a rectangle of 80 mm long (the grip-to-grip length), 15 mm wide and 3 mm thick. To visualize the Lüders' front to the naked eye, we polished the rear surface of the specimen with an abrasive paper of grit number 2000. [6]. The front surface was first polished with an abrasive paper of grit number 100 and then painted white to visualize WBs at a high contrast.

We illuminated the rear surface of the specimen with a fluorescent lamp, and recorded the image with a video recorder. The ESPI setup used for the front surface of the specimen was a typical dual-beam configuration [3]. We took specklegrams of the specimen with a CCD (Charged Coupled Device) camera at a sampling rate of 1 frame/s, and stored the image data into computer memory at the same rate. The tensile load was applied at constant crosshead speeds. The applied load and the stroke of the dynamic grip were recorded at a rate of 10 sample/s. Rulers were attached to the lower grip of the test machine on both sides, so that the locations of the Lüders' line and WB could be read directly on the respective images. We formed interferometric fringe patterns by subtracting the specklegram taken at one time step from the specklegram taken several time steps before, which corresponded to the total elongation of 10–30 μm.

3 Results and Discussions

Figure 1 shows the Lüders' front and WB observed in the first type of specimen at the crosshead speed of 2.5 μm/s. In this particular run, a pair of Lüders' front and WB appeared near the lower end of the specimen and propagated upward. Except that the Lüders' front and the corresponding WB are slanting in mutually opposite orientations because they are observed from the opposite sides of the specimen, they propagate in a very similar fashion. Figure 2 plots the locations of the centers of the Lüders' front and WB measured from a reference point near the lower grip as a function of time. The solid lines are the best fits to the data points. From the slopes of these lines, the propagation velocities of the Lüders' front and the WB are found to be 1.11×10^{-4} m/s and 1.13×10^{-4} m/s, respectively. Considering that the error associated with the reading of the locations of the Lüders' line and WBs is estimated to be ±2%, it is fair to say that the Lüders' front and the WB propagate at the same velocity.

Figure 3 shows the stress-strain diagram in the plastic regime along with its expanded view near the yield point. This diagram was recorded at the same time as Fig. 1. The extents where the Lüders' fronts and WB appear are marked, respectively. From the observed yield elongation (region between points A and B) and the specimen length of 50 mm, the yield strain ε_L can be estimated to be 0.0197. Using this yield strain, the crosshead speed $V_c=2.5\times10^{-6}$ m/s and the relationship $V_L=V_C/\varepsilon_L$ derived by Sylwestrowicz and Hall [7], the theoretical velocity of the Lüders' front can be estimated to be $V_L=1.27\times10^{-4}$ m/s. This value shows a good agreement with the measurement, as the dashed line in Fig. 2 indicates.

While the theoretical velocity of the Lüders' front based on the yield strain agrees with experiment, Fig. 3 indicates that the Lüders' front or WB does not extend the whole span of the yield elongation. A possible explanation of this observation is that while the plastic deformation front begins to propagate at the lower end

 Copyright © 2005 by ASME

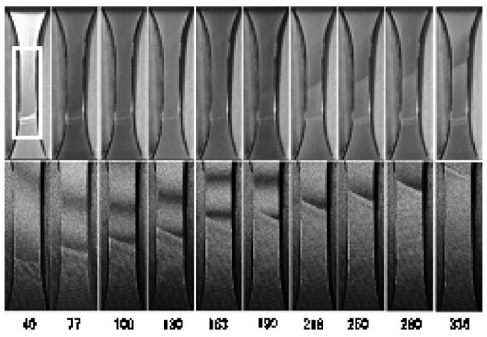


Fig. 1 Lüders' front (upper) and WB (lower) observed simultaneously at the crosshead speed of 2.5 μm/s. Numbers represent the elapsed time in s from a reference time set before the yield point. The window inserted in the leftmost image of the upper row indicates the view of the lower images approximately.

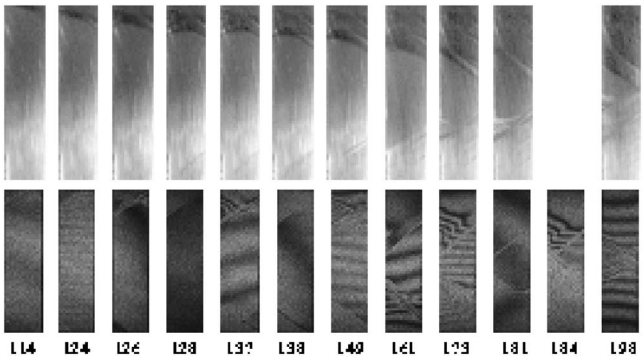


Fig. 4 Lüders' front (upper) and WB (lower) observed simultaneously at the crosshead speed of 25 μm/s. The Lüders' front taken at time step 184 is not shown because of poor quality.

of the specimen at the yield stress and completes the propagation at the upper end of the specimen at the end of the yield elongation, the initial and final parts of the propagation are not manifested as the Lüders' front or WB. The fact that the WB appears earlier and disappears later than the Lüders' front indicates that the former has higher sensitivity as an indicator of the plastic deformation

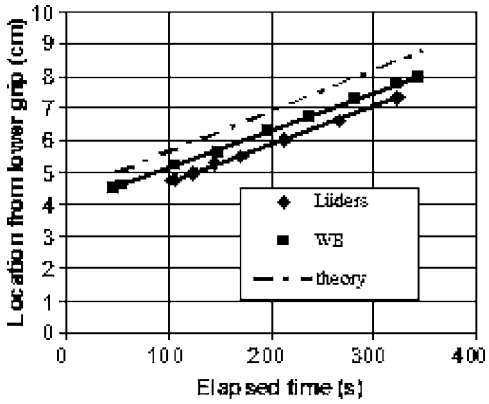


Fig. 2 Locations of Lüders' front and WB as a function of time. The crosshead speed is 2.5 μm/s.

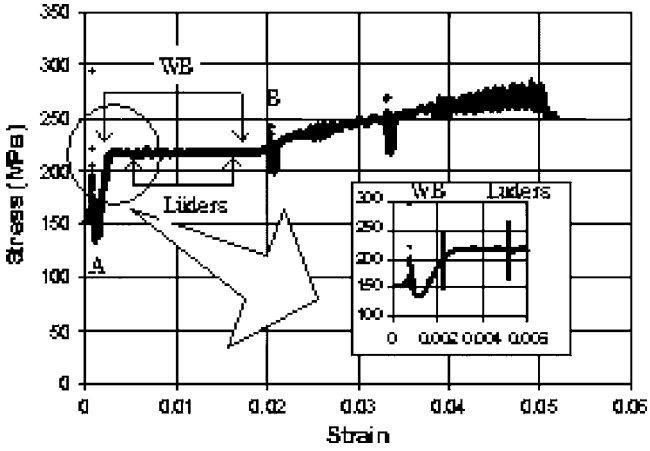


Fig. 3 The stress-strain diagram in the plastic regime recorded at the same time as Fig. 1. The expanded view indicates the stress values when the Lüders' front and WB begin to appear.

front. The expanded view in Fig. 3 shows that the WB begins to appear at 8% lower stress than the Lüders' front [8]. This observation is consistent with the previous finding by Funamoto [9] for the same material that the WB begins to appear near a circular hole at a stress value 11% lower than the corresponding stretcher-strain.

At the crosshead speed of 25 μm/s, a similar relationship between the Lüders' front and WB was observed. Figure 4 shows the Lüders' front and WB observed in the second type of specimen at this crosshead speed. In this run, two pairs of Lüders' front and WB began to appear at the upper and lower end of the specimen, respectively, within a time lag less than 30 s. At this crosshead speed, the Lüders' fronts show less sharp edges, as is normally the case. The two pairs of Lüders' front and WB propagated along the length of the specimen in mutually opposite directions. Figure 5 shows the locations of the pairs as a function of time. Like Fig. 2, the Lüders' front and WB of the same pair propagate at the same velocity, which is reasonably close to the theoretical Lüders' front velocity (dashed lines) estimated in the same fashion as Fig. 2. (The lower Lüders' front and WB show somewhat higher velocity than the theoretical value, and its reason is not clear.) Unlike Fig. 2, however, both the Lüders' front and the WB fluctuate around the fitted straight lines, indicating that when the Lüders' front forms a less sharp edge, it does not propagate at a constant velocity but somewhat fluctuates around a mean value. The WB observed under this condition follows a similar, fluctuating trend, indicating that it still represents the same phenomenon as the Lüders' front. We examined the second type of specimen at the crosshead speed of 2.5 μs/s and observed a straight trend similar

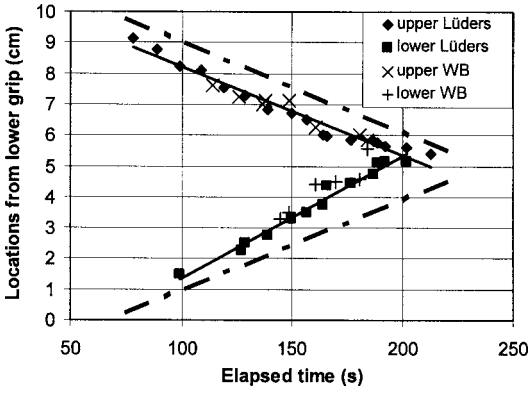


Fig. 5 Locations of Lüders' front and WB as a function of time. The crosshead speed is 25 μm/s.

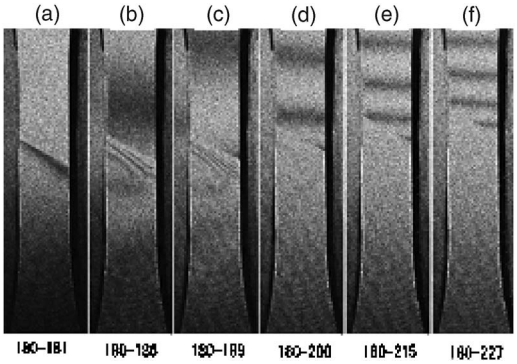


Fig. 6 WB obtained for a common specklegram with various subtraction increments

to Fig. 2. Judging from this result, the fluctuating trend observed in Fig. 5 is not caused by the difference in the shape of the specimen but the difference in the crosshead speed.

