Journal of Applied Mechanics

Published Bimonthly by ASME

VOLUME 71 • NUMBER 6 • NOVEMBER 2004

TECHNICAL PAPERS

BRIEF NOTES

(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.

Editor
ROBERT M. McMEEKING

Assistant to the Editor
LIZ MONTANA

APPLIED MECHANICS DIVISION

Executive Committee
(Chair) **M. C. BOYCE**
W.-K. LIU
T. N. FARRIS
K. RAVI-CHANDAR

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, **HARRY ARMEN**
Executive Director, **V. R. CARTER**
Treasurer, **R. E. NICKELL**

PUBLISHING STAFF
Managing Director, Engineering
THOMAS G. LOUGHLIN
Director, Technical 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.) The American Society of Mechanical Engineers, Three Park Avenue, New York, NY 10016. Periodicals postage paid at New York, NY and additional mailing office. 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.

INDEXED by Applied Mechanics Reviews and Engineering Information, Inc. Canadian Goods & Services Tax Registration #126148048. ded Piezoelectric Material Strip

(**Contents continued**)

Journal of Applied Mechanics **Volume 71, Number 6** **NOVEMBER 2004**

ANNOUNCEMENTS AND SPECIAL NOTICES

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

Dynamic Instability of an Elastic Disk Under the Action of a Rotating Friction Couple

H. Ouyang

J. E. Mottershead

Department of Engineering,
University of Liverpool,
Liverpool L69 3GH, UK

This paper investigates the instability of the transverse vibration of a disk excited by two corotating sliders on either side of the disk. Each slider is a mass-spring-damper system traveling at the same constant speed around the disk. There are friction forces acting in the plane of the disk at the contact interfaces between the disk and each of the two sliders. The equation of motion of the disk is established by taking into account the bending couple acting in the circumferential direction produced by the different friction forces on the two sides of the disk. The normal forces and the friction couples produced by the rotating sliders are moving loads and are seen to bring about dynamic instability. Regions of instability for parameters of interest are obtained by the method of state space. It is found that the moving loads produced by the sliders are a mechanism for generating unstable parametric resonances in the subcritical speed range. The existence of stable regions in the parameter space of the simulated example suggests that the disk vibration can be suppressed by suitable assignment of the parameter values of the sliders.

[DOI: 10.1115/1.1795815]

1 Introduction

Disks are basic mechanical components and can be found in car disk brakes, clutches, saws, computer disk drives, and many other applications. In these mechanical devices, friction usually plays an essential role. Apart from its useful purposes, the friction force acting at the moving contact interface among mechanical components can induce excessive vibration or offensive noise, such as squeal in a car disk brake. The vibration and noise emanating from a disk brake causes diskomfort to passengers and raises their concern about the safety and reliability of the vehicle. Excessive vibration of a wood saw can cause damage and poor quality of the work piece. Therefore, it is of both academic interest and practical significance to understand how unstable vibration and noise are generated in disks. The objective of this paper is to explore the friction-induced dynamic instability in a disk modeled as a flat circular plate. The friction mechanism incorporated is intended to explain the unstable vibration appearing in components in moving frictional contact, for example, the squeal in a car disk brake.

The study of the vibration of disks subject to moving loads was initiated in the 1970s. Mote [1] investigated the vibration of a disk under a pointwise moving load and showed that instability might occur in the supercritical speed range. Iwan and Moeller [2] studied the dual problem of a disk spinning past a stationary mass-spring-damper system in the supercritical speed range. Many papers have been published since. The parametric excitation of asymmetric disks was analyzed by Yu and Mote [3]. Jiang et al. [4] looked at axially moving loads acting on a disk. Shen and Mote [5] explained why a stationary disk could be destabilized by a rotating damper. Shen [6] used the method of multiple scales for analyzing different types of parametric resonances of a stationary disk excited by a rotating slider. Large damping forces were examined by Huang and Mote [7].

Ono et al. [8] introduced friction as a follower force in their spinning disk model for a computer disk drive. Friction modeled as a follower force appeared to originate from North [9] for a very simple model of a disk brake. Lee and Waas [10] studied a rotating multilayer disk subjected to a stationary frictional follower force. Chan et al. [11] investigated the parametric resonances of a stationary disk under a rotating frictional follower force. Mottershead et al. [12] extended the work to a distributed system of mass, stiffness, damping, and friction. Ouyang et al. [13] combined the frictional follower force and the negative slope of friction-velocity curve and identified some new parametric resonances. Tseng and Wickert [14] included the friction-induced in-plane stresses through theory of elasticity into the equation of motion of a spinning disk. Ouyang et al. [15] found rich dynamic behavior of a stationary disk excited by a rotating slider experiencing stick-slip vibration due to friction. Ouyang et al. [16] derived a numerical-analytical combined method for the stability analysis of a car disk brake with the pads treated as moving loads. The genetic algorithm was used to find optimal placement of stabilizers to suppress friction-induced vibration of a stationary disk [17]. Tian and Hutton [18] presented a general approach for stability analysis of disks involving various nonconservative forces acting on spinning disks. The vibration and dynamic stability of disks were reviewed by Mottershead [19].

In addition to the follower force hypothesis, the stick-slip scenario and the mechanism of the decreasing friction against relative velocity, there can be other means whereby friction results in unstable vibration. North [20] showed in his two-degrees-of-freedom model of a car disk brake that the different friction forces on either side of a vibrating disk (modeled as a rigid body) could lead to instability. Hulten and Flint [21] extended this simple idea to a much more refined model of disk brakes in which the disk and the two pads were modeled as beams.

This paper presents a study of the transverse vibration of a disk under two co-rotating sliders. Each slider, driven at the same constant speed, is a mass-spring-damper system, which may represent a pad in the case of a car disk brake. There are friction forces acting in the plane of the disk at the contact interfaces between the disk and the two sliders. The differential friction forces on the two sides of the disk produce a bending couple in the circumferential direction, which fluctuates as it moves around the vibrating disk. This investigation furthers North's idea to a more realistic model of the disk as an annular plate and, more importantly, treats the

---

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 15, 2001; final revision, April 12, 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 of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

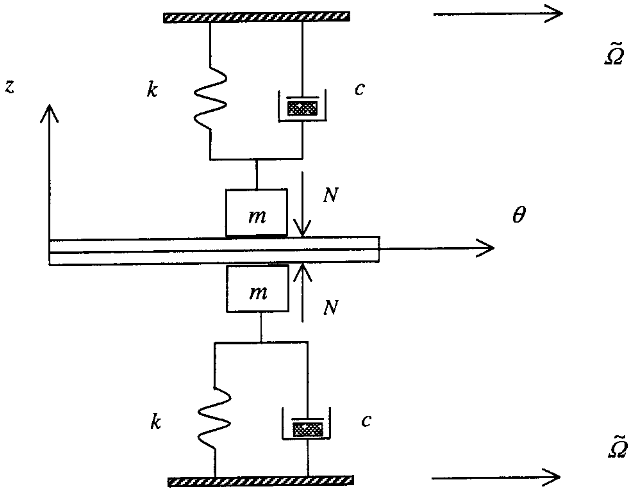


Fig. 1 Circular plate under rotating sliders on both sides

rotating frictional sliders as moving loads. Moving loads are commonly found in many mechanical and civil engineering systems. Fryba's monograph summarized many analytical solutions of simple moving load problems [22].

Friction is a very complicated phenomenon, and it constitutes a challenging research topic in its own right. In this paper, the authors make no attempt to use very sophisticated friction laws. Instead they adopt the simple Coulomb friction model, which is able to generate interesting results of instability. Furthermore, it is important that excitation by a friction couple can be justified physically, whereas the follower-force model, though it produces an effect similar to what is observed in squealing brakes, has been the subject of some debate, since it is an internal force. The reader is referred to two comprehensive review papers on the physics and modeling of friction and simulation of friction-induced vibration by Oden and Martins [23] and Ibrahim [24].

2 Theoretical Development

When a car disk brake is applied, two normal forces are applied through the pads, modeled as sliders in this paper, onto either side of the disk. These forces are initially equal in magnitude, but acting in opposite directions. As the disk rotates past the sliders and when the transverse vibration of the disk is initiated, the two normal forces from the sliders vary with time and thus cause two different friction forces acting on the top and the bottom surfaces of the disk in contact with the two sliders, as illustrated in Fig. 1. The net friction force acting in the plane of the disk plate is taken to be constant. This is simple and assumed by most researchers. However, the two different, time-varying friction forces on the two sides of the disk, which make up a non-time-varying constant friction force in the plane of the disk, also produce a bending couple, which has been neglected or omitted by most researchers. This couple produced by the two uneven friction forces was first considered by North [20] and the idea was recently extended by Hulten [25] in the context of a drum brake modeled as beams. Hulten and Flint [21] later applied the idea to a disk brake and treated the disk as a beam and the friction force and normal force were acting on the entire length of the beam. Popp and Rudolph [26] introduced the same nonconservative friction force into their 14-parameter lumped-mass model of a disk brake. The vibration of all these models was not considered as a moving load problem.

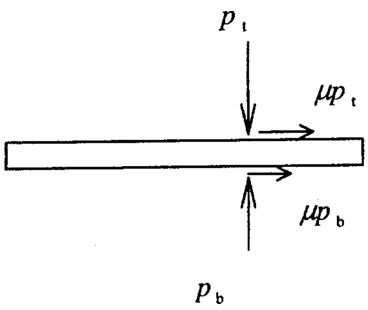


Fig. 2 Forces acting on the disk under sliders

Since car disk-brake squeal tends to appear at low road speeds when centripetal and gyroscopic effects may be omitted, the problem is treated instead as a stationary disk excited by two rotating sliders in this paper, as shown in Fig. 1.

The sliders are initially located at polar coordinates of $(r_0,0)$. As the disk vibrates, the two sliders experience a vertical displacement $u(t)$ in the z direction as [27]

$$u(t)=w(r_0,\tilde{\Omega}t,t) \tag{1}$$

which is the transverse deflection of the disk where it is in instantaneous contact with the sliders. Equation (1) implies that the sliders always maintain contact with the disk. As a result of the transverse vibration of the disk, the normal forces from the sliders acting on either side of the disk become

$$p_{\mathrm{b}}=N-m\ddot{u}-c\dot{u}-ku, \quad p_{\mathrm{t}}=N+m\ddot{u}+c\dot{u}+ku \tag{2}$$

Consequently, the friction forces illustrated in Fig. 2 on either side of the disk are

$$\begin{aligned} f_{\mathrm{b}}&=\mu p_{\mathrm{b}}=\mu(N-m\ddot{u}-c\dot{u}-ku), \\ f_{\mathrm{t}}&=\mu p_{\mathrm{t}}=\mu(N+m\ddot{u}+c\dot{u}+ku) \end{aligned} \tag{3}$$

These uneven friction forces produce a couple [20,25] in the circumferential direction of the disk. In the present model, this couple is

$$M=h(f_{\mathrm{t}}-f_{\mathrm{b}})/2=\mu h(p_{\mathrm{t}}-p_{\mathrm{b}})/2=\mu h(m\ddot{u}+c\dot{u}+ku) \tag{4}$$

which brings about a second-order singularity to the equation of motion of the disk modeled as a Kirchoff plate. It should be pointed out that both the two normal forces and the friction couple are functions of only *t*. This equation of motion in the space-fixed cylindrical coordinate system shown in Fig. 1 is

$$\begin{aligned} \rho h\frac{\partial^2 w}{\partial t^2}+D^*\frac{\partial w}{\partial t}+D\nabla^4 w&=\frac{1}{r}\left\{\left[p_{\mathrm{b}}-p_{\mathrm{t}}+\mu(p_{\mathrm{b}}+p_{\mathrm{t}})\frac{\partial w}{r\partial\theta}\right]\delta(r-r_0)\,\delta(\theta-\tilde{\Omega}t)+\frac{\partial}{r\partial\theta}[M(t)\,\delta(\theta-\tilde{\Omega}t)]\,\delta(r-r_0)\right\} \\ &=\frac{1}{r}\left\{\left[p_{\mathrm{b}}-p_{\mathrm{t}}+2\mu N\frac{\partial w}{r\partial\theta}\right]\delta(r-r_0)\,\delta(\theta-\tilde{\Omega}t)+\frac{\partial}{r\partial\theta}[M(t)\,\delta(\theta-\tilde{\Omega}t)]\,\delta(r-r_0)\right\} \end{aligned} \tag{5}$$

where the biharmonic differential operator in the cylindrical coordinate is

$$\nabla^4=\left(\frac{\partial^2}{\partial r^2}+\frac{\partial}{r\partial r}+\frac{\partial^2}{r^2\partial\theta^2}\right)^2 \tag{6}$$

Equation (5) describes a circular thin plate under moving concentrated loads manifested by the δ functions in the θ direction. The third term in the first square bracket on the right-hand side of Eq. (5) represents the in-plane friction as a follower force.

The transverse vibration of the disk can be expressed by summation of its modes and modal coordinates as

$$w(r,\theta,t)=\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}\psi_{mn}(r,\theta)q_{mn}(t) \tag{7}$$

where the mode shape functions are

$$\psi_{mn}(r,\theta)=R_{mn}(r)\exp(\mathrm{i}n\theta)/(\rho h b^2)^{1/2}$$
$$(m=0,1,2,\ldots;n=0,-1,1,-2,2,\ldots) \tag{8}$$

which satisfy the ortho-normality conditions,

$$\int_a^b \rho h\overline{\psi_{kl}}\psi_{mn}r\mathrm{d}r\mathrm{d}\theta=\delta_{km}\delta_{ln}$$
$$\int_a^b D\overline{\psi_{kl}}\nabla^4\psi_{mn}r\mathrm{d}r\mathrm{d}\theta=\omega_{mn}^2\delta_{km}\delta_{ln} \tag{9}$$

where the bar over a symbol denotes the complex conjugation.

Multiplying Eq. (5) with $\overline{\psi_{kl}}(r,\theta)$, integrating the resultant equation yields

$$\int\int\overline{\psi_{kl}}\left[\rho h\frac{\partial^2 w}{\partial t^2}+D^*\frac{\partial w}{\partial t}+D\nabla^4 w\right]r\mathrm{d}\theta\mathrm{d}r$$
$$=\left[\overline{\psi_{kl}}\left(p_{\mathrm{b}}-p_{\mathrm{t}}+2\mu N\frac{\partial w}{r\partial\theta}\right)\right]_{r=r_0,\theta=\tilde{\Omega}t}$$
$$+\frac{\mu h}{2r_0}\left[\frac{\partial\overline{\psi_{kl}}}{\partial\theta}(p_{\mathrm{b}}-p_{\mathrm{t}})\right]_{r=r_0,\theta=\tilde{\Omega}t} \tag{10}$$

Substituting Eqs. (2), (3), (7), and (8) into (10) and making use of Eq. (9), the equation of motion of the disk in terms of the modal coordinates is derived as

$$\ddot{q}_{kl}+2\xi_{\mathrm{v}}\omega_{kl}\dot{q}_{kl}+\omega_{kl}^2q_{kl}=-\frac{2}{\rho h b^2}\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}R_{mn}(r_0)R_{kl}(r_0)\exp[\mathrm{i}(n-l)\tilde{\Omega}t]\left\{\left(1-\frac{\mu h}{2r_0}\mathrm{i}l\right)[m(\ddot{q}_{mn}+2\mathrm{i}n\tilde{\Omega}\dot{q}_{mn}-n^2\tilde{\Omega}^2q_{mn})+c(\dot{q}_{mn}+\mathrm{i}n\tilde{\Omega}q_{mn})+kq_{mn}]+\frac{\mathrm{i}n\mu N}{r_0}q_{mn}\right\}\quad(k=0,1,2,\ldots;l=0,-1,1,-2,2,\ldots) \tag{11}$$

Equation (11) indicates that the rotating sliders introduce time-dependent mass, stiffness, and damping into the disk. The last term in the curly bracket on the right-hand side of Eq. (11) is for the constant in-plane friction modeled as a follower force. In the case of a disk brake, $\mu N/r_0$ is many times smaller than k. In addition, the dynamic instability generated by this term has been studied in detail by Chan et al. [11]. For these reasons, this term is omitted in the subsequent derivation and the final calculation. In other applications, computer disk drives for example, this term can be relatively large because k and h are both small.

The critical speed for the relative rotation of the disk is defined as

$$\tilde{\Omega}_{\mathrm{cr}}=\min_{l}(\omega_{kl}/l)\quad(l=1,2,\ldots) \tag{12}$$

Nondimensional variables are defined as

$$\tau=\tilde{\Omega}_{\mathrm{cr}}t,\quad\beta_{kl}=\omega_{kl}/\tilde{\Omega}_{\mathrm{cr}},\quad\Omega=\tilde{\Omega}/\tilde{\Omega}_{\mathrm{cr}} \tag{13}$$

Introduce a scaling parameter ε such that

$$\varepsilon\gamma=2m/(\rho h b^2),\quad\varepsilon\varsigma=2c/(\rho h b^2\tilde{\Omega}_{\mathrm{cr}}),$$
$$\varepsilon\kappa=2k/(\rho h b^2\tilde{\Omega}_{\mathrm{cr}}^2),\quad\varepsilon\xi=\xi_{\mathrm{v}}\tilde{\Omega}_{\mathrm{cr}} \tag{14}$$

Thus Eq. (11) becomes

$$\ddot{q}_{kl}+2\varepsilon\xi\beta_{kl}\dot{q}_{kl}+\beta_{kl}^2q_{kl}$$
$$=-\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}R_{mn}(r_0)R_{kl}(r_0)$$
$$\times\exp[\mathrm{i}(n-l)\Omega\tau][1-\mathrm{i}l\mu h/(2r_0)]$$
$$\times[\varepsilon\gamma(\ddot{q}_{mn}+2\mathrm{i}n\Omega\dot{q}_{mn}-n^2\Omega^2q_{mn})$$
$$+\varepsilon\varsigma(\dot{q}_{mn}+\mathrm{i}n\Omega q_{mn})+\varepsilon\kappa q_{mn}]$$
$$(k=0,1,2,\ldots;l=0,-1,1,-2,2,\ldots) \tag{15}$$

where the time derivative is now with respect to the new time variable τ.

It is impossible to get closed-form solutions of Eq. (15). So it is usually solved by numerical methods, such as the method of state space for general cases or by perturbation methods when ε is small.

3 Numerical Solutions by the Method of State Space

To use the method of state space, new variables are introduced as

$$x_{kl}(\tau)=\exp(\mathrm{i}l\Omega\tau)q_{kl}(\tau)\quad(k=0,1,2,\ldots;l=0,-1,1,-2,2,\ldots) \tag{16}$$

Equation (15) then becomes

$$\ddot{x}_{kl}+2(\varepsilon\xi\beta_{kl}-\mathrm{i}l\Omega)\dot{x}_{kl}+(\beta_{kl}^2-\mathrm{i}2\varepsilon\xi\beta_{kl}l\Omega-l^2\Omega^2)x_{kl}$$
$$=-\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}R_{mn}(r_0)R_{kl}(r_0)[1-\mathrm{i}l\mu h/(2r_0)](\varepsilon\gamma\ddot{x}_{mn}$$
$$+\varepsilon\varsigma\dot{x}_{mn}+\varepsilon\kappa q_{mn})$$
$$(k=0,1,2,\ldots;l=0,-1,1,-2,2,\ldots) \tag{17}$$

Table 1 Natural frequencies and nondimensional frequencies of the disk

Modes (k,l)	0, 0	0, ±1	0, ±2	0, ±3	0, ±4	0, ±5	0, ±6
ω_{kl} (rad/s)	7203	4678	7532	16,391	26,535	37,111	48,080
β_{kl}	1.91	1.24	2	4.35	7.05	9.85	12.77

which no longer has time-dependent coefficients. Equation (17) can be rewritten in matrix form as

$$(\mathbf{I}+\varepsilon\gamma\mathbf{Q})\ddot{\mathbf{x}}+(\mathbf{A}+\varepsilon\varsigma\mathbf{Q})\dot{\mathbf{x}}+(\mathbf{B}+\varepsilon\kappa\mathbf{Q})\mathbf{x}=0 \tag{18}$$

where

$$\mathbf{A}=2\,\mathrm{diag}[\varepsilon\xi\beta_{kl}-\mathrm{i}l\Omega],\quad \mathbf{B}=\mathrm{diag}[\beta_{kl}^2-\mathrm{i}2\varepsilon\xi\beta_{kl}l\Omega-l^2\Omega^2]$$
$$\mathbf{x}^T=\{x_{kl}\}\quad (k=0,1,2,\ldots;l=0,-1,1,-2,2,\ldots) \tag{19}$$

and the element of matrix $\mathbf{Q}$ on the lth row and the nth column is

$$Q_{ln}=R_{kl}(r_0)R_{mn}(r_0)[1-\mathrm{i}l\mu h/(2r_0)]$$
$$(k,m=0,1,2,\ldots;l,n=0,-1,1,-2,2,\ldots) \tag{20}$$

When a new vector is introduced as $\mathbf{y}=\dot{\mathbf{x}}$, Eq. (21) can be derived from (18) as

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\begin{Bmatrix}\mathbf{x}\\ \mathbf{y}\end{Bmatrix}=\begin{bmatrix}0 & \mathbf{I}\\ -(\mathbf{I}+\varepsilon\gamma\mathbf{Q})^{-1}(\mathbf{B}+\varepsilon\kappa\mathbf{Q}) & -(\mathbf{I}+\varepsilon\gamma\mathbf{Q})^{-1}(\mathbf{A}+\varepsilon\varsigma\mathbf{Q})\end{bmatrix}\times\begin{Bmatrix}\mathbf{x}\\ \mathbf{y}\end{Bmatrix} \tag{21}$$

If the real part of an eigenvalue of the matrix in Eq. (21) is positive for some parameter values, the whole system is unstable. Otherwise, the system is stable. Those parameter values that lead to zero real part of at least one eigenvalue and negative real parts for all the rest of the eigenvalues form the boundaries of the regions of instability, which are of particular interest to the authors and the designers. Equation (21) provides numerical solutions for general cases. Being numerical, the state-space formulation does not normally lead to any definite conclusions about the effects of a parameter before a computation is made.