The fringe patterns observed in the regions divided by the WBs can be characterized as consisting of nearly equidistant and horizontally parallel fringes (see Figs. 1, 4, and 6). A system of perfectly equidistant and horizontally parallel fringes observed in a horizontally sensitive ESPI setup represents a rigid body rotation of the object [10]. This observation, therefore, indicates that the deformation is concentrated at the Lüders' front in such a way that the regions divided by the front experience rigid body like rotations [5]; i.e., there is a stress concentration at the Lüders' front.

Figure 6 shows fringe patterns formed by subtracting several specklegrams from a common specklegram (frame #180) varying the time-step increment. The number shown underneath each fringe pattern denotes the number of the frames involved in the subtraction. The fringe systems seen in these patterns have the following features. The region above the WB (region 1) shows nearly the equidistant, horizontally parallel system representing a rigid body like rotation. The region inside the WB (region 2) shows a dense fringes parallel to the WB, representing a concentrated strain normal to the WB. The region below the WB (region 3) does not show a clear fringe structure. As the number of the time-step increment increases in going from (a) to (f), the number of fringes in region 1 increases rather constantly. The number of fringes in region 2 also increases, at a rate much higher than region 1. In going from (a) to (c), the fringe density in region 2 becomes four or five times higher, while the number of fringes in region 1 increases by one. The fringe pattern in region 3 remains more or less the same in (a) through (f). These features can be interpreted as follows. When the Lüders' front was formed, the material in the vicinity of the front experienced a localized deformation. As the time went by, the part of the specimen held by the dynamic grip (region 3) displaced in the vertical direction as a rigid body (without substantial deformation), forming no clear fringe structure. This motion sustained the localized deformation in region 2. While this happened, the localized deformation moved up by some mechanism as the other side of the material (region 1) rotated as nearly a rigid body at an approximately constant rate.

4 Conclusion

The result of this study confirms that the WB represents the same phenomenon as the Lüders' front, supporting our previous interpretation. Now that this has been confirmed, further investigation on the plastic deformation front can be pursued through analysis on the WB. It is observed that the WB appears at about 10% lower stress than the Lüders' front, indicating that the WB has higher sensitivity than the Lüders' front as an indicator of the plastic deformation front.

Acknowledgment

This study was supported by the Tokyo Denki University, Research Institute for Science and Technology through project 2Q334. The data analysis software used in this study was written by T. Mita. We thank S. Arikawa, A. Arai, M. Shimozaki, R. Takanashi, T. Togashi, and R. Haraguchi for their support in carrying out the experiment.

References

[1] Yoshida, S., Suprapedi, Widiastuti, R., Pardede, M., Hutagalon, S., Marpaung, J., Faizal, A., and Kusnowo, A., 1996, "Dircet Observation of Developed Plastic Deformation and Its Application to Nondestructive Testing," Jpn. J. Appl. Phys., Part 2 **35**, pp. L854–857.

[2] As described in Ref. [1], this optical band usually appears to be more whitish than the bright peak of the surrounding interferometric fringes. That is why it is called the white band.

[3] Løkberg, O. J., 1993 "Recent Development in Video Speckle Interferometry," in *Speckle Metrology*, edited by R. Shiroi, Dekker, New York, USA, pp. 157–194.

[4] Hall, E. O., 1970, *Yield point phenomena in metals and alloys*, Plenum Press, New York.

[5] Yoshida, S., Muhamed, I., Pardede, M., Widiastuti, R., Muchiar, Siahan, B., and Kusnowo, A., 1997, "Optical Interferometry Applied to Analyze Deformation and Fracture of Aluminum Alloys," Theor. Appl. Fract. Mech. **27**, pp. 85–98.

[6] Ichinose, K., Kosaka, Y., Fukuda, K., and Taniuchi, K., 1999, "Detection of the Failure Zone Caused by Cyclic Loading," J. Intell. Mater. Syst. Struct., **10**, pp. 214–220.

[7] Sylwestrowicz, W. and Hall, E. O., 1951, "The Deformation and Ageing of Mild Steel," Proc. Phys. Soc. London, Sect. B **64**, pp. 495–502.

[8] Note that in Fig. 5 the WBs appear to begin later than the Lüders' fronts. This is due to the fact that the view area of the CCD camera to take the specklegrams was smaller than the view area of the video camera to take the Lüders' fronts and that the locations where the WBs begin to appear were out of the view area of the CCD camera.

[9] Funamoto, Y., 2002, "The Observation of Yielding Region of the Circular Hole Vicinity under the Gradually Increase Cyclic Loading and Application of ESPI in the Failure Region of Low Carbon Steel," Master's thesis, Tokyo Denki University, Tokyo, Japan, p. 102 (in Japanese).

[10] Yoshida, S., Suprapedi, Widiastuti, R., Astuti, E. T., and Kusnowo, A., 1995, "Phase Evaluation for Electronic Speckle-Pattern Interferometry Deformation Analysis," Opt. Lett. **20**, pp. 755–757.

On the Buckling of a Circular Plate on an Elastic Foundation

C. Y. Wang
Professor
Departments of Mathematics and Mechanical Engineering, Michigan State University, East Lansing, MI 48824
Member ASME

The buckling of a circular plate on an elastic foundation is studied analytically. The buckling mode may not be axisymmetric as previously assumed. [DOI: 10.1115/1.1988347]

Introduction

The prediction of buckling of structural members restrained laterally is important in the design of engineering components which are attached to a foundation. The buckling of beams on an elastic foundation has been well studied (see Hetenyi [1]). It was found that for stiffer foundations, the buckling load corresponds to a mode with increased nodes. The buckling of rectangular plates was discussed by Seide [2] and Vlasov and Leontev [3] but only for very narrow plates. Analytic solution for the buckling loads for a rectangular plate is possible if two opposite sides are simply supported [4], otherwise numerical means is necessary [5]. The buckling of circular plates was studied by several authors [6–9]. However, these sources only considered axisymmetric buckling, which may not lead to the correct buckling load. The purpose of the present Note is to complete the results of the buckling of circular plates by including the nonaxisymmetric buckling modes, thus correctly determining the buckling loads.

Formulation

Following Yu [6], the governing equation for a thin (Kirchkoff) plate attached to a Winkler foundation is

$$D\nabla^4 w + N\nabla^2 w + kw = 0 \tag{1}$$

where w is the lateral displacement, D is the flexural rigidity, N is the uniform compressive load at the edge, and k is the spring constant of the foundation. If we normalize lengths by the radius of the plate R, Eq. (1) can be written as

$$\nabla^4 w + \lambda^2 \nabla^2 w + \gamma^4 w = 0 \tag{2}$$

where $\lambda^2 \equiv NR^2/D$ is a load parameter and $\gamma^4 \equiv kR^4/D$ is a stiffness parameter. Suppose there are n nodal diameters. In polar coordinates (r, θ) set

$$w = u(r)\cos(n\theta) \tag{3}$$

Equation (2) becomes

$$L^2 u + \lambda^2 L u + \gamma^4 u = 0 \tag{4}$$

where L is the operator

$$L \equiv \frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{n^2}{r^2} \tag{5}$$

The general solution which is bounded at the origin is as follows. Let J_n be the Bessel function of order n. If $\lambda > \sqrt{2}\gamma$

$$u = C_1 J_n(\alpha r) + C_2 J_n(\beta r) \tag{6}$$

where

$$\alpha = \left(\frac{\lambda^2 + \sqrt{\lambda^4 - 4\gamma^4}}{2}\right)^{1/2}, \quad \beta = \left(\frac{\lambda^2 - \sqrt{\lambda^4 - 4\gamma^4}}{2}\right)^{1/2} \tag{7}$$

If $\lambda = \sqrt{2}\gamma$ the solution is

$$u = C_1 J_n(\gamma r) + C_2 r J_{n+1}(\gamma r) \tag{8}$$

If $\lambda < \sqrt{2}\gamma$ the solution is

$$u = C_1 \,\mathrm{Re}[J_n(i\delta r)] + C_2 \,\mathrm{Im}[J_n(i\delta r)] \tag{9}$$

where $i = \sqrt{-1}$ and

$$\delta = \left(\frac{-\lambda^2 + \sqrt{4\gamma^4 - \lambda^4 i}}{2}\right)^{1/2} \tag{10}$$

These solutions reduce to those of the axisymmetric cases studied by previous authors when $n=0$. In what follows we shall present the results for the clamped boundary, the simply supported boundary, and the free boundary.

Results and Discussion

For the clamped boundary, the edge conditions are zero displacement and zero slope

$$u(1) = 0, \quad u'(1) = 0 \tag{11}$$

For nontrivial solutions of C_1, C_2, Eqs. (11) give an exact determinant for the three cases Eqs. (6), (8), and (9). Then for each n and γ, the buckling load, represented by λ, is found from the characteristic equation by simple root search. Figure 1 shows λ rises monotonically with the foundation stiffness parameter γ. We see that the $n=0$ symmetric solution weaves with the $n=1$ nonsymmetric solution. When stiffness is zero, the untethered plate buckles axisymmetrically with $\lambda = 3.8317$ which is the first zero of $J_1(\lambda) = 0$. When γ is increased beyond 3.644 ($\lambda = 6.022$), the $n=1$ nonsymmetric mode gives the correct lower buckling load. This persists until $\gamma = 5.185$ ($\lambda = 7.994$) where the $n=0$ mode again determines the buckling load. The next two switches are at $\gamma = 6.912$ ($\lambda = 10.265$) and $\gamma = 8.446$ ($\lambda = 12.355$). Kline and Hancock [8] obtained only the $n=0$ curve, thus overestimated the buckling load where the $n=1$ mode gives the buckling load. In the range of stiffness considered, the higher modes do not affect the buckling loads but may be a factor when the foundation is extremely stiff. Notice that the buckling loads all satisfy $\lambda > \sqrt{2}\gamma$.

For the simply supported boundary, the edge conditions are zero deflection and zero moment

$$u(1) = 0, \quad u''(1) + \nu[u'(1) - n^2 u(1)] = 0 \tag{12}$$

where ν is the Poisson ratio. When the foundation is absent, the buckling load corresponds to the axisymmetric mode and is given by the first root of

$$\lambda J_0(\lambda) - (1 - \nu) J_1(\lambda) = 0 \tag{13}$$

For $\nu = 0.3$ (metals) we find $\lambda = 2.0489$. Figure 2 shows the results are similar to that of the clamped case. The buckling loads are lower, and all three forms of Eqs. (6), (8), and (9) are used. The transition points between the $n=0$ modes and the $n=1$ modes are at $\gamma = 3.047$ ($\lambda = 4.359$), $\gamma = 4.604$ ($\lambda = 6.605$), $\gamma = 6.228$ ($\lambda = 8.843$), $\gamma = 7.795$ ($\lambda = 11.072$), etc. Again, the buckling loads found by the previous authors who considered only the symmetric modes may be erroneous.

The plate with free edge on an elastic foundation has not been considered before. Note that without the foundation, the buckling problem does not exist—the buckling force is always zero. The edge boundary conditions are zero moment and zero shear

$$u''(1) + \nu[u'(1) - n^2 u(1)] = 0 \tag{14}$$

Contributed by the Applied Mechanics Division of The American Society of Mechanical Engineers for publication in the ASME Journal of Applied Mechanics. Manuscript received by the ASME Applied Mechanics Division, October 11, 2004; final revision, November 22, 2004. Associate Editor: N. Triantafyllides.

 Copyright © 2005 by ASME

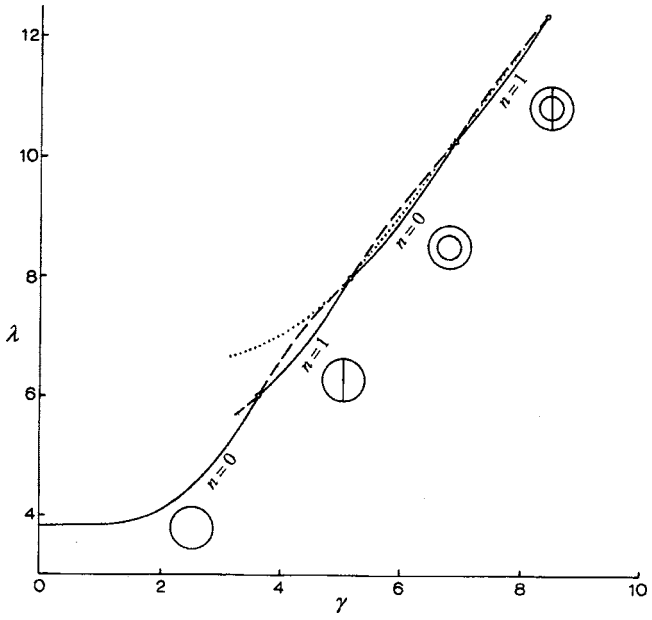


Fig. 1 Buckling load parameter λ versus foundation stiffness parameter γ for the clamped edge. Only the solid line represents the buckling load. Small circles denote a change in buckling modes. The dotted line is from the $n=2$ mode. The buckling modes are shown schematically as disks with nodal lines.

$$u'''(1) + u''(1) - [1 + n^2(2-\nu) - \lambda^2]u'(1) + n^2(3-\nu)u(1) = 0 \tag{15}$$

The term with λ^2 in Eq. (15) is due to the compressive force component [10]. We find the buckling load is low, such that $\lambda < \sqrt{2}\gamma$. Figure 3 shows, for $\nu=0.3$, the buckling load starts from zero, corresponding to the $n=1$ mode, until $\gamma=2.059$ ($\lambda=2.046$)

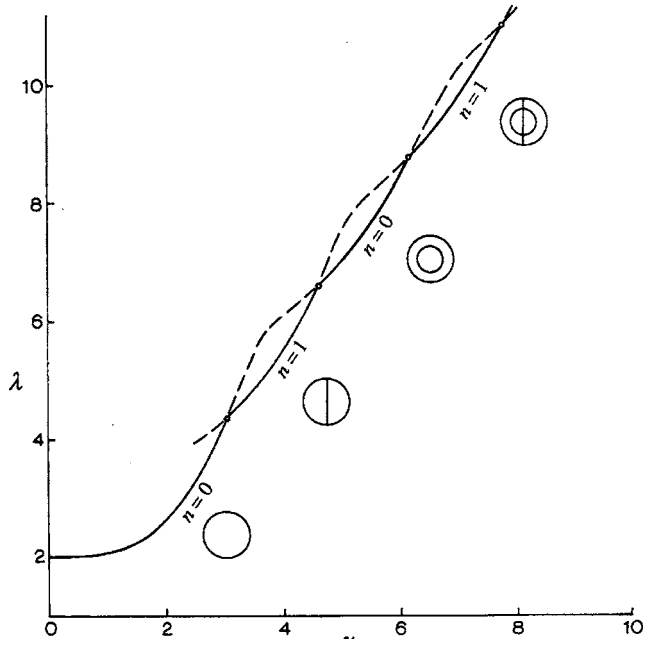


Fig. 2 λ versus γ for the simply-supported edge. Legend is similar to that of Fig. 1.

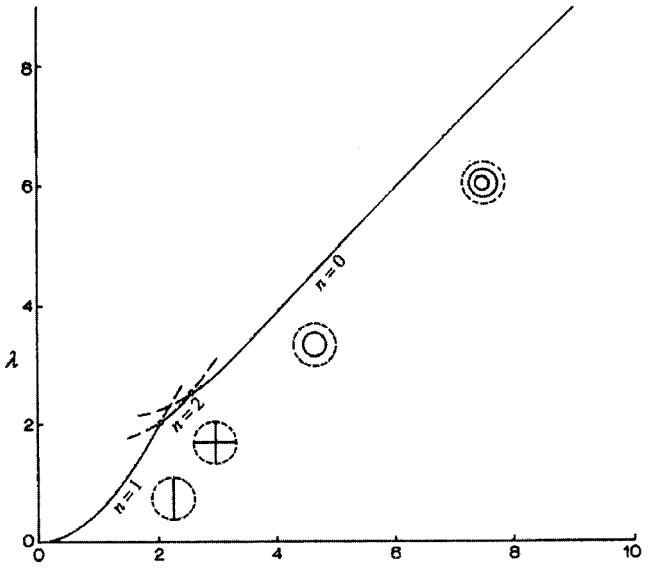


Fig. 3 λ versus γ for the free edge. Legend is similar to that of Fig. 1.

where the buckling load is determined by the $n=2$ mode. Then as γ reaches 2.550 ($\lambda=2.502$) the buckling mode becomes symmetric ($n=0$) and stays symmetric for the range of stiffness considered.

The buckling load of a thin circular plate on an elastic foundation is now correctly determined. We found that the buckling load is determined not only by the axisymmetric modes considered by previous authors, but also by nonaxisymmetric modes as well. Our figures will be useful in the basic design of embedded circular plates.

Last, in this paper the characteristic equations are exact, and the solutions can be found to any accuracy. Thus exact solutions can be used to check approximate or numerical results [11]. On the other hand, numerical solutions such as the Ritz method rely on the a priori perception of the correct buckling mode. This would be difficult in view of the many buckling modes found in this paper.

References

[1] Hetenyi, M., 1948, *Beams on Elastic Foundation*, University of Michigan Press, Ann Arbor, MI.

[2] Seide, P., 1958, "Compressive Buckling of a Long Simply Supported Plate on an Elastic Foundation," J. Aerosp. Sci., **25**, pp. 382–394.

[3] Vlasov, V. Z., and Leontev, U. N., 1966, *Beams, Plates and Shells on Elastic Foundations*, Translated from the Russian by Israel Prog. Sci. Trans., Jerusalem.

[4] Warren, R. C., 1980, "Buckling of a Rectangular Plate on an Elastic Foundation, Compressed in Two Directions," Ph.D. thesis, Michigan State University.

[5] Raju, K. K., and Rao, G. V., 1988, "Thermal Post-Buckling of a Square Plate Resting on an Elastic Foundation by Finite Element Method," Comput. Struct., **28**, pp. 195–199.

[6] Yu, Y. Y., 1957, "Axisymmetrical Bending of Circular Plates Under Simultaneous Action of Lateral Load, Force in the Middle Plane, and Elastic Foundation," J. Appl. Mech., **24**, pp. 141–143.

[7] Galletly, G. D., 1959, "Circular Plates on a Generalized Elastic Foundation," J. Appl. Mech., **26**, p. 297.

[8] Kline, L. V., and Hancock, J. O., 1965, "Buckling of Circular Plate on Elastic Foundation," J. Eng. Ind., **87**, pp. 323–324.

[9] Wolkowisky, J. H., 1969, "Buckling of the Circular Plate Embedded in Elastic Springs, an Application to Geophysics," Commun. Pure Appl. Math., **22**, pp. 367–667.

[10] Yamaki, N., 1958, "Buckling of a Thin Annular Plate Under Uniform Compression," J. Appl. Mech., **25**, pp. 267–273.

[11] Wang, C. M., Wang, C. Y., and Reddy, J. N., 2005, *Exact Solutions for Buckling of Structural Members*, CRC Press, Boca Raton.