If ε is small, the method of multiple scales [28] can be used to derive approximate analytic solutions. The parametric resonances in the subcritical speed range may appear as [11]

$$(n\pm l)\Omega=\beta_{mn}-\beta_{kl}+\varepsilon\sigma\quad (n>l,l\geqslant 0;k,m=0,1,2,\ldots) \tag{22}$$

when the method of multiple scales is used. The detailed formulation is not presented here.

4 Simulated Example and Analysis

In the subsequent parametric analysis, one of the five parameters $\varepsilon\gamma$, $\varepsilon\varsigma$, $\varepsilon\kappa$, $\varepsilon\xi$, and μ is varied while other parameters are kept constant. The regions of instability dependent on the varying parameter versus Ω are found. When each one of these parameters is studied in this way, their roles in the instability of the whole system are established.

To provide a detailed picture of the effects of the parameters of interest in terms of the regions of instability, an example is analyzed. The dimensions and properties of a real brake disk made of gray cast iron are, $a=0.0805$ m, $b=0.1305$ m, $h=0.02$ m, $E=120$ GPa, $\nu=0.211$, $\rho=7200$ kg m^{-3}. The plate model of the disk is tuned to fit the numerical frequencies and modes of a very detailed finite element model of the disk. The first 13 (one single and six double) frequencies of the disk are computed by an analytical method and are listed in Table 1 ($\tilde{\Omega}_{\mathrm{cr}}=3766$ rad·s^{-1}).

The rotating sliders are located at $r_0=0.11$ m. The damping coefficient of the sliders is taken to be 2.5%. The friction coefficient is taken to be a linear function of the relative velocity between the disk and the sliders with negative slope, as $\mu=\mu_s-0.0576\tilde{\Omega}r_0$ and $\mu=0.3$ if the formula produces $\mu<0.3$. The value of negative slope is taken from measured data [26].

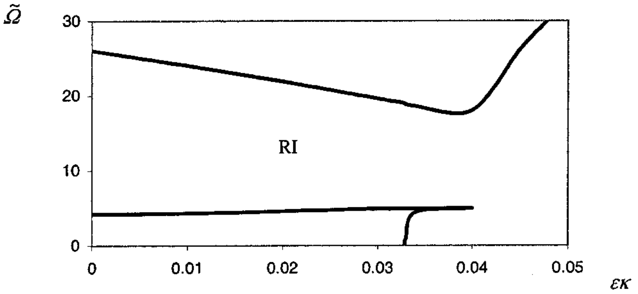


Fig. 3 Region of instability ($\mu_s=0.55$, $\varepsilon\xi=0.0001$, $\varepsilon\varsigma=0$, $\varepsilon\gamma=0.002$)

The vertical spring constant of the sliders, which can represent the aggregate result of the stiffness of the pads, caliper and brake fluid, is a very interesting parameter. As a first example, take $\varepsilon\gamma=0.002$, $\varepsilon\xi=10^{-4}$, $\varepsilon\varsigma=0$, $\mu_s=0.55$. The region of instability is presented in Fig. 3. The range of rotating speed is limited to 30 rad/s since squeal tends to occur at low speeds below 30 rad/s.

It spreads over the entire range of the stiffness values. However, a small increase of the disk damping can greatly reduce the region of instability, as shown in Fig. 4.

This indicates that the damping of the disk is stabilizing and is very effective in suppressing the disk vibration. Of course, the addition of damping of the slider can achieve the same effect, as shown in Fig. 5. The damping breaks the single region of instability into two smaller halves. Further increase of the damping of the slider can eliminate the first region of instability at smaller $\varepsilon\kappa$ in Fig. 5 and leads to only one region of instability like that of Fig. 4.

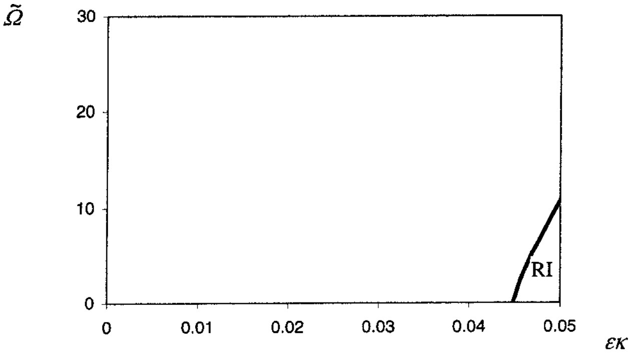


Fig. 4 Region of instability ($\mu_s=0.55$, $\varepsilon\xi=0.00014$, $\varepsilon\varsigma=0$, $\varepsilon\gamma=0.002$)

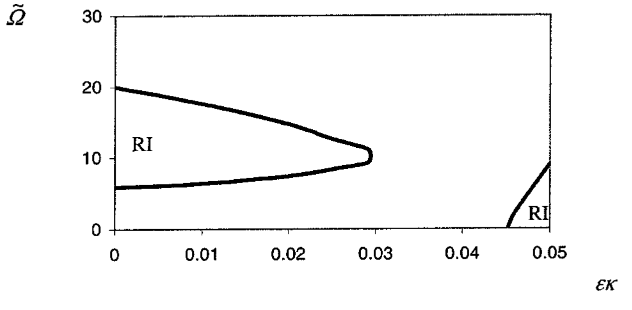


Fig. 5 Regions of instability ($\mu_s=0.55$, $\varepsilon\xi=0.0001$, $\varepsilon\varsigma=0.025$, $\varepsilon\gamma=0.002$)

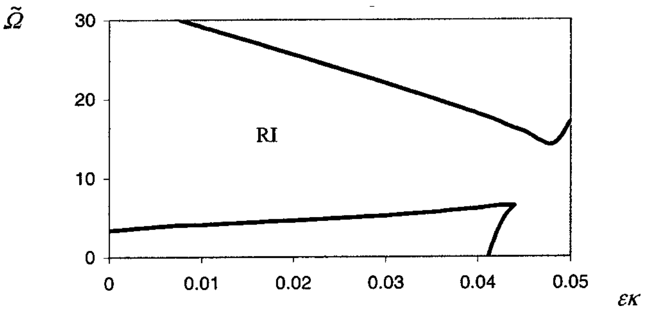


Fig. 6 Region of instability (μ_s=0.6, $\varepsilon\xi$=0.0001, $\varepsilon\varsigma$=0, $\varepsilon\gamma$=0.002)

Both figures reveal that with adequate damping from the disk or from the slider, there is a region of stability within the range of stiffness values. Higher damping can remove instability for the range of parameter values shown in these two figures.

If the friction coefficient is increased, the region of instability in Fig. 3 expands and the region of instability in Fig. 5 expands and may even reconnect to form a larger, single region of instability. An example of using $\mu_s = 0.6$ is presented in Fig. 6.

Interestingly, the influence of the mass of the sliders on the stability of the system shows a similar trend. The region of instability obtained by using $\varepsilon\xi = 10^{-4}$, $\varepsilon\varsigma = 0$, $\mu_s = 0.55$, and $\varepsilon\kappa = 0.02$ is given in Fig. 7.

If either the damping of the disk or the damping of the sliders is increased, the region of instability in Fig. 7 reduces, reminiscent of what has happened for $\varepsilon\kappa$. Figure 8 shows the reduced region of instability when the damping of the disk is increased.

By examining Figs. 3–8, it can be seen that either damping of the disk or the damping of the sliders is stabilizing. Presence of large enough damping can completely eliminate unstable vibration. Even if the damping is not great, there are stable regions for the ranges of mass parameter values and stiffness parameter values. At low speeds, small vertical stiffness of the slider is harmful. Large stiffness values are even more detrimental because the region of instability expands to the right. The same is true for the mass of the sliders. The location and size of these stable regions depend on the specific parameter values involved. This means that by suitable design of the rotating sliders the transverse vibration of the disk can be suppressed even though the dampings are not great. This offers a new way for vibration suppression of disks under a rotating frictional load, other than damping, which may not always be feasible to enhance.

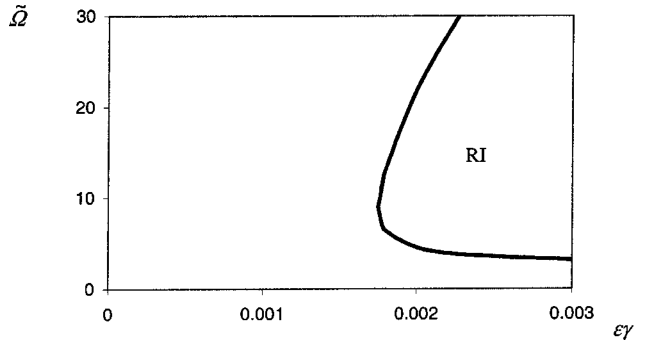


Fig. 7 Region of instability (μ_s=0.55, $\varepsilon\xi$=0.0001, $\varepsilon\varsigma$=0, $\varepsilon\kappa$=0.02)

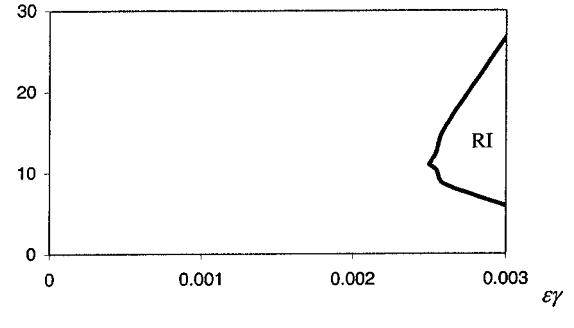


Fig. 8 Region of instability (μ_s=0.55, $\varepsilon\xi$=0.00014, $\varepsilon\varsigma$=0, $\varepsilon\kappa$=0.02)

It should be noted that the above results do not apply to high rotating speeds, where the stabilizing centripetal and gyroscopic effect needs to be considered. It should also be stressed that the instability displayed by this friction mechanism is highly speed-dependent. The consequence is that the omission of the relative rotation between the disk and the sliders (which is a common practice among researchers of brake squeal)—that is, when the vibration and squeal of a disk brake is not modeled as a moving load problem—some significant regions of instability will be missed. Figures 3–8 are obtained using the particular parameter values specified. If the damping of the disk or of the sliders is reduced or the friction coefficient is increased, the regions of instability will widen to reach even lower rotational speeds than shown in these figures.

5 Conclusions

In this paper, the instability of the transverse vibration of a circular disk under rotating load of two sliders of mass, spring, and damper is studied as a moving-load problem. The friction between the disk and the sliders is modeled as generating a fluctuating couple as the sliders move around the top and bottom surfaces of the disk. The friction coefficient is assumed to be a linear function of the relative speed between the disk and the sliders with negative gradient and then becomes a constant at sufficient relative speed. The model is meant to demonstrate a mechanism for unstable vibration of a car disk brake and similar applications. Parametrical analysis allows the following conclusions to be drawn:

1. At very low damping values, the regions of instability spread over the whole range of the stiffness parameter values. In the range of parameter considered, the unstable region tends to lie within the higher mass values.
2. At normal level of damping of the sliders or with a small increase of the disk damping, the regions of instability contract and may reveal stable regions.
3. The size and location of the regions of stability depend on the specific parameter values. This phenomenon can be exploited in designing suitable sliders to suppress the vibration of the disk.
4. The damping of the slider and the damping of the disk, in particular, are both stabilizing in the speed range studied.
5. The friction coefficient is very destabilizing.
6. The mechanism whereby friction destabilizes the vibration of the disk, incorporated in this moving load problem, is seen to be able to generate speed-dependent instability over a wide range of parameter values. The instability of the vibration of the disk is speed-dependent and, therefore, should be modeled as a moving-load problem.