Free Vibrations of Thick, Complete Conical Shells of Revolution From a Three-Dimensional Theory

Jae-Hoon Kang[1]
Associate Professor
e-mail: jhkang@cau.ac.kr
Department of Architectural Engineering, Chung-Ang University, 221 Heuksuk-Dong, Dongjak-Ku, Seoul 156-756, South Korea

Arthur W. Leissa
Adjunct Professor
Fellow ASME
Colorado State University, Fort Collins, CO 80523-1374

A three-dimensional (3D) method of analysis is presented for determining the free vibration frequencies and mode shapes of thick, complete (not truncated) conical shells of revolution in which the bottom edges are normal to the midsurface of the shells based upon the circular cylindrical coordinate system using the Ritz method. Comparisons are made between the frequencies and the corresponding mode shapes of the conical shells from the authors' former analysis with bottom edges parallel to the axial direction and the present analysis with the edges normal to shell midsurfaces. [DOI: 10.1115/1.1989355]

1 Introduction

Recently the present authors [1] demonstrated a three-dimensional (3D) method of analysis for determining the free vibration frequencies and mode shapes of thick, complete (not truncated) conical shells of revolution. An edge of a shell is typically assumed to be normal to the midsurface of the shell. However, the bottom edge of the conical shell that they used in [1] was not normal to the midsurface, but parallel to the axial (z) direction (or, to the axis of revolution) in order to simplify the evaluation of energy functionals.

In the present paper, the conical shells of revolution with the bottom edge normal to the midsurface of the shell are analyzed by a 3D method based upon the circular cylindrical coordinate system (r,z,θ) using the Ritz method as in [1]. All the symbols and variables used in the present paper are the same as those used in the former analysis. The only difference in analysis between the present paper and [1] is that the domain for the shell, and therefore the limits of energy integrals, are changed. Convergence studies were made for the present work, which are similar to those shown in [1], but are omitted here to save space.

2 Method of Analysis

A representative cross section of a complete conical shell of revolution with the vertex half-angle α, the radius of midsurface of the cone at the bottom face R, and the uniform wall thickness $h(=H\sin\alpha)$, where H is the vertical slant thickness, is shown in Fig. 1. The cylindrical coordinate system (r,z,θ), also shown in the figure, is used in the analysis. The axes of r, z, and θ are the radial, axial, and circumferential coordinates, respectively. The origin of the (r,z) coordinates is located at the vertex of the midsurface of the conical shell. The straight lines $z_{i,m,o}$ in Fig. 1 are the inner, middle, and outer surfaces of the conical shell, respectively, and their equations are expressed as

$$z_m = r\cot\alpha, \tag{1}$$

$$z_{i,o} = r\cot\alpha \pm H/2. \tag{2}$$

The bottom edge is normal to the midsurface of the conical shell and its equation is expressed as

$$z_b = -(r-R)\tan\alpha + R\cot\alpha, \tag{3}$$

while the equation of the edge in [1] was $r=R$, which is parallel to the axial (z) direction. However, it is clear that the areas in the cross sections of the shells used in the present and former [1] analyses are equal. It can also be shown that the material volumes are the same for both types of configuration. The line z_b intersects the lines $z_{i,o}$ at $r=R_{i,o}$, respectively, where

$$R_{i,o} = R \mp (H\sin\alpha\cos\alpha)/2. \tag{4}$$

The domain (Ω) of the shell is therefore described by

$$0 \leqslant r \leqslant R_i, \quad z_o \leqslant z \leqslant z_i, \quad 0 \leqslant \theta \leqslant 2\pi, \tag{5a}$$

and

$$R_i \leqslant r \leqslant R_o, \quad z_o \leqslant z \leqslant z_b, \quad 0 \leqslant \theta \leqslant 2\pi. \tag{5b}$$

The domain (Ω) expressed in terms of the nondimensional cylindrical coordinates (Ψ,ζ,θ), defined by $\psi \equiv r/R$ and $\zeta \equiv z/H$, is given by

$$0 \leqslant \psi \leqslant \psi_i, \quad \zeta_o \leqslant \zeta \leqslant \zeta_i, \quad 0 \leqslant \theta \leqslant 2\pi, \tag{6a}$$

and

$$\psi_i \leqslant \psi \leqslant \psi_o, \quad \zeta_o \leqslant \zeta \leqslant \zeta_b, \quad 0 \leqslant \theta \leqslant 2\pi, \tag{6b}$$

where

$$\psi_{i,o} \equiv R_{i,o}/R = 1 \mp (H\sin\alpha\cos\alpha)/2R, \tag{7}$$

$$\zeta_{i,o} \equiv z_{i,o}/H = (R\psi\cot\alpha)/H \pm 1/2, \tag{8}$$

$$\zeta_b \equiv z_b/H = R[\cot\alpha - (\psi-1)\tan\alpha]/H. \tag{9}$$

The maximum potential (strain) energy ($V_{\max}$) and the maximum kinetic energy ($T_{\max}$) functionals in a cycle of vibratory motion are as follows:

$$V_{\max} = \frac{GR^2}{2H}\left[\int_0^{\psi_i}\int_{\zeta_o}^{\zeta_i} I_V\psi\,d\zeta\,d\psi + \int_{\psi_i}^{\psi_o}\int_{\zeta_o}^{\zeta_b} I_V\psi\,d\zeta\,d\psi\right], \tag{10}$$

$$T_{\max} = \frac{\rho\omega^2 HR^2}{2}\left[\int_0^{\psi_i}\int_{\zeta_o}^{\zeta_i} I_T\psi\,d\zeta\,d\psi + \int_{\psi_i}^{\psi_o}\int_{\zeta_o}^{\zeta_b} I_T\psi\,d\zeta\,d\psi\right], \tag{11}$$

where

$$I_V \equiv [(\lambda/G)(\kappa_1+\kappa_2+\kappa_3)^2 + 2(\kappa_1^2+\kappa_2^2+\kappa_3^2) + \kappa_4^2]\Gamma_1 + (\kappa_5^2+\kappa_6^2)\Gamma_2, \tag{12}$$

$$I_T \equiv (U_r^2+U_z^2)\Gamma_1 + U_\theta^2\Gamma_2, \tag{13}$$

and ρ is mass density per unit volume, ω is a natural frequency, λ and G are the Lamé parameters, $\kappa_{1,\ldots,6}$, and $\Gamma_{1,2}$ are defined as in Eqs. (14) and (15) of [1], respectively; and the displacement functions U_r, U_z, U_θ are also assumed as algebraic polynomials as in Eqs. (17) of [1].

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of The American Society of Mechanical Engineers for publication in the ASME Journal of Applied Mechanics. Manuscript received by the ASME Applied Mechanics Division, September 16, 2004; final revision, January 20, 2005. Associate Editor: O. M. O'Reilly.

 Copyright © 2005 by ASME

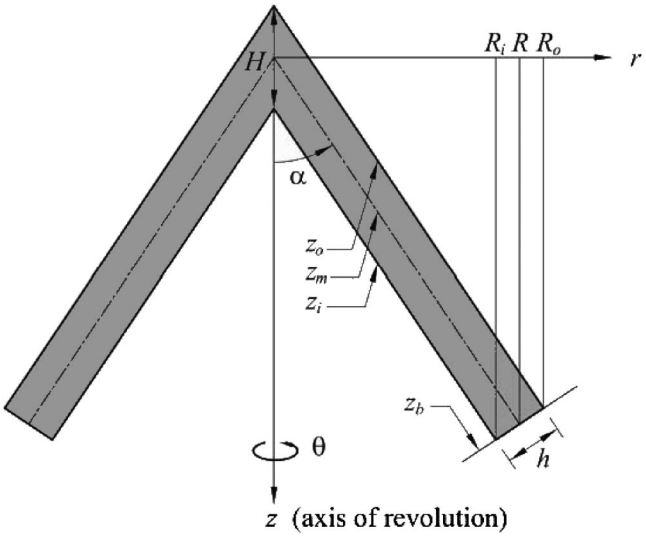


Fig. 1 A cross section of thick conical shell of revolution with the bottom edge *normal* to the midsurface, and the circular cylindrical coordinate system (r, z, θ)

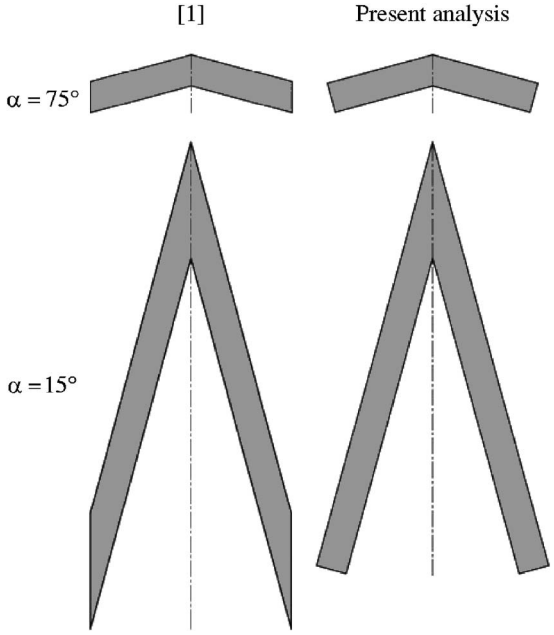


Fig. 2 Comparison of thick (h/R=0.3) conical shells having bottom edges of present analysis with those of [1]

3 Numerical Results and Comparisons

Tables 1 and 2 present the nondimensional frequencies in $\omega R\sqrt{\rho/G}$ of completely free, very thick (h/R=0.3), and moderately thick (h/R=0.1), complete conical shells of revolution, respectively, with the vertex half-angles of α=15 deg, 30 deg, 45 deg, 60 deg, and 75 deg for Poisson's ratio (ν)=0.3. Thirty frequencies are given for each configuration, which arise from six circumferential wave numbers (n=0^T, 0^A, 1, 2, 3, 4) and the first five modes (s=1, 2, 3, 4, 5) for each value of n, where the superscripts T and A indicate torsional and axisymmetric modes, respectively. The numbers in parentheses identify the first five frequencies for each configuration. The zero frequencies of rigid body modes are omitted from the tables. Tables 1 and 2 are corresponding to Tables 3 and 4 in [1].