Acknowledgments

The authors are grateful for the support by the Engineering and Physical Sciences Research Council (Grant No. GR/L91061), TRW Automotive, and TMD Friction UK. The suggestions for improving the manuscript by the referees are also gratefully acknowledged.

Nomenclature

a, b = inner and outer radii of the disk
c, k, m = viscous damping, stiffness, mass of a rotating slider
h = thickness of the disk
i = $\sqrt{-1}$
$p_{\mathrm{b}}, p_{\mathrm{t}}$ = the normal forces acting onto the disk from the bottom and the top sliders, respectively
$q_{kl}\,(q_{rs})$ = modal coordinate for $k(r)$ nodal circles and $l(s)$ nodal diameters for the disk
r = radial coordinate in the cylindrical coordinate system
r_0 = initial radial position of the rotating sliders
t = time
w = deflection of the disk in the cylindrical coordinate system
D = flexural rigidity
D^* = viscous damping of the disk
E = Young's modulus
M = a couple produced by the two unequal friction forces on either side of the disk
$R_{kl}\,(R_{rs})$ = combination of Bessel functions to represent the mode shape of the disk in the radial direction, corresponding to $q_{kl}\,(q_{rs})$
$\delta(\bullet)$ = Dirac delta function
δ_{kl} = Kronecker delta
μ = kinetic friction coefficient between the disk and the sliders
μ_s = static friction coefficient between the disk and the sliders
θ = circumferential coordinate in the cylindrical coordinate system
ν = Poisson ratio of the disk material
ξ_{v} = damping coefficient $(=D^*/(2\rho h\omega_{kl}))$ of the disk
ρ = mass-density of the disk
ψ_{kl} = mode shape function for the transverse vibration of the disk corresponding to q_{kl}
ω_{kl} = natural (circular) frequency corresponding to q_{kl}
$\tilde{\Omega}$ = rotating speed of the sliders in radians per second

References

[1] Mote, C. D., 1970, "Stability of Circular Plates Subjected to Moving Loads," J. Franklin Inst., **290**, pp. 329–344.
[2] Iwan, W. D., and Moeller, T. L., 1976, "The Stability of a Spinning Elastic Disc With a Transverse Load System," ASME J. Appl. Mech., **43**, pp. 485–496.
[3] Yu, R. L., and Mote, C. D., 1987, "Vibration and Parametric Excitation in Asymmetric Plates Under Moving Loads," J. Sound Vib., **119**, pp. 409–427.
[4] Jiang, Z. W., Chonan, S., and Abe, H., 1990, "Dynamic Response of a Read/Write Head Floppy Disc System Subjected to Axial Excitation," ASME J. Vibr. Acoust., **112**, pp. 53–58.
[5] Shen, Y., and Mote, C. D., 1991, "On the Mechanism of Instability of a Circular Plate Under a Rotating Spring-Mass-Dashpot System," J. Sound Vib., **148**, pp. 307–318.
[6] Shen, I. Y., 1993, "Response of a Stationary Damped Circular Plate Under a Rotating Slider Bearing System," ASME J. Vibr. Acoust., **115**, pp. 65–69.
[7] Huang, F. Y., and Mote, C. D., 1996, "Mathematical Analysis of Stability of a Spinning Disc Under Rotating, Arbitrary Large Damping Force," ASME J. Vibr. Acoust., **118**, pp. 657–662.
[8] Ono, K., Chen, J. S., and Bogy, D. B., 1991, "Stability Analysis of the Head-Disk Interface in a Flexible Disc Drive," ASME J. Appl. Mech., **58**, pp. 1005–1014.
[9] North, N. R., 1972, "Disc Brake Squeal—A Theoretical Model," M.I.R.A. Research Report, No. 1972/5.
[10] Lee, D., and Waas, A. M., 1997, "Stability Analysis of a Rotating Multi-Layer Annular Plate With a Stationary Frictional Follower Load," Int. J. Mech. Sci., **39**(10), pp. 1117–1138.
[11] Chan, S. N., Mottershead, J. E., and Cartmell, M. P., 1994, "Parametric Resonances at Subcritical Speeds in Discs With Rotating Frictional Loads," Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci., **208**, pp. 417–25.
[12] Mottershead, J. E., Ouyang, H., Cartmell, M. P., and Friswell, M. I., 1997, "Parametric Resonances in an Annular Disc, With a Rotating System of Distributed Mass and Elasticity; and the Effects of Friction and Damping," Proc. R. Soc. London, Ser. A, **453**(1), pp. 1–19.
[13] Ouyang, H., Mottershead, J. E., Cartmell, M. P., and Friswell, M. I., 1998, "Friction-Induced Parametric Resonances in Discs: Effect of a Negative Friction-Velocity Relationship," J. Sound Vib., **209**(2), pp. 251–264.
[14] Tseng, J.-G., and Wickert, J. A., 1998, "Nonconservative Stability of a Friction Loaded Disc," ASME J. Vibr. Acoust., **120**, pp. 922–929.
[15] Ouyang, H., Mottershead, J. E., Cartmell, M. P., and Brookfield, D. J., 1999, "Friction-Induced Vibration of an Elastic Slider on a Vibrating Disc," Int. J. Mech. Sci., **41**, pp. 325–336.
[16] Ouyang, H., Mottershead, J. E., Brookfield, D. J., James, S., and Cartmell, M. P., 2000, "A Methodology for the Determination of Dynamic Instabilities in a Car Disc Brake," Int. J. Veh. Des., **23**, pp. 241–262.
[17] Ouyang, H., and Mottershead, J. E., 2001, "Optimal Suppression of Parametric Vibration in Discs Under Rotating Frictional Loads," Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci., **215**(1), pp. 65–75.
[18] Tian, J., and Hutton, S. G., 1999, "Self-Excited Vibration in Flexible Rotating Disks Subjected to Various Transverse Interactive Forces: A General Approach," ASME J. Appl. Mech., **66**(3), pp. 800–805.
[19] Mottershead, J. E., 1998, "Vibration and Friction-Induced Instability in Discs," Shock Vib. Dig., **30**(1), pp. 14–31.
[20] North, N. R., 1976, "Disc Brake Squeal," Proc. Inst. Mech. Eng., **C38/76**, pp. 169–176.
[21] Hulten, J., and Flint, J., 2002, "Lining-Deformation-Induced Modal Coupling as Squeal Generator in a Distributed Parameter Disc Brake Model," J. Sound Vib., **254**(1), pp. 1–21.
[22] Fryba, L., 1972, *Vibration of Solids and Structures Under Moving Loads*, Noordhoff, Groningen.
[23] Oden, J. T., and Martins, J. A. C., 1985, "Models and Computational Methods for Dynamic Friction Phenomena," Comput. Methods Appl. Mech. Eng., **52**, pp. 527–643.
[24] Ibrahim, R. A., 1994, "Friction-Induced Vibration, Chatter, Squeal, and Chaos. Part 1: Mechanics of Friction; Part 2: Dynamics of Modeling," Appl. Mech. Rev., **47**, pp. 209–253.
[25] Hulten, J. O., 1995, "Some Drum Brake Squeal Mechanisms," *Proc. of the 1995 Noise and Vibration Conference*, Traverse City, Michigan, Vol. 1, pp. 377–388.
[26] Popp, K., and Rudolph, M., 2001, "Brake Squeal," *In Detection, Utilization and Avoidance of Nonlinear Dynamical Effects in Engineering Applications*, K. Popp (ed.) Aachen: Shaker Verlag, pp. 197–225.
[27] Iwan, W. D., and Stahl, K. J., 1973, "The Response of an Elastic Disc With a Moving Mass System," ASME J. Appl. Mech., **40**, pp. 445–451.
[28] Nayfeh, H., and Mook, D. T., 1979, *Nonlinear Oscillation*, Wiley, New York.

Stochastic Stability of Coupled Oscillators in Resonance: A Perturbation Approach

N. Sri Namachchivaya
Department of Aerospace Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801

H. J. Van Roessel
Department of Mathematical Sciences,
University of Alberta,
Edmonton, Alberta T6G 2G1, Canada

A perturbation approach is used to obtain an approximation for the moment Lyapunov exponent of two coupled oscillators with commensurable frequencies driven by a small intensity real noise with dissipation. The generator for the eigenvalue problem associated with the moment Lyapunov exponent is derived without any restriction on the size of pth moment. An orthogonal expansion for the eigenvalue problem based on the Galerkin method is used to derive the stability results in terms of spectral densities. These results can be applied to study the moment and almost-sure stability of structural and mechanical systems subjected to stochastic excitation. [DOI: 10.1115/1.1795813]

1 Introduction

This paper is concerned with stochastic stability of two-degree-of-freedom (dof) systems in the presence of *internal resonances*. In linear multi-degree-of-freedom systems, resonances occur due to commensurability of the frequencies $\omega_i(\mu)$ where μ represents the system parameters. A relationship of the form $m_1\omega_1(\mu) + \ldots + m_n\omega_n(\mu) = 0$ exists in this case for some integer values of $m_1, \ldots, m_n$. The resonance effects would in this case, persist in all regions of the phase space and is due to a particular choice of the system parameters μ. Almost-sure and moment stability of coupled oscillators with commensurable frequencies has been a difficult problem to tackle because resonance gives rise to a nontrivial approximate foliation of the phase space. As a result, the equations cannot be reduced to a lower dimension via standard stochastic approximation methods. Thus, the determination of the moment and maximal Lyapunov exponents of two coupled oscillators in resonance at present is an open problem. The challenge has been to extend the existing techniques [1–3] in order to explicitly evaluate the moment Lyapunov exponents for such systems with *commensurable frequencies*. It is this need and challenge that we shall address in this paper.

In the study of stability of solutions of random dynamical systems, the exponential growth rate of $\mathbb{E}\|x(t;x_0)\|^p$ is provided by the moment Lyapunov exponent defined as

$$g(p;x_0) = \lim_{t\to\infty}\frac{1}{t}\log \mathbb{E}\|x(t;x_0)\|^p$$

where $x(t;x_0)$ is the solution process of a linear random dynamical system. If $g(p;x_0) < 0$, then, by definition, $\mathbb{E}\|x(t;x_0)\|^p \to 0$ as $t\to\infty$ and this is referred to as *p*th moment stability. The connection between moment stability and almost-sure stability for an undamped linear oscillator under real noise excitation was established for the first time by Molčanov [4]. These results were extended for an arbitrary *d*-dimensional system by Arnold [5] where a concise formulation of the relation between almost-sure sample stability and *p*th mean stability is presented. The complete set of results on the so-called moment Lyapunov exponent, its properties, and generators is obtained in two consecutive papers by Arnold et al. [6,7] for white- and real-noise situations, respectively.