It is seen from Tables 1 and 2 that the fundamental (lowest) frequencies of all the configurations are for modes having two (n=2) circumferential waves irrespective of the thickness ratio

Table 1 Nondimensional frequencies in $\omega R\sqrt{\rho/G}$ of completely free, conical shells with h/R=0.3 for ν=0.3

n	s	α=15°	α=30°	α=45°	α=60°	α=75°
	1	1.274	2.479	3.544	4.391	4.943
	2	2.044	3.986	5.720	7.128	8.076
$\mathbf{0}^T$	**3**	2.777	5.418	7.784	9.723	10.50
	4	3.498	6.826	9.804	10.56	11.07
	5	4.223	8.236	10.63	11.67	11.71
	1	1.431	**1.621(3)**	**1.580(3)**	**1.461(2)**	**1.280(2)**
	2	1.734	2.416	2.855	3.174	3.378
$\mathbf{0}^A$	**3**	2.169	3.222	3.864	4.016	3.949
	4	2.763	3.706	4.908	6.003	6.711
	5	2.832	4.997	6.844	8.027	8.652
	1	**0.7055(2)**	**1.751(5)**	**2.055(4)**	**2.217(4)**	**2.320(4)**
	2	**1.220(5)**	2.033	2.495	2.650	2.715
1	**3**	1.548	2.658	3.660	4.556	5.147
	4	1.740	2.870	3.949	4.997	5.680
	5	2.198	3.880	5.525	6.180	6.648
	1	**0.5043(1)**	**0.5781(1)**	**0.6353(1)**	**0.6790(1)**	**0.7074(1)**
	2	**1.124(3)**	**1.692(4)**	**2.078(5)**	**2.264(5)**	**2.330(5)**
2	**3**	1.949	2.723	3.022	3.313	3.525
	4	2.384	3.369	4.006	4.143	4.213
	5	2.756	4.021	4.853	5.852	6.508
	1	**1.207(4)**	**1.322(2)**	**1.410(2)**	**1.474(3)**	**1.514(3)**
	2	2.072	2.778	3.235	3.468	3.571
3	**3**	3.005	3.707	4.113	4.495	4.759
	4	3.329	4.627	5.503	5.691	5.784
	5	4.019	5.428	6.168	7.177	7.828
	1	2.049	2.196	2.307	2.387	2.436
	2	3.102	3.862	4.306	4.538	4.652
4	**3**	4.079	4.715	5.236	5.690	5.985
	4	4.350	5.889	6.956	7.232	7.358
	5	5.327	6.805	7.495	8.460	9.079

Notes: T=torsional mode; A=axisymmetric mode. Numbers in parentheses identify frequency sequence.

Table 2 Nondimensional frequencies in $\omega R\sqrt{\rho/G}$ of completely free, conical shells with $h/R=0.1$ for $\nu=0.3$

n	s	$\alpha=15°$	$\alpha=30°$	$\alpha=45°$	$\alpha=60°$	$\alpha=75°$
	1	1.323	2.558	3.622	4.442	4.959
	2	2.162	4.182	5.926	7.274	8.126
$\mathbf{0}^T$	**3**	2.975	5.760	8.168	10.03	11.21
	4	3.778	7.317	10.38	12.76	14.29
	5	4.577	8.864	12.59	15.50	17.43
	1	1.442	1.523	**1.383(5)**	**1.141(4)**	**0.7770(3)**
	2	1.672	2.019	2.135	2.119	1.937
$\mathbf{0}^A$	**3**	1.968	2.730	3.187	3.400	3.458
	4	2.349	3.251	3.856	4.057	3.906
	5	2.758	3.552	4.625	5.520	6.048
	1	**0.7509(4)**	1.557	1.546	**1.366(5)**	**1.108(5)**
	2	1.278	2.034	2.434	2.618	2.655
1	**3**	1.538	2.151	2.558	2.664	2.719
	4	1.671	2.812	3.584	4.296	4.751
	5	1.985	2.875	4.048	5.084	5.741
	1	**0.1931(1)**	**0.2271(1)**	**0.2476(1)**	**0.2568(1)**	**0.2582(1)**
	2	**0.5571(3)**	**0.9301(4)**	**1.219(4)**	1.424	1.562
2	**3**	1.100	1.889	2.349	2.436	2.383
	4	1.553	2.548	2.789	3.180	3.502
	5	2.019	2.767	3.933	4.191	4.234
	1	**0.4647(2)**	**0.5172(2)**	**0.5519(2)**	**0.5737(2)**	**0.5854(2)**
	2	0.9300	**1.384(5)**	1.771	2.079	2.286
3	**3**	1.473	2.431	3.252	3.586	3.606
	4	2.009	3.434	3.634	4.045	4.484
	5	2.704	3.527	4.949	5.756	5.816
	1	**0.8314(5)**	**0.9036(3)**	**0.9544(3)**	**0.9891(3)**	**1.010(4)**
	2	1.425	1.979	2.458	2.839	3.091
4	**3**	2.037	3.161	4.164	4.640	4.682
	4	2.720	4.354	4.577	5.003	5.521
	5	3.606	4.445	6.018	7.298	7.411

Notes: T=torsional mode; A=axisymmetric mode. Numbers in parentheses identify frequency sequence.

(h/R) and the vertex half-angle (α) as in [1]. For $h/R=0.1$, the second frequencies are all for $n=3$ as in the former analysis of [1].

Table 3 compares the first five nondimensional frequencies in $\omega R\sqrt{\rho/G}$ and the corresponding mode shapes (n,s) of completely free, complete conical shells of revolution from the former analysis in [1] with bottom edges parallel to the axial (z) direction and the present analysis with the edges normal to shell midsurfaces for $\nu=0.3$. It is observed that most of the frequencies from the present analysis are larger than ones from the former one [1] with some exceptions where the frequencies by the present method are un-

Table 3 Comparisons of the first five frequencies in $\omega R\sqrt{\rho/G}$ and the corresponding mode shapes (n,s) of completely free, complete conical shells of revolution from the former analysis [1] with bottom edges *parallel* to the axial (z) direction and the present analysis with the bottom edges *normal* to shell midsurfaces ($\nu=0.3$)

		$h/R=0.3$				$h/R=0.1$			
α	Order	Reference [1]		Present		Reference [1]		Present	
	1	0.3949	(2,1)	0.5043	(2,1)	0.1831	(2,1)	0.1931	(2,1)
	2	0.6938	(1,1)	0.7055	(1,1)	0.4151	(3,1)	0.4647	(3,1)
15°	3	0.7627	(3,1)	1.124	**(2,2)**	0.5495	(2,2)	0.5571	(2,2)
	4	1.008	(2,2)	1.207	**(3,1)**	0.7001	(4,1)	0.7509	**(1,1)**
	5	1.061	(4,1)	1.220	**(1,2)**	0.7495	(1,1)	0.8314	**(4,1)**
	1	0.5155	(2,1)	0.5781	(2,1)	0.2209	(2,1)	0.2271	(2,1)
	2	1.101	(3,1)	1.322	(3,1)	0.4931	(3,1)	0.5172	(3,1)
30°	3	1.651	(2,2)	1.621	$(0^A,1)$	0.8487	(4,1)	0.9036	(4,1)
	4	1.655	$(0^A,1)$	1.692	**(2,2)**	0.9244	(2,2)	0.9301	(2,2)
	5	1.711	(4,1)	1.751	(1,1)	1.362	(3,2)	1.384	(3,2)
	1	0.5999	(2,1)	0.6353	(2,1)	0.2434	(2,1)	0.2476	(2,1)
	2	1.299	(3,1)	1.410	(3,1)	0.5390	(3,1)	0.5519	(3,1)
45°	3	1.608	$(0^A,1)$	1.580	$(0^A,1)$	0.9283	(4,1)	0.9544	(4,1)
	4	2.024	(1,1)	2.055	(1,1)	1.216	(2,2)	1.219	(2,2)
	5	2.075	(4,1)	2.078	**(2,2)**	1.394	$(0^A,1)$	1.383	$(0^A,1)$
	1	0.6626	(2,1)	0.6790	(2,1)	0.2545	(2,1)	0.2568	(2,1)
	2	1.427	(3,1)	1.461	$(0^A,1)$	0.5678	(3,1)	0.5737	(3,1)
60°	3	1.481	$(0^A,1)$	1.474	**(3,1)**	0.9782	(4,1)	0.9891	(4,1)
	4	2.203	(1,1)	2.217	(1,1)	1.149	$(0^A,1)$	1.141	$(0^A,1)$
	5	2.292	(4,1)	2.264	**(2,2)**	1.370	(1,1)	1.366	(1,1)
	1	0.7032	(2,1)	0.7074	(2,1)	0.2575	(2,1)	0.2582	(2,1)
	2	1.289	$(0^A,1)$	1.280	$(0^A,1)$	0.5839	(3,1)	0.5854	(3,1)
75°	3	1.503	(3,1)	1.514	(3,1)	0.7804	$(0^A,1)$	0.7770	$(0^A,1)$
	4	2.317	(1,1)	2.320	(1,1)	1.007	(4,1)	1.010	(4,1)
	5	2.339	(2,2)	2.330	(2,2)	1.110	(1,1)	1.108	(1,1)

derlined. Most of the frequencies of the former [1] configuration are significantly lower than those of the present one, because the former one is longer (for the same R) and its bottom edge is thin, compared with the present shell (Fig. 2). As expected, the frequencies for $\alpha=15$ deg are considerably different (see Fig. 2) for the two types of bottom edges. For $\alpha=75$ deg the conical shell is shallow, the two types are nearly the same, and the frequencies are nearly the same. It is interesting to note that as the shell thickness becomes thicker and α becomes smaller the order of the mode shapes (n,s) is changed, particularly in higher modes. Such mode shapes (n,s) from the present method are in boldface type. However, the (n,s) for the first mode shapes, irrespective of configurations and all the (n,s) for $h/R=0.1$ and $\alpha\leqslant 30$ deg are not changed.

References

[1] Kang, J.-H., and Leissa, A. W., 2004, "Three-Dimensional Vibration Analysis of Thick Complete Conical Shells," ASME J. Appl. Mech., **71**(4), pp. 502–507.