The aim of this paper is to determine an asymptotic expansion for the moment Lyapunov exponent of a noisy 2-dof system given by

$$\ddot{q}_i + \omega_i^2 q_i + 2\varepsilon^2\zeta\omega_i\dot{q}_i + \varepsilon\sum_{j=1}^{2} k_{ij}q_j f(\xi(t)) = 0, \quad i,j = 1,2 \tag{1}$$

where the q_i's are generalized coordinates, ω_i is the *i*th natural frequency, $\xi(t)$ is a stationary stochastic process, and $\varepsilon^2\zeta$ represents a small viscous damping coefficient. This scaling of damping and noise terms is such that the leading-order diffusion part balances the leading-order drift term, after transforming to polar coordinates as shown in Sec. 2. Under the assumption that the natural frequencies are *noncommensurable*, the small-noise expansions of the moment Lyapunov exponent for this 2-dof system were obtained by Sri Namachchivaya et al. [2] for small *p*. Sri Namachchivaya and Van Roessel [3], extended these to obtain an asymptotic representation of the moment Lyapunov exponent for finite *p*. Our interest is in the case when the two frequencies are *commensurable*, i.e., there exists a relation of the form $m_1\omega_1 = m_2\omega_2$, where m_1 and m_2 are integers. As before [2,3] we consider a real-noise excitation with specific infinitesimal generator *G*. It is assumed that *G* has an isolated simple zero eigenvalue. In this paper, we derive the generator $L(p)$ for finite *p*, whose principal eigenvalue is the moment Lyapunov exponent, by an asymptotic expansion similar to that presented in Sri Namachchivaya and Van Roessel [3]. The second method is based on stochastic averaging, and it is presented in Sri Namachchivaya et al. [8].

Section 2 presents the formulation of the problem and describes the general formula of the moment Lyapunov exponent for linear systems with real noise. In Sec. 3, a small-noise expansion is constructed and we obtain the appropriate eigenvalue problem for the moment Lyapunov exponent. The generator for the eigenvalue problem is derived without any restriction on the size of *p*. This equation along with appropriate boundary conditions forms the eigenvalue problem for which moment Lyapunov exponent is the principal eigenvalue. An orthogonal expansion for the eigenvalue problem based on Galerkin method is presented in Sec. 4. In Sec. 5, various cases of interest are obtained numerically based on Galerkin approximation. One-to-one internal resonances arise in many physical systems. However, the semi-simple form that is examined in this paper is common in coupled oscillators where there is geometric symmetry present in the physical system. The moment Lyapunov exponent in terms of spectral densities is calculated for this special case, which is common in many conservative structural and mechanical systems.

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 30, 2002; final revision, March 18, 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 of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Problem Formulation

The stochastic term $\xi(t)$ is a real-noise process on a smooth connected Riemannian manifold M (with or without boundary) with f a smooth nonconstant function defined on M. The associated infinitesimal generator is assumed to have the form

$$G(\xi)=\sum_{i=1}^{n}\mu_i(\xi)\frac{\partial}{\partial\xi_i}+\frac{1}{2}\sum_{k=1}^{r}\left[\sum_{i=1}^{n}\sigma_i^k(\xi)\frac{\partial}{\partial\xi_i}\right]\left[\sum_{i=1}^{n}\sigma_i^k(\xi)\frac{\partial}{\partial\xi_i}\right] \tag{2}$$

The almost-sure stability of the equilibrium state $q=\dot{q}=0$ of (1) is to be investigated. Using the transformation $q_i=x_{2i-1}$, $\dot{q}_i=\omega_i x_{2i}$, $i=1$, 2, Eq. (1) may be represented by the following system of Stratonovich differential equations:

$$\begin{gathered}\dot{x}=Ax+\varepsilon f(\xi)Bx,\quad x_0\in\mathbb{R}^4\\ d\xi=\mu(\xi)dt+\sum_{k=1}^{r}\sigma^k(\xi)\circ dW_k,\quad \xi\in M\end{gathered} \tag{3}$$

where

$$A=\begin{bmatrix}0 & \omega_1 & 0 & 0\\ -\omega_1 & -2\varepsilon^2\zeta\omega_1 & 0 & 0\\ 0 & 0 & 0 & \omega_2\\ 0 & 0 & -\omega_2 & -2\varepsilon^2\zeta\omega_2\end{bmatrix},$$

$$B=\begin{bmatrix}0 & 0 & 0 & 0\\ -p_{11} & 0 & -p_{12} & 0\\ 0 & 0 & 0 & 0\\ -p_{21} & 0 & -p_{22} & 0\end{bmatrix},\quad p_{ij}=\frac{k_{ij}}{\omega_i}$$

Consider the transformation $\{x_1,x_2,x_3,x_4\}\rightarrow\{\rho,\phi_1,\phi_2,\theta\}$ given by

$$\begin{gathered}x_1=e^{\rho}\cos\phi_1\cos\theta,\quad x_2=-e^{\rho}\sin\phi_1\cos\theta\\ x_3=e^{\rho}\cos\phi_2\sin\theta,\quad x_4=-e^{\rho}\sin\phi_2\sin\theta\end{gathered} \tag{4}$$

with the determinant

$$\frac{\partial(\rho,\phi_1,\phi_2,\theta)}{\partial(x_1,x_2,x_3,x_4)}=2e^{-\rho}\csc 2\theta$$

where $\rho\in\mathbb{R}$, $0\leqslant\phi_i\leqslant 2\pi$, $0<\theta<\pi/2$. The transformation is singular at $\theta=0$, $\pi/2$. Applying the transformation (4) to (3) yields the following set of equations for the logarithm of the amplitude ρ, phase variables (ϕ_1,ϕ_2,θ), and noise process ξ:

$$\begin{gathered}\dot{\rho}=\varepsilon q_1(\phi_1,\phi_2,\theta,\xi)+\varepsilon^2 q_2(\phi_1,\phi_2,\theta,\delta)\\ \dot{\theta}=\varepsilon s_1(\phi_1,\phi_2,\theta,\xi)+\varepsilon^2 s_2(\phi_1,\phi_2,\theta,\delta)\\ \dot{\phi}_i=\omega_i+\varepsilon h_{i1}(\phi_1,\phi_2,\theta,\xi)+\varepsilon^2 h_{i2}(\phi_1,\phi_2,\theta,\delta)\\ d\xi=\mu(\xi)dt+\sum_{k=1}^{r}\sigma^k(\xi)\circ dW_k,\quad \xi\in M\end{gathered} \tag{5}$$

where

$$q_1(\phi_1,\phi_2,\theta,\xi)=\tfrac{1}{4}f(\xi)[q_1^0(\phi_1,\phi_2)+q_1^c(\phi_1,\phi_2)\cos 2\theta+q_1^s(\phi_1,\phi_2)\sin 2\theta]$$

$$q_2(\phi_1,\phi_2,\theta,\delta)=q_2^0(\phi_1,\phi_2)+q_2^c(\phi_1,\phi_2)\cos 2\theta$$

$$s_1(\phi_1,\phi_2,\theta,\xi)=\tfrac{1}{4}f(\xi)[s_1^0(\phi_1,\phi_2)+s_1^c(\phi_1,\phi_2)\cos 2\theta+s_1^s(\phi_1,\phi_2)\sin 2\theta]$$

$$s_2(\phi_1,\phi_2,\theta,\delta)=s_2^s(\phi_1,\phi_2)\sin 2\theta,$$

$$h_{11}(\phi_1,\phi_2,\theta,\xi)=\tfrac{1}{2}f(\xi)[h_{11}^0(\phi_1,\phi_2)+h_{11}^\theta(\phi_1,\phi_2)\tan\theta]$$

$$h_{21}(\phi_1,\phi_2,\theta,\xi)=\tfrac{1}{2}f(\xi)[h_{21}^0(\phi_1,\phi_2)+h_{21}^\theta(\phi_1,\phi_2)\cot\theta]$$

$$h_{12}(\phi_1,\phi_2,\theta,\delta)=h_{12}(\phi_1,\delta)=-\delta_1\sin 2\phi_1$$

$$h_{22}(\phi_1,\phi_2,\theta,\delta)=h_{22}(\phi_2,\delta)=-\delta_2\sin 2\phi_2$$

$$q_1^0(\phi_1,\phi_2)=p_{11}\sin 2\phi_1+p_{22}\sin 2\phi_2,\quad s_1^0(\phi_1,\phi_2)=p_{21}^-\sin\phi^+-p_{21}^+\sin\phi^-$$

$$q_1^c(\phi_1,\phi_2)=p_{11}\sin 2\phi_1-p_{22}\sin 2\phi_2,\quad s_1^c(\phi_1,\phi_2)=p_{21}^+\sin\phi^+-p_{21}^-\sin\phi^-$$

$$q_1^s(\phi_1,\phi_2)=p_{21}^+\sin\phi^+-p_{21}^-\sin\phi^-,\quad s_1^s(\phi_1,\phi_2)=p_{22}\sin 2\phi_2-p_{11}\sin 2\phi_1$$

$$q_2^0(\phi_1,\phi_2)=-\tfrac{1}{2}[\delta_1(1-\cos 2\phi_1)+\delta_2(1-\cos 2\phi_2)]$$

$$q_2^c(\phi_1,\phi_2)=-\tfrac{1}{2}[\delta_1(1-\cos 2\phi_1)-\delta_2(1-\cos 2\phi_2)]$$

$$s_2^s(\phi_1,\phi_2)=\tfrac{1}{2}[\delta_1(1-\cos 2\phi_1)+\delta_2(1-\cos 2\phi_2)]$$

$$h_{11}^0(\phi_1)=p_{11}(1+\cos 2\phi_1),\quad h_{11}^\theta(\phi_1,\phi_2)=p_{12}(\cos\phi^++\cos\phi^-)$$

$$h_{21}^0(\phi_1)=p_{22}(1+\cos 2\phi_2),\quad h_{21}^\theta(\phi_1,\phi_2)=p_{21}(\cos\phi^++\cos\phi^-)$$

$$p_{21}^{\pm}=p_{21}\pm p_{12},\quad \phi^{\pm}=\phi_1\pm\phi_2,\quad \delta_i=\zeta\omega_i$$

Since the processes $(\phi_1,\phi_2,\theta,\xi)$ do not depend on ρ, the processes $(\phi_1,\phi_2,\theta,\xi)$ alone form a diffusive Markov process, and the associated generator is given by

$$\mathcal{L}^{\varepsilon}=\mathcal{L}_0+\varepsilon\mathcal{L}_1+\varepsilon^2\mathcal{L}_2$$

where

$$\mathcal{L}_0=G+\omega_1\frac{\partial}{\partial\phi_1}+\omega_2\frac{\partial}{\partial\phi_2}$$

$$\mathcal{L}_1=s_1\frac{\partial}{\partial\theta}+h_{11}\frac{\partial}{\partial\phi_1}+h_{21}\frac{\partial}{\partial\phi_2}$$

$$\mathcal{L}_2=s_2\frac{\partial}{\partial\theta}+h_{12}\frac{\partial}{\partial\phi_1}+h_{22}\frac{\partial}{\partial\phi_2}$$

It is worthy to note that the influence of small $\mathcal{O}(\epsilon^2)$ deviations from the exact commensurable frequencies, i.e., $m_1\omega_1(\mu)+\dots+m_n\omega_n(\mu)=\mathcal{O}(\epsilon^2)$, on the moment Lyapunov exponent can be easily incoperated by adding an appropriate detuning term in $h_{i2}(\phi_1,\phi_2,\theta,\delta)$ of (5).

Then we obtain the following expression for the norm of the response:

$$\|x(t;x_0)\|=\|x_0\|\exp\left\{\int_0^t q^{\varepsilon}(\xi(\tau),\phi_1(\tau),\phi_2(\tau),\theta(\tau))d\tau\right\}$$

where

$$q^{\varepsilon}(\phi_1(t),\phi_2(t),\theta(t),\xi(t),\delta)\overset{\text{def}}{=}\varepsilon q_1(\phi_1(t),\phi_2(t),\theta(t),\xi(t))+\varepsilon^2 q_2(\phi_1(t),\phi_2(t),\theta(t),\delta)$$

Combining the above result with the definition of moment Lyapunov exponent yields

$$g(p;x_0)=\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}\left[\exp\left\{p\int_0^t q^{\varepsilon}(\xi(\tau),\phi_1(\tau),\phi_2(\tau),\theta(\tau),\delta)d\tau\right\}\right]\quad\text{for } p\in\mathbb{R},\ \text{and fixed } x_0\in\mathbb{R}^4\backslash\{0\}\tag{6}$$

For $p\in\mathbb{R}$, it was shown by Arnold et al. [7] that $g(p)$ is the principal eigenvalue of

$$L^{\varepsilon}(p)\overset{\text{def}}{=}\mathcal{L}^{\varepsilon}+pq^{\varepsilon}(\xi,\phi_1,\phi_2,\theta,\delta)=L_0(p)+\varepsilon L_1(p)+\varepsilon^2L_2(p)\tag{7}$$

where $L^{\varepsilon}(p)$ acts on $C(M\times\mathbb{S}^3)$ and

$$L_0(p)=G(\xi)+\sum_{i=1}^{2}\omega_i\frac{\partial}{\partial\phi_i}+pq_0(\phi_1,\phi_2,\theta,\xi)=\mathcal{L}_0+pq_0$$

$$L_1(p)=s_1(\phi_1,\phi_2,\theta,\xi)\frac{\partial}{\partial\theta}+\sum_{i=1}^{2}h_{i1}(\phi_1,\phi_2,\theta,\xi)\frac{\partial}{\partial\phi_i}+pq_1(\phi_1,\phi_2,\theta,\xi)=\mathcal{L}_1+pq_1$$

$$L_2(p)=s_2(\phi_1,\phi_2,\theta,\delta)\frac{\partial}{\partial\theta}+\sum_{i=1}^{2}h_{i2}(\phi_1,\phi_2,\theta,\delta)\frac{\partial}{\partial\phi_i}+pq_2(\phi_1,\phi_2,\theta,\delta)=\mathcal{L}_2+pq_2$$

Consider the operator $L^{\varepsilon}(p)$ and its adjoint $L^{\varepsilon *}(p)$. Then by the main theorem in Arnold et al. [7], $g^{\varepsilon}(p)$ is an isolated simple eigenvalue of $L^{\varepsilon}(p)$ with non-negative eigenfunction $\psi^{\varepsilon}(p)$ such that $\|\psi^{\varepsilon}(p)\|=1$. The adjoint operator $L^{\varepsilon *}(p)$ has an eigenfunction $\mu^{\varepsilon}(p)$ corresponding to $g^{\varepsilon}(p)$, which is unique and has the property $\langle\psi^{\varepsilon}(p),\mu^{\varepsilon}(p)\rangle=1$, i.e.,

$$L^{\varepsilon}(p)\psi^{\varepsilon}(p)=g^{\varepsilon}(p)\psi^{\varepsilon}(p),\quad\langle\psi^{\varepsilon}(p),\mu^{\varepsilon}(p)\rangle=1\quad\forall p\in\mathbb{R}\tag{8}$$

3 Asymptotic Results for Coupled Oscillators

Here we consider an expansion of the moment Lyapunov exponent in powers of ε

$$g_{\varepsilon}(p)=g_0(p)+\varepsilon g_1(p)+\varepsilon^2g_2(p)+\mathcal{O}(\varepsilon^2)$$

It has been shown that such an expansion is asymptotic by Sri Namachchivaya et al. [2], Arnold et al. [9], Khasminskii and Moshchuk [10], and Pardoux and Wihstutz [11]. Insertion of these expansions into (8) leads to the following sequence of Poisson equations:

$$[L_0(p)-g_0(p)]\psi_0=0\tag{9}$$

$$[L_0(p)-g_0(p)]\psi_1=g_1(p)\psi_0-L_1(p)\psi_0\tag{10}$$

$$[L_0(p)-g_0(p)]\psi_2=g_2(p)\psi_0+g_1(p)\psi_1-L_2(p)\psi_0-L_1(p)\psi_1\tag{11}$$

$$\cdots$$

$$\cdots$$

$$[L_0(p)-g_0(p)]\psi_n=g_2(p)\psi_{n-2}+g_1(p)\psi_{n-1}-L_2(p)\psi_{n-2}-L_1(p)\psi_{n-1}$$

$$\cdots$$

$$\cdots$$

We construct a formal expansion of the adjoint problem, i.e.,

$$\mu^{\varepsilon}=\mu_0+\varepsilon\mu_1+\cdots+\varepsilon^N\mu_N+\cdots$$

Substituting this expansion and the expansion for L^{ε} into the Fokker-Planck equation yields the following sequence of Poisson equations to be solved for $\mu_0,\ \mu_1,\ \mu_2,\dots$:

$$L_0^*\mu_0=0\tag{12}$$

$$L_0^*\mu_1=-L_1^*\mu_0\tag{13}$$

$$L_0^*\mu_2=-L_1^*\mu_1-L_2^*\mu_0\tag{14}$$

$$\cdots$$

$$\cdots$$

Our interest is in the case when the two frequencies are *commensurable*, i.e., there exists a relation of the form $m_1\omega_1=m_2\omega_2$, where m_1 and m_2 are integers. To solve the above sequence of Poisson equations, it is convenient to consider another transformation, $\{\phi_1, \phi_2\} \rightarrow \{\gamma_1, \gamma_2\}$ given by

$$\gamma_1 = \omega_2 \phi_1 - \omega_1 \phi_2, \quad \gamma_2 = \omega_1 \phi_2$$

and the inverse transformation is

$$\phi_1 = \Phi_1(\gamma_1, \gamma_2) = (\gamma_1 + \gamma_2)/\omega_2, \quad \phi_2 = \Phi_2(\gamma_1, \gamma_2) = \gamma_2/\omega_1$$

Hence, for any $\psi \in C^2(M \times \mathbb{S}^3)$, we define the differential operators in $(\xi, \theta, \gamma_1, \gamma_2)$ as

$$L_0\psi = G + \omega_i \frac{\partial \psi}{\partial \gamma_j} \frac{\partial \gamma_j}{\partial \phi_i} + pq_0\psi = G + \omega_1\omega_2 \frac{\partial \psi}{\partial \gamma_2} + pq_0\psi$$

$$L_1\psi = s_1 \frac{\partial \psi}{\partial \theta} + h_{11} \frac{\partial \psi}{\partial \gamma_1} \frac{\partial \gamma_1}{\partial \phi_1} + h_{21} \frac{\partial \psi}{\partial \gamma_1} \frac{\partial \gamma_1}{\partial \phi_2} + pq_1\psi = s_1 \frac{\partial \psi_0}{\partial \theta} + H_1 \frac{\partial \psi}{\partial \gamma_1} + \omega_1 h_{21} \frac{\partial \psi}{\partial \gamma_2} + pq_1\psi$$

$$L_2\psi = s_2 \frac{\partial \psi}{\partial \theta} + h_{12} \frac{\partial \psi}{\partial \gamma_1} \frac{\partial \gamma_1}{\partial \phi_1} + h_{22} \frac{\partial \psi}{\partial \gamma_1} \frac{\partial \gamma_1}{\partial \phi_2} + pq_2\psi = s_2 \frac{\partial \psi}{\partial \theta} + H_2 \frac{\partial \psi}{\partial \gamma_1} + \omega_1 h_{22} \frac{\partial \psi}{\partial \gamma_2} + pq_2\psi$$

where

$$q_1(\gamma_1, \gamma_2, \theta, \xi) \overset{\text{def}}{=} q_1(\phi_1(\gamma_1, \gamma_2), \phi_2(\gamma_1, \gamma_2), \theta, \xi)$$

$$q_2(\gamma_1, \gamma_2, \theta, \xi) \overset{\text{def}}{=} q_2(\phi_1(\gamma_1, \gamma_2), \phi_2(\gamma_1, \gamma_2), \theta, \xi)$$

$$s_1(\gamma_1, \gamma_2, \theta, \xi) \overset{\text{def}}{=} s_1(\phi_1(\gamma_1, \gamma_2), \phi_2(\gamma_1, \gamma_2), \theta, \delta)$$

$$s_2(\gamma_1, \gamma_2, \theta, \xi) \overset{\text{def}}{=} s_2(\phi_1(\gamma_1, \gamma_2), \phi_2(\gamma_1, \gamma_2), \theta, \delta)$$

$$H_1(\gamma_1, \gamma_2, \theta, \xi) \overset{\text{def}}{=} \omega_2 h_{11}(\phi_1(\gamma_1, \gamma_2), \phi_2(\gamma_1, \gamma_2), \theta, \xi) - \omega_1 h_{21}(\phi_1(\gamma_1, \gamma_2), \phi_2(\gamma_1, \gamma_2), \theta, \xi)$$

$$H_2(\gamma_1, \gamma_2, \theta, \xi) \overset{\text{def}}{=} \omega_2 h_{12}(\phi_1(\gamma_1, \gamma_2), \phi_2(\gamma_1, \gamma_2), \theta, \delta) - \omega_1 h_{22}(\phi_1(\gamma_1, \gamma_2), \phi_2(\gamma_1, \gamma_2), \theta, \delta)$$

Before proceeding with the determination of solutions, let us state some facts about the periodicity and commensurability in the new variables γ_1, γ_2.

Lemma 3.1 *If (i)* ω_1 *and* ω_2 *are commensurable (i.e.,* $\omega_1/\omega_2 \in \mathbb{Q}$ *); and* (ii) $f(x)$ *is both* ω_1 *and* ω_2*-periodic then* f *is* ω*-periodic where* $\omega = \gcd(\omega_1, \omega_2)$.

Proof: It follows from (i) that $\exists n_1, n_2 \in \mathbb{Z}$ such that $\omega_1/\omega_2 = n_1/n_2$ with n_1 and n_2 relatively prime, which in turn implies that $\exists m_1, m_2 \in \mathbb{Z}$ such that $m_1 n_1 + m_2 n_2 = 1$. Now let $\omega =^{\text{def}} \omega_1/n_1 = \omega_2/n_2$. Then $f(x+\omega) = f(x + (m_1 n_1 + m_2 n_2)\omega) = f(x + m_1\omega_1 + m_2\omega_2) = f(x)$. ☐

Then the following result is easily obtained.