The Virtual Mass of a Rotating Sphere in Fluids

Abdullah Abbas Kendoush
Department of Heat Transfer, Center of Engineering Physics, Ministry of Sciences and Technology, Baghdad, Iraq

An analytical solution to the virtual mass of a rotating fluid or solid sphere is obtained. The solution is valid at Reynolds number <10. *The solution was based on integrating the kinetic energy of the fluid round the rotating sphere. The value of the virtual mass coefficient of the rotating sphere was found to be equal to 5.*
[DOI: 10.1115/1.1989357]

Introduction

When a sphere rotates in a fluid, a force is generated due to the acceleration imparted to the fluid that rotates with the sphere. A sphere here implies either fluid or solid sphere. This force must be considered in the general equation of momentum. This force is called the virtual or "added mass" force. Similar force occurs also when a sphere oscillates in fluid or starts to move linearly from rest. Davis [1] derived the following equation for the virtual mass coefficient of an oscillating sphere:

$$C_v = \frac{1}{2} - \frac{3g}{16\Omega^2 a} \tag{1}$$

where g is the acceleration due to gravity, Ω is the angular velocity, and a is the radius of the sphere.

Stokes [2] calculated C_v of a sphere oscillating sinusoidally at low Reynolds number ($\mathrm{Re}<1$) as follows:

$$C_v = \frac{1}{2} + \frac{9(2\nu/\Omega)^{1/2}}{4a} \tag{2}$$

where ν is the kinematic viscosity of the fluid. Based on the above, Drew and Lahey [3] conjectured that the C_v of a rotating sphere must depend on the angular speed. Chan et al. [4] acknowledged the absence of any precise information on the magnitude of the C_v of the rotating sphere. They were forced to use the classical value of half for the accelerating and nonrotating sphere in their analysis.

Kurose and Komori [5] analyzed numerically the drag and lift forces on the rotating sphere in a linear shear flow without mentioning the forces of the virtual mass. The virtual mass of the accelerating and nonrotating sphere derived earlier (Milne-Thomson [6], p. 491) was based on potential flow. This derivation that resulted in $C_v=1/2$ was used later in viscous flow computations by numerous authors, e.g., Bagchi and Balachandar [7] and Magnaudet et al. [8].

The purpose of the present paper is to provide an analytical solution for the virtual mass of a rotating sphere in quiescent fluid.

Theoretical Analysis

Consider a rotating sphere (solid or fluid) in a mass of fluid of density ρ. Let Ω be the speed of rotation common to both the fluid particles and the sphere. Consider the fluid to be stationary at infinity. The flow field generated by the rotation of the sphere is given as follows in spherical coordinates with the origin at the center of the sphere (Lugt [9])

$$U_r = U_\theta = 0$$
$$U_\phi = \Omega \frac{a^3}{r^2} \sin\theta \tag{3}$$

The kinetic energy of the fluid round the sphere is given by the following equation that has been used previously by the author [10]:

$$K = \int_V \frac{1}{2}\rho U_\phi^2 dV \tag{4}$$

where V denotes the volume occupied by the fluid. The volume dV of an elementary annular region about the line $\theta=0$ through the sphere is given by the product of its perimeter ($2\pi r \sin\theta$), width ($r d\theta$) and depth dr. Thus

$$K = \int_{\theta=0}^{\pi} \int_{r=a}^{\infty} \frac{1}{2}\rho U_\phi^2 2\pi r \sin\theta r d\theta dr \tag{5}$$

Substituting Eq. (3) into Eq. (5) and remembering that $\int_0^\pi \sin^3\theta d\theta = 4/3$ we get

$$K = (4\pi/3)\Omega^2 \rho a^5 \tag{6}$$

The basic physical principles state that the kinetic energy of a rotating sphere with the same speed Ω is given by the following:

$$K = (1/2) I \Omega^2 \tag{7}$$

where I is the moment of inertia of the sphere which is equal to $(2/5)Ma^2$ and M is the mass of the sphere, hence Eq. (4) becomes

$$K = (1/5) M a^2 \Omega^2 \tag{8}$$

Clearly, from Eqs. (6) and (8) we get

$$M = 5\left(\frac{4}{3}\pi\rho a^3\right) \tag{9}$$

which we note is precisely five times the mass of the fluid displaced by the sphere. This mass is the virtual mass of the rotating sphere. It is the effective mass of the fluid that must be added to that of the rotating sphere in applying the equation of motion. The virtual mass coefficient could be defined according to Cheng, Drew, and Lahey [11] by the following:

$$C_v = \frac{\text{Volume of ''virtual mass''}}{\text{Volume of fluid displaced by the sphere}} \tag{10}$$

Accordingly, the virtual mass coefficient of the rotating sphere becomes

$$C_v = 5 \tag{11}$$

To the best of our knowledge this is a result that has not been reported before. The range of applicability of this equation is that the Reynolds number $\mathrm{Re}<10$ where $\mathrm{Re}=(2a)^2\Omega/2\upsilon$. This restriction is dictated by the flow field of Eq. (3). Unfortunately, there are neither experimental data nor theoretical solutions to validate the present result.

It is well known that initial acceleration of a spherical gas bubble in liquid is equal to 2 g where g is the acceleration due to gravity [12]. This arises from the value of the virtual mass coefficient of the nonrotating sphere which is equal to half. Thus the initial acceleration of the bubble that has been introduced into a purely vortical flow, would be (1/5)g. This brand of research has been presented by Sirdhar and Katz [13].

The contribution of the virtual mass to the drag, lift, and history forces is given by the following intergrodifferential equation of

Contributed by the Applied Mechanics Division of The American Society of Mechanical Engineers for publication in the ASME Journal of Applied Mechanics. Manuscript received by the ASME Applied Mechanics Division, September 3, 2004; final revision, February 5, 2005. Associate Editor: B. A. Younis.

 Copyright © 2005 by ASME

momentum for the solid rotating sphere of density ρ_s accelerating with a velocity U in a quiescent fluid of viscosity v and density ρ

$$(V\rho_s + C_v m)\frac{dU}{dt} = V(\rho_s - \rho)g - \frac{1}{2}C_D\rho U^2 S + \frac{1}{2}C_L\rho U^2 S - \frac{9m}{2a}\sqrt{\frac{v}{\pi}}\int_0^t \frac{dU}{d\tau}\frac{d\tau}{\sqrt{t-\tau}} \quad (12)$$

where $V=4/3\pi a^3$, $m=V\rho$, $S=\pi a^2$, $C_D=$ drag coefficient, and $C_L=$ lift coefficient. The first and the fourth terms on the right side of Eq. (12) represent the buoyancy and the history forces, respectively.

References

[1] Davis, A. M. J., 1977, "High Frequency Limiting Virtual Mass Coefficients of Heaving Half-Immersed Spheres," J. Fluid Mech., **80**, pp. 305–319.

[2] Stokes, G. G., 1851, *Mathematical and Physical Papers*, 3, Johnson Reprint Corp., NY, p. 34.

[3] Drew, D. A., and Lahey, Jr., R. T., 1987, "The Virtual Mass and Lift Force on a Sphere in Rotating and Straining Inviscid Fluid," Int. J. Multiphase Flow, **13**, pp. 113–121.

[4] Chan, K. W., Baird, M. H. I., and Round, G. F., 1974, "Motion of a Solid Sphere in a Horizontally Oscillating Liquid," Chem. Eng. Sci., **29**, pp. 1585–1592.

[5] Kurose, R., and Komori, S., 1999, "Drag and Lift Forces on a Rotating Sphere in a Linear Shear Flow," J. Fluid Mech., **384**, pp. 183–206.

[6] Milne-Thomson, L. M., 1972, *Theoretical Hydrodynamics*, 5th ed., Macmillan, Glasgow, U.K.

[7] Bagchi, P., and Balachandar, S., 2003, "Inertial and Viscous Forces on a Rigid Sphere in Straining Flows at Moderate Reynolds Numbers," J. Fluid Mech., **481**, pp. 105–148.

[8] Magnaudet, J., Rivero, M., and Fabre, J., 1995, "Accelerated Flows Past a Rigid Sphere or a Spherical Bubble. Part 1. Steady Straining Flow," J. Fluid Mech., **284**, pp. 97–135.

[9] Lugt, H. J., 1983, *Vortex Flow in Nature and Technology*, Wiley, NY.

[10] Kendoush, A. A., 2003, "The Virtual Mass of a Spherical-Cap Bubble" Phys. Fluids, **15**(9), pp. 2782–2785; Erratum, 2004, **16**(7), p. 2713.

[11] Cheng, L. Y., Drew, D. A., and Lahey, Jr., R. T., 1978, "Virtual Mass Effects in Two-Phase Flow," Report No. NUREG/CR-0020, Division of Reactor Safety Research, Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission.

[12] Batchelor, G. K., 1967, *An Introduction to Fluid Dynamics*, Cambridge University Press, Cambridge, England, p. 454.

[13] Sirdhar, G., and Katz, J., 1995, "Drag and Lift Forces on Microscopic Bubbles Entrained by a Vortex," Phys. Fluids, **7**(2), pp. 389–399.

Buckling of Shallow Spherical Caps Subjected to External Pressure

J. Btachut
Department of Engineering,
The University of Liverpool,
Liverpool L69 3GH, UK

This contribution details buckling tests and numerical results for shallow spherical caps subjected to static and uniform external pressure. Six mild steel caps were carefully CNC-machined from a solid billet. Three caps were designed to fail elastically with the remaining three failing within the plastic range. Caps' shallowness parameter, λ, varied from 3.5 to 5.5, and radius-to-thickness ratio varied from 300 to 1800. [DOI: 10.1115/1.1993667]

1 Background

Buckling of spherical caps subjected to uniform and static external pressure has been researched for decades. References [1–3] list most of the relevant work spanning a century or so. When reviewing the experimental work on buckling of spherical caps it becomes transparent that only small number of tests was carried out on metallic caps. One specific set of seventeen tests is reported in Ref. [4]. They were carried out on models with the base diameter between 20 mm and 50 mm. All these models were machined from a thick aluminum alloy plate. A drop in the buckling strength was confirmed in these tests for small magnitude of caps' geometrical parameter lambda, i.e., for $\lambda \approx 4.0$. A number of subsequent tests did not follow the same trend. Various reasons were given in order to explain this scatter of results around $\lambda = 4.0$. By-and-large there appears to be no definitive answer to this dilemma. This limited experimental and numerical study re-examines spherical caps' static stability around $\lambda = 4.0$. The base diameter of tested caps is about 176 mm, i.e., much larger than reported in Ref. [4].

2 Geometry and Modeling Details

Let us consider a spherical cap of radius, R, base radius, a, and constant wall thickness, t,—as sketched in Fig. 1. The shell is to be clamped at the base and subjected to static, uniform external pressure, p. The issue of boundary conditions, at the cap's base, has been the subject of wide discussions due to their effect on the load carrying capacity. In the current study it has been decided to have an integral base ring of 'substantial dimensions' attached to each model—as sketched in Fig. 2. During testing the base ring was fixed to test vessel's heavy base by an additional fixing collar rather then directly by bolts. The use of collar secured an evenly distributed clamping force. Numerical simulations have shown that clamping caps directly by bolts causes propagation of stresses directly into the caps' wall. The use of additional collar prevents this from occurring. Details about this arrangement are depicted later.

Material of the caps was mild steel which remained linear-elastic up to the yield, followed by Lüders bands for about 2.0%. Mechanical properties of mild steel were the same as reported in Ref. [5]. That paper should be consulted for further details. Average mechanical properties of the mild steel were found from tests on round and flat specimens and they were Young's modulus, $E = 207.0$ GPa, yield point of material, $\sigma_{yp} = 303.5$ MPa, and Poisson's ratio, $\nu = 0.28$.

3 Experimental Details

A series of six cups, designated here as D1, …, D6, were CNC-machined from 245 mm diameter mild steel billet. The outer diameter of the base flange was machined down to 206 mm in order to accommodate the caps in the existing test vessel. Shells had a heavy edge ring (20 mm thick and integral with the wall). Its role was to model the fully clamped boundary conditions. The shallowness parameter, λ,

$$\lambda \equiv 2[3(1-\nu^2)]^{1/4}(H/t)^{1/2} \tag{1}$$

was chosen to be between 3.5 and 5.5. The sequence of manufacturing for caps D1, …, D6 was the same as described in Ref. [5]. It is worth noting only that after final machining caps were stress-relieved in a vacuum furnace. Next, shape and wall thickness were measured. Shape was measured using (XYZ)-coordinate measuring table—as shown in Fig. 3. Measurements were taken along 14 equally spaced meridians and at 10 mm arc-length intervals. The wall thickness was measured at the same grid as shape using a customary "two-ball-bearing" arrangement. Table 1 contains average values for geometry (columns 2–4). It is seen from Table 1 that the height-to-wall thickness ratio varied approximately from 2.5 to 4.0 and the radius-to-thickness ratio varied from 300 to 1800. Examining the scatter of wall thickness, and radial deviations from perfect geometry it was concluded that all caps were, geometrically, near-perfect. As mentioned earlier caps were fixed to test vessel's base plate by means of an additional collar and by a set of eight bolts—as shown in Fig. 4(*b*). The pressure tightness was secured by an "O-ring". The inner surface of caps was open to atmosphere, and the pressurization medium was oil.

The above models were buckled through the single, incremental application of quasi-static external pressure in a small pressure vessel—see Fig. 4. Loading was carried out by a hand operated pump and magnitude of pressure increments was 0.25% of anticipated buckling load. A time break of 60 s between two load steps was maintained during all tests. All caps failed suddenly through a snap-through mechanism resulting in a large, and axisymmetric, indent at the apex. After unloading there was no snap-back in any of tested caps. Typical view of a collapsed cap is depicted in Fig. 5. Values of experimental buckling pressures are given in Table 1 (column 5).

4 Discussion of Experimental and Numerical Results

In view of near-perfect shape of caps it was decided to employ axisymmetric modeling for prediction of buckling loads. Two pieces of existing software were used to obtain buckling loads. BOSOR5 [6] provided the magnitudes of asymmetric bifurcation and snap-through loads while ABAQUS [7] was used to check the magnitude of the snap-through. Predictions of the latter two loads by both codes were identical. In the analyses geometrical and material nonlinearity were used. In the latter, elastic perfectly-plastic modeling of the stress-strain curve was adopted. Table 2 summarizes experimental and numerical results which are normalized by linear classical buckling pressure, p_{cl}, i.e.,

$$p_{cl} \equiv 2[3(1-\nu^2)]^{-1/2}E(t/R)^2. \tag{2}$$

Comparison of experimental failure pressures with numerical predictions, as detailed in Table 2, is found to be good (with the ratio of p_{expt}/p_{num} varying from 0.92 to 1.04). Figure 6 depicts position of test values, for D1, …, D6, on a plot of dimensionless buckling pressure versus the shallowness parameter, λ.

The values of the slenderness parameter, $\bar{\lambda}$, given by

Contributed by the Applied Mechanics Division of The American Society of Mechanical Engineers for publication in the ASME Journal of Applied Mechanics. Manuscript received by the ASME Applied Mechanics Division, September 14, 2004; final revision, February 18, 2005. Associate Editor: N. Triantafyllides.

 Copyright © 2005 by ASME

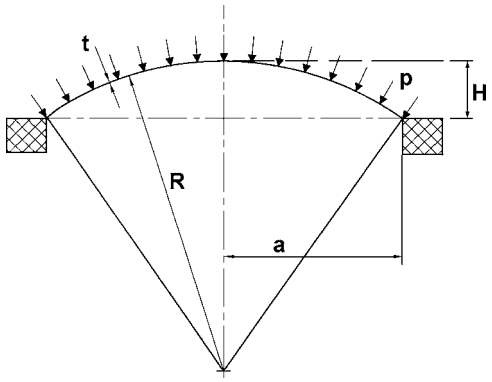


Fig. 1 Geometry of spherical cap

$$\bar{\lambda} = [3(1-\nu^2)]^{1/4}\sqrt{\frac{\sigma_{yp}}{E}}\sqrt{\frac{R}{t}} \tag{3}$$

are provided in column two of Table 2. It is seen from Fig. 6 that for $3.2 < \lambda \approx < 5.2$ the load carrying capacity of considered caps is determined by axisymmetric collapse irrespective whether this is elastic or elastic-plastic collapse. Results shown in Fig. 6 indicate that indeed there is a minimum of the load carrying capacity

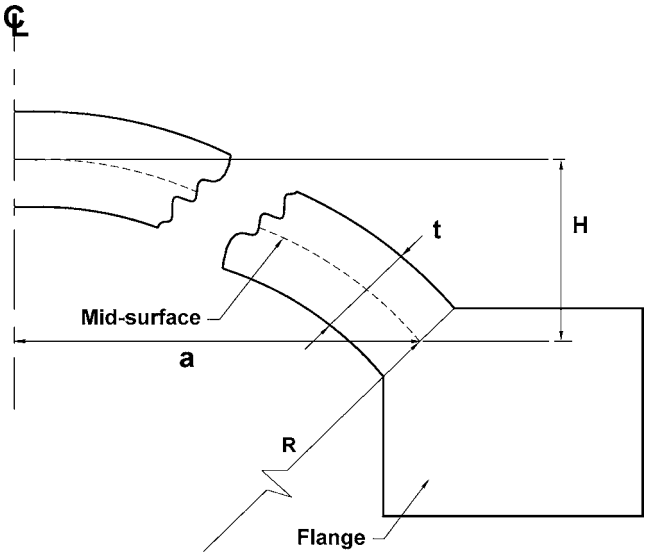


Fig. 2 Joining arrangements between spherical shell and integral base flange

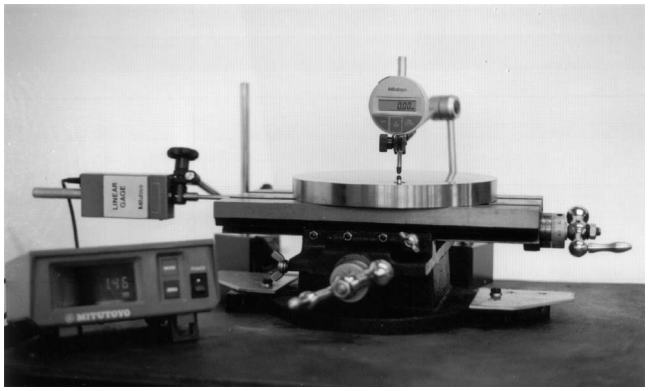


Fig. 3 Shape measuring table

Table 1 Average experimental dimensions (mid-surface where appropriate). Also, experimental buckling pressures.

Cap	2a (mm)	H (mm)	t (mm)	R (mm)	p_{expt} (MPa)
D1	166.12	1.90	1.0	1816.5	0.0458
D2	166.13	2.15	1.02	1605.7	0.0523
D3	166.10	3.93	1.03	878.6	0.211
D4	166.12	2.96	1.76	1166.7	0.330
D5	166.10	4.56	1.76	759.3	0.650
D6	166.18	6.20	1.76	563.4	1.172

for $\lambda \approx 4.0$, again irrespective whether it is elastic or elastic-plastic case. The transition from axisymmetric collapse to asymmetric bifurcation mode appears to be nonsmooth, i.e., of the cusp-type. The load-deflection curves for all six caps are depicted in Fig. 7. Purely elastic behavior, up to the collapse, is noted for

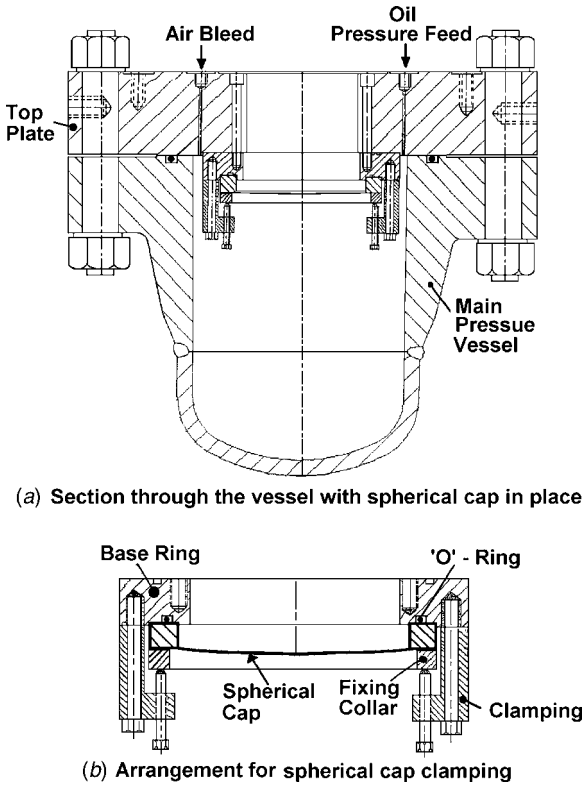


Fig. 4 Section through the pressure vessel showing test arrangements (Fig. 4(*a*)). Also, details of a cap fixing to the base (Fig. 4(*b*)).