Lemma 3.2 *If* $w(\phi_1, \phi_2)$ *is* 2π*-periodic in* ϕ_1*, then* $\Phi^* w$ *is* $\overline{\omega_2}$*-periodic in* γ_1*, where* $\overline{\omega_2} =^{\text{def}} 2\pi\omega_2$.

Proof: Let $W(\gamma_1, \gamma_2) = (\Phi^* w)(\gamma_1, \gamma_2)$. Then

$$\begin{aligned} W(\gamma_1 + \overline{\omega_2}, \gamma_2) &= w[\Phi_1(\gamma_1 + \overline{\omega_2}, \gamma_2), \Phi_2(\gamma_1 + \overline{\omega_2}, \gamma_2)] \\ &= w[\Phi_1(\gamma_1, \gamma_2) + 2\pi, \Phi_2(\gamma_1, \gamma_2)] \\ &= w[\Phi_1(\gamma_1, \gamma_2), \Phi_2(\gamma_1, \gamma_2)] = W(\gamma_1, \gamma_2) \end{aligned}$$

☐

3.1 Solution to $\mathcal{O}(1)$.

Since $q_0(\gamma_1, \gamma_2, \theta, \xi) \equiv 0$, it follows from the definition of $g(p)$ that $g_0(p) \equiv 0$. Thus the operator $L_0(p)$ reduces to $\mathcal{L}_0$ and (9) becomes

$$\mathcal{L}_0 \psi_0 = 0$$

Since the equations to be solved involve the differential operator $\mathcal{L}_0$ at each stage, the solution of the corresponding adjoint problem $\mathcal{L}_0^* \mu_0 = 0$, along with periodic boundary conditions $\mu_0(\phi_1 + 2\pi, \phi_2, \theta, \xi) = \mu_0(\phi_1, \phi_2 + 2\pi, \theta, \xi) = \mu_0(\phi_1, \phi_2, \theta, \xi)$, is required. However, from Lemma 3.2, it follows that in the new coordinates, $\mu_0(\gamma_1, \gamma_2, \theta, \xi)$ is $2\pi\omega_1$ and $2\pi\omega_2$-periodic in γ_2 and γ_1, respectively. In order to make the problem tractable, G will be assumed to have an *isolated simple zero eigenvalue*. Hence, the only solution of $Gu = 0$ is $u \equiv$ constant. It follows that the associated adjoint operator G^* also has zero as a simple, isolated eigenvalue, and the normalized invariant measure $v(\xi)d\xi$ satisfies $G^* v(\xi) = 0$.

Since the frequencies ω_1 and ω_2 are commensurable and G has an isolated simple zero eigenvalue, the solution to $\mathcal{L}_0^* \mu_0 = 0$ is

$$\mu_0(\gamma_1, \theta, \xi) = \frac{\nu(\xi)\mathcal{F}(\gamma_1, \theta)}{2\pi}$$

where $\mathcal{F}$ is an arbitrary function of (γ_1, θ), which has yet to be determined. By a similar argument it follows that:

$$\psi_0 \in \ker(\mathcal{L}_0) = \{C(\gamma_1, \theta) : C \text{ is an arbitrary function of } \gamma_1, \theta\}$$

Therefore, $\psi_0 = \psi_0(\gamma_1, \theta)$, a function of (γ_1, θ) which has yet to be determined.

3.1.1 *Solution to* $\mathcal{O}(\varepsilon)$.

Inserting the above expression for ψ_0 into (10) results in

$$\mathcal{L}_0 \psi_1 = g_1(p)\psi_0 - s_1(\gamma_1, \gamma_2, \theta, \xi) \frac{\partial \psi_0}{\partial \theta} - H_1(\gamma_1, \gamma_2, \theta, \xi) \frac{\partial \psi_0}{\partial \gamma_1} - pq_1(\gamma_1, \gamma_2, \theta, \xi)\psi_0 \tag{15}$$

Premultiplying by $\nu(\xi)$ and integrating with respect to ξ and γ_2 yields

$$\int_0^{2\pi\omega_1}\int_M \nu(\xi)\cdot G\psi_1 d\xi d\gamma_2+\int_M \nu(\xi)\int_0^{2\pi\omega_1}\omega_1\omega_2\frac{\partial\psi_1}{\partial\gamma_2}d\gamma_2 d\xi$$

$$=g_1(p)\psi_0-\int_0^{2\pi\omega_1}\int_M \nu(\xi)s_1(\gamma_1,\gamma_2,\theta,\xi)\frac{\partial\psi_0}{\partial\theta}d\xi d\gamma_2$$

$$-\int_0^{2\pi\omega_1}\int_M \nu(\xi)H_1(\gamma_1,\gamma_2,\theta,\xi)\frac{\partial\psi_0}{\partial\gamma_1}d\xi d\gamma_2$$

$$-\int_0^{2\pi\omega_1}\int_M p\nu(\xi)q_1(\gamma_1,\gamma_2,\theta,\xi)\psi_0 d\xi d\gamma_2$$

Since, $G^*\nu(\xi)=0$ and ψ_1 is $2\pi\omega_1$-periodic in γ_2 the left-hand side of the above equation is identically zero, and the last three terms on the right-hand side of the above equation are zero due to the fact $f(\xi)$ is mean zero. Hence, the eigenvalue $g_1(p)$ is zero.

The solution ψ_1, of (15) with $g_1(p)=0$, can be expressed in terms of the Green's function $g(\xi,t;\eta,0)$ for the operator G. Rewrite (15) as

$$L_0\psi_1=-\tfrac{1}{4}f(\xi)R_1(\gamma_1,\gamma_2,\theta)$$

where

$$R_1(\gamma_1,\gamma_2,\theta)\overset{\text{def}}{=}p[q_1^0(\gamma_1,\gamma_2)+q_1^c(\gamma_1,\gamma_2)\cos 2\theta+q_1^s(\gamma_1,\gamma_2)\sin 2\theta]\psi_0+[s_1^0(\gamma_1,\gamma_2)+s_1^c(\gamma_1,\gamma_2)\cos 2\theta+s_1^s(\gamma_1,\gamma_2)\sin 2\theta]\frac{\partial\psi_0}{\partial\theta}$$
$$+2[\omega_2h_{11}^0(\gamma_1,\gamma_2)-\omega_1h_{21}^0(\phi_1,\phi_2)+\omega_2h_{11}^\theta(\gamma_1,\gamma_2)\tan\theta-\omega_1h_{21}^\theta(\gamma_1,\gamma_2)\cot\theta]\frac{\partial\psi_0}{\partial\gamma_1} \tag{16}$$

If $g(\eta,\tau;\xi,0)$ is a solution of

$$\left(\frac{\partial}{\partial\tau}-G\right)g=0,\quad \text{with}\quad g(\eta,0;\xi,0)=\delta(\eta-\xi)$$

then

$$\psi_1(\gamma_1,\gamma_2,\theta,\xi)=\frac{1}{4}\int_0^\infty K(\xi,T)\times R_1(\gamma_1,\gamma_2+\omega_1\omega_2T,\theta)dT$$

where

$$K(\xi,T)\overset{\text{def}}{=}\int_M f(\eta)g(\eta,T;\xi,0)d\eta$$

3.1.2 Solution to $\mathcal{O}(\varepsilon^2)$. Employing the above results, Poisson equation (11) for ψ_2 becomes

$$\mathcal{L}_0\psi_2=-\left(pq_2(\gamma_1,\gamma_2,\theta;\xi)\psi_0+s_2(\gamma_1,\gamma_2,\theta,\xi)\frac{\partial\psi_0}{\partial\theta}\right.$$
$$\left.+H_2(\gamma_1,\gamma_2,\theta,\xi)\frac{\partial\psi_0}{\partial\gamma_1}\right)-L_1\psi_1(\gamma_1,\gamma_2,\theta,\xi;p)+g_2(p)\psi_0 \tag{17}$$

the first term on the right-hand side of (17) depends on damping alone while the second term depends on ξ.

Definition 3.3 (Averaging Operator) *Fix $\psi\in C^\infty(\mathbb{S}^3\times M)$, which is $2\pi\omega_1$-periodic in γ_2 in its second argument. Define $\overline{\psi}\in C^\infty(\mathbb{R}^2\times M)$ by*

$$(\overline{\psi})(\gamma_1,\theta,\xi)\overset{def}{=}\frac{1}{2\pi\omega_1}\int_0^{2\pi\omega_1}\psi(\gamma_1,\gamma_2,\theta,\xi)d\gamma_2$$

for all $\gamma_1,\ \theta\in\mathbb{S}^2$ and $\xi\in M$.

Once again, premultiplying (17) by $\nu(\xi)$ and integrating with respect to ξ and γ_2 yields

$$g_2(p)\psi_0-\left(p\overline{q_2}(\gamma_1,\theta)\psi_0+\overline{s_2}(\gamma_1,\theta)\frac{\partial\psi_0}{\partial\theta}+\overline{H_2}(\gamma_1,\theta)\frac{\partial\psi_0}{\partial\gamma_1}\right)$$
$$-\int_M \nu(\xi)\cdot\overline{L_1\psi_1}(\gamma_1,\theta,\xi;p)d\xi=0 \tag{18}$$

We consider the three terms in the bracket in (18), then

$$\overline{q_2}=-\tfrac{1}{2}(\delta_1+\delta_2)-\tfrac{1}{2}(\delta_1-\delta_2)\cos 2\theta$$

$$\overline{s_2}=\tfrac{1}{2}(\delta_1-\delta_2)\sin 2\theta,\quad \overline{H_2}=0$$

Hence,

$$\left(p\overline{q_2}\psi_0+\overline{s_2}\frac{\partial\psi_0}{\partial\theta}+\overline{H_2}\frac{\partial\psi_0}{\partial\gamma_1}\right)(\gamma_1,\theta)$$
$$=p\left\{-\frac{1}{2}(\delta_1+\delta_2)-\frac{1}{2}(\delta_1-\delta_2)\cos 2\theta\right\}\psi_0(\gamma_1,\theta)$$
$$+\frac{1}{2}(\delta_1-\delta_2)\sin 2\theta\frac{\partial\psi_0}{\partial\theta}(\gamma_1,\theta) \tag{19}$$

We now consider the last term in (18)

$$\int_M \nu(\xi)\cdot\overline{L_1\psi_1}(\gamma_1,\theta,\xi;p)d\xi=\frac{1}{4}\int_M\int_0^\infty \nu(\xi)K(\xi,T)\overline{(L_1R_1^T)}(\gamma_1,\theta)dTd\xi=\frac{1}{2\pi\omega_1}\int_0^\infty\int_0^{2\pi\omega_1}\frac{\mathcal{R}_\xi(\tau)}{4}\left\{s_1\frac{\partial R_1^\tau}{\partial\theta}(\gamma_1,\gamma_2,\theta)\right.$$
$$\left.+H_1\frac{\partial R_1^\tau}{\partial\gamma_1}(\gamma_1,\gamma_2,\theta)+\omega_1h_{21}\frac{\partial R_1^\tau}{\partial\gamma_2}(\gamma_1,\gamma_2,\theta)+pq_1R_1^\tau(\gamma_1,\gamma_2,\theta)\right\}d\gamma_2d\tau \tag{20}$$

where

$$R_1^T(\gamma_1,\gamma_2,\theta)\overset{\text{def}}{=}R_1(\gamma_1+\omega_1 T,\gamma_2+\omega_2 T,\theta)$$

and

$$\mathcal{R}_\xi(\tau)\overset{\text{def}}{=}\int_M f(\xi)\nu(\xi)K(\xi,\tau)d\xi$$