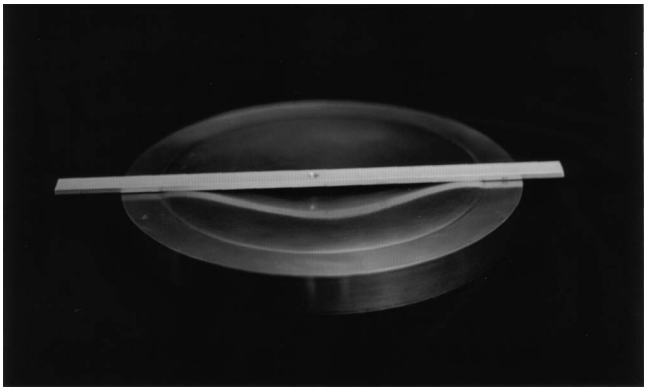


Fig. 5 Photograph of buckled spherical cap

Table 2 Comparison of experimental buckling pressure, p_{expt}, with classical, p_{cl}, and numerical, p_{num}, predictions. Also, the first yield pressures, p_{yp}, and elastic collapse pressures, p_{el}.

Cap	$\bar{\lambda}$	λ	p_{yp} (MPa)	p_{el} (MPa)	p_{cl} (MPa)	p_{expt}/p_{cl}	p_{num}/p_{cl}	p_{expt}/p_{num}
D1	2.104	3.56	--	0.0440	0.0755	0.61	0.59	1.04
D2	1.959	3.74	--	0.0571	0.101	0.52	0.57	0.92
D3	1.442	5.04	--	0.210	0.342	0.62	0.61	1.01
D4	1.271	3.35	0.308	0.358	0.567	0.58	0.59	0.98
D5	1.026	4.11	0.613	0.744	1.338	0.49	0.450	0.98
D6	0.883	4.84	0.929	1.432	2.430	0.48	0.49	0.99

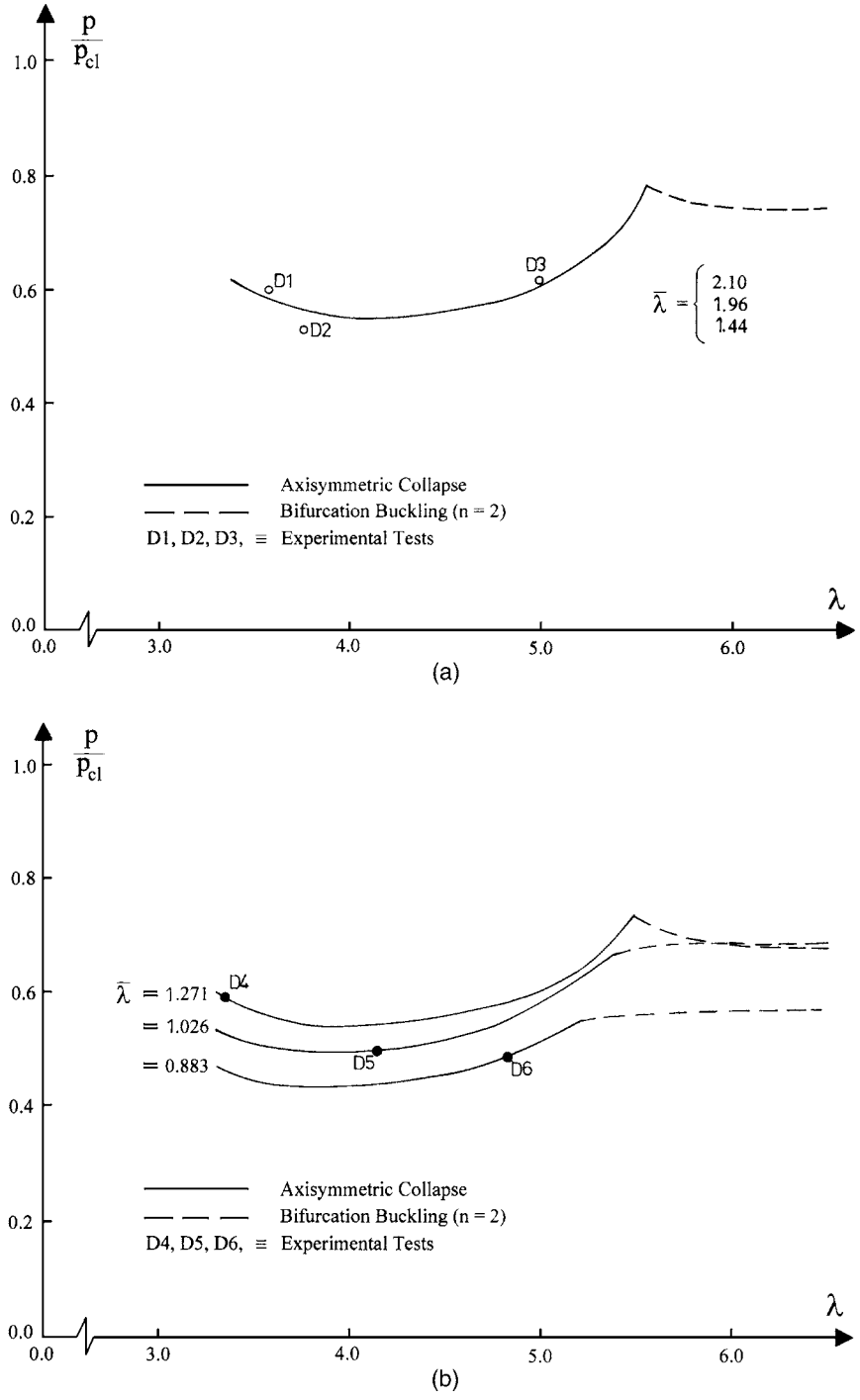


Fig. 6 Plot of buckling load vs shallowness parameter, λ, for six values of slenderness, $\bar{\lambda}$

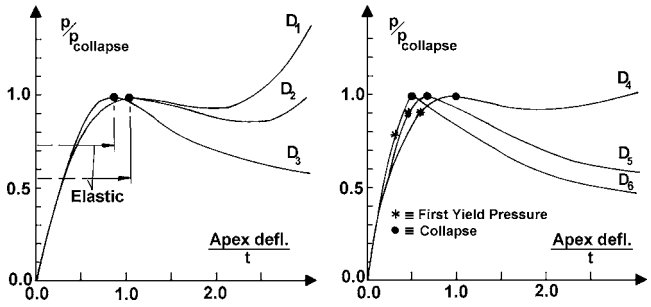


Fig. 7 Plots of external pressure versus apex deflection. Note elastic behavior of models D_1, D_2, and D_3 at the collapse; and magnitudes of first yield pressures for D_4, D_5, and D_6.

D1, D2, and D3 models. Plastic straining occurs below the collapse in the remaining models. Magnitudes of the corresponding first yield pressures can be found in Table 2.

5 Conclusions

Results of this study confirm that the magnitude of axisymmetric collapse pressure, of near-perfect spherical caps, reaches minimum at shallowness parameter $\lambda \approx 4.0$. This observation remains true for both elastic and elastic-plastic behavior. In the case of elastic behavior, the axisymmetric collapse remains the controlling mode of failure for the shallowness parameter, $\lambda < 5.5$. For larger values of lambda the asymmetric bifurcation mode controls the failure. In the case of elastic-plastic behavior the transition from bifurcation to axisymmetric collapse mode remains dependent on the value of slenderness parameter, $\bar{\lambda}$, and it moves towards smaller values of λ as the slenderness becomes smaller. Finally, it is worth noting that the minimum of buckling strength around $\lambda = 4.0$ is not a sharp one.

References

[1] Singer, J., Arbocz, J., and Weller, T., 2002, *Buckling Experiments-Experimental Methods in Buckling on Thin-Walled Structures*, Vol. 2, John Wiley & Sons, NY, Chap. 9.

[2] Bushnell, D., 1985, *Computerized Buckling Analysis of Shells*, Martinus Nijhoff Publishers, Dordrecht.

[3] Uchiyama, M., and Yamada, S., 2003, "Nonlinear buckling simulation of imperfect shell domes by mixed finite elements," J. Eng. Mech., **129**, pp. 707–714.

[4] Krenzke, M. A., and Kiernan, T. J., 1963, "Elastic stability of near-perfect shallow spherical shells," AIAA J., **1**, pp. 2855–2857.

[5] Btachut, J., Galletly, G. D., and Moreton, D. N., 1990, "Buckling of near-perfect steel torispherical and hemispherical shells subjected to external pressure," AIAA J., **28**, pp. 1971–1975.

[6] Bushnell, D., 1976, "BOSOR5-program for buckling of elastic-plastic complex shells of revolution including large deflections and creep," Comput. Struct., **6**, pp. 221–239.

[7] Hibbitt, Karlsson and Sorensen Inc., *ABAQUS User Manual, Ver. 5.8*, Pawtucket, RI 02860-4847, USA.