Then, after a lengthy calculation, we get

$$\int_M \nu(\xi)\cdot\overline{L_1\psi_1}(\gamma_1,\theta,\xi;p)d\xi=[pQ_{\psi_1}(\gamma_1,\theta,S_\xi,T_\xi)+p^2\hat{Q}_{\psi_1}(\gamma_1,\theta,S_\xi,T_\xi)]\psi_0(\gamma_1,\theta)+[\mu_{\psi_1}(\gamma_1,\theta,S_\xi,T_\xi)$$
$$+p\hat{\mu}_{\psi_1}(\gamma_1,\theta,S_\xi,T_\xi)]\frac{\partial\psi_0}{\partial\theta}(\gamma_1,\theta)+[\nu_{\psi_1}(\gamma_1,\theta,S_\xi,T_\xi)+p\hat{\nu}_{\psi_1}(\gamma_1,\theta,S_\xi,T_\xi)]\frac{\partial\psi_0}{\partial\gamma_1}(\gamma_1,\theta)$$
$$+\frac{1}{2}\sigma^2_{\theta\theta}(\gamma_1,\theta,S_\xi,T_\xi)\frac{\partial^2\psi_0}{\partial\theta^2}(\gamma_1,\theta)+\frac{1}{2}\sigma^2_{\gamma_1\gamma_1}(\gamma_1,\theta,S_\xi,T_\xi)\frac{\partial^2\psi_0}{\partial\gamma_1^2}(\gamma_1,\theta)$$
$$+\sigma^2_{\theta\gamma_1}(\gamma_1,\theta,S_\xi,T_\xi)\frac{\partial^2\psi_0}{\partial\theta\partial\gamma_1}(\gamma_1,\theta) \tag{21}$$

where a typical term, take for example a diffusion coefficient, is evaluated as

$$\sigma^2_{\theta\theta}=\int_M \nu(\xi)\cdot\{\overline{s_1^0 s_1^{0T}}+(\overline{s_1^0 s_1^{cT}+s_1^{0T}s_1^c})\cos 2\theta$$
$$+(\overline{s_1^0 s_1^{sT}+s_1^{0T}s_1^s})\sin 2\theta+\overline{s_1^c s_1^{cT}}\cos^2 2\theta$$
$$+(\overline{s_1^c s_1^{sT}+s_1^{cT}s_1^s})\cos 2\theta\sin 2\theta+\overline{s_1^s s_1^{sT}}\sin^2 2\theta\}d\xi$$

Here the superscript T in the coefficient s_1^{0T} denotes that the variable γ_2 has been shifted to $\gamma_2+\omega_1\omega_2 T$. The explicit expressions for the coefficients Q_{ψ_1}, $\hat{Q}_{\psi_1}$, μ_{ψ_1}, $\hat{\mu}_{\psi_1}$, ν_{ψ_1}, $\hat{\nu}_{\psi_1}$, $\sigma^2_{\theta\theta}$, $\sigma^2_{\gamma_1\gamma_1}$, $\sigma^2_{\theta\gamma_1}$ are given in the Appendix. For physical systems with symmetry, one-to-one resonance is the rule rather than the exception. Hence, while averaging these coefficients in (20), we have made use of this specific resonance condition called *one-to-one* semi-simple resonance, where we define

$$\omega\overset{\text{def}}{=}\omega_1=\omega_2$$

and the results for the other resonance cases can easily be obtained. We have also made use of the correlation function of $f(\xi)$ and the cosine and sine spectrum given, respectively, by

$$S_\xi(\omega)\overset{\text{def}}{=}2\int_0^\infty \mathcal{R}_\xi(\tau)\cos(\omega\tau)d\tau$$

and

$$T_\xi(\omega)\overset{\text{def}}{=}2\int_0^\infty \mathcal{R}_\xi(\tau)\sin(\omega\tau)d\tau$$

Combining the the right-hand sides of Eqs. (19) and (21) and defining

$$\gamma\overset{\text{def}}{=}\frac{\gamma_1}{\omega}$$

yields the partial differential equation at $\mathcal{O}(\varepsilon^2)$, whose principal eigenvalue is the moment Lyapunov exponent, i.e.,

$$\tilde{L}(p)\psi_0=g_2(p)\psi_0 \tag{22}$$

where

$$\tilde{L}(p)\psi_0\overset{\text{def}}{=}\frac{1}{2}\sigma^2_{\theta\theta}(\theta,\gamma)\frac{\partial^2\psi_0}{\partial\theta^2}(\theta,\gamma)+\frac{1}{2}\sigma^2_{\gamma\gamma}(\theta,\gamma)\frac{\partial^2\psi_0}{\partial\gamma^2}(\theta,\gamma)+\sigma^2_{\gamma\theta}(\theta,\gamma)\frac{\partial^2\psi_0}{\partial\theta\partial\gamma}(\theta,\gamma)+[\mu(\theta,\gamma)+p\hat{\mu}(\theta,\gamma)]\frac{\partial\psi_0}{\partial\theta}(\theta,\gamma)+[\nu(\theta,\gamma)$$
$$+p\hat{\nu}(\theta,\gamma)]\frac{\partial\psi_0}{\partial\gamma}(\theta,\gamma)+\left[pQ(\theta,\gamma)+\frac{1}{2}p^2\hat{Q}(\theta,\gamma)\right]\psi_0(\theta,\gamma) \tag{23}$$

We reiterate that (22) was derived without any restriction on the size of p. This equation along with appropriate boundary conditions forms the eigenvalue problem for which $g_2(p)$ is the principal eigenvalue. In order to reduce the number of cases to be evaluated, we can simplify the coefficients p_{ij} by a suitable scaling of the state variables x. It can be shown as in [3] that it is always possible to take $p_{12}=\pm p_{21}=\kappa$. Since the semi-simple linear form $(\omega_1=\omega_2=\omega)$ natuarally occurs in conservative systems, we shall only consider $p_{12}=p_{21}=\kappa$, in which case the coefficients reduce to

$$\mu(\theta,\gamma)=[\sigma^2_{\theta,\theta}(\theta,\gamma)+\tfrac{1}{4}\kappa^2 S_\xi(0)\cos(2\gamma)]\cot 2\theta-\tfrac{1}{2}\Lambda\sin 2\theta+\tfrac{1}{8}S_\xi(2\omega)(p_{11}+p_{22})\kappa\cos(\gamma)\cos 2\theta+\tfrac{1}{8}T_\xi(2\omega)(p_{11}+p_{22})\kappa\sin(\gamma)$$
$$-\tfrac{1}{8}S_\xi(0)(p_{11}-p_{22})\kappa\cos(\gamma)$$

$$\hat{\mu}(\theta,\gamma)=\tfrac{1}{2}(2A(\gamma)\cos 2\theta-E_0)\sin 2\theta+\tfrac{1}{16}\kappa[2(p_{11}-p_{22})\cos^2 2\theta+(p_{11}+p_{22})\cos 2\theta-(p_{11}-p_{22})]S_\xi(2\omega)\cos(\gamma)$$

$$\nu(\theta,\gamma)=-\tfrac{1}{8}(p_{11}-p_{22})(p_{11}+p_{22})T_\xi(2\omega)-\tfrac{1}{2}S_\xi(0)\sin(2\gamma)\kappa^2\csc^2 2\theta+\tfrac{1}{4}\kappa[\{(p_{11}+p_{22})T_\xi(2\omega)\cos(\gamma)$$
$$+(p_{11}-p_{22})S_\xi(0)\sin(\gamma)\}\cos 2\theta-(p_{11}+p_{22})S_\xi(2\omega)\sin(\gamma)]\csc 2\theta+\tfrac{1}{4}S_\xi(0)\sin(2\gamma)\kappa^2$$

$$\hat{\nu}(\theta,\gamma)=-\tfrac{1}{8}\kappa[\{(p_{11}-p_{22})+(p_{11}+p_{22})\}S_{\xi}(2\omega)\sin(\gamma)]\csc 2\theta-\tfrac{1}{8}p_{11}p_{22}S_{\xi}(0)\sin(2\gamma)$$

$$\sigma^2_{\theta,\theta}(\theta,\gamma)=[A_0+\tfrac{1}{16}p_{11}p_{22}S_{\xi}(2\omega)\cos(2\gamma)]\cos^2 2\theta+C_0-\tfrac{1}{16}\{p_{11}p_{22}S_{\xi}(2\omega)+2\kappa^2S_{\xi}(0)\}\cos(2\gamma)-\tfrac{1}{8}\kappa(p_{11}-p_{22})S_{\xi}(2\omega)\cos(\gamma)\cos 2\theta\sin 2\theta$$

$$\sigma^2_{\gamma,\gamma}(\theta,\gamma)=\tfrac{1}{2}\omega^2\kappa^2\left\{S_{\xi}(0)+S_{\xi}(2\omega)+S_{\xi}(0)\cos\left(2\frac{\gamma_1}{\omega}\right)\right\}\cot^2(2\theta)-\tfrac{1}{2}\omega^2\kappa(p_{11}-p_{22})\{2S_{\xi}(0)+S_{\xi}(2\omega)\}\cos\left(\frac{\gamma_1}{\omega}\right)\cot(2\theta)+\tfrac{1}{8}\omega^2\left[(p_{11}^2+p_{22}^2)S_{\xi}(2\omega)+2(p_{11}-p_{22})^2S_{\xi}(0)-2p_{11}p_{22}S_{\xi}(2\omega)\cos\left(2\frac{\gamma_1}{\omega}\right)\right]$$

$$\sigma^2_{\theta,\gamma}(\theta,\gamma)=\tfrac{1}{4}\kappa^2S_{\xi}(0)\sin(2\gamma)\cot 2\theta-\tfrac{1}{4}\kappa(p_{11}-p_{22})S_{\xi}(0)\sin(\gamma)$$

$$Q(\theta,\gamma)=\sigma^2(\theta,\gamma)+\tfrac{1}{2}\Lambda\cos 2\theta+D+\tfrac{1}{8}S_{\xi}(0)\kappa^2\cos(2\gamma)+\frac{1}{8}S_{\xi}(2\omega)(p_{11}+p_{22})\kappa\cos(\gamma)\sin 2\theta$$

$$\hat{Q}(\theta,\gamma)=-[A_0+\tfrac{1}{16}p_{11}p_{22}S_{\xi}(2\omega)\cos(2\gamma)]\cos^2 2\theta+E_0\cos 2\theta+F_0+\tfrac{1}{16}S_{\xi}(2\omega)p_{11}p_{22}\cos(2\gamma)+\tfrac{1}{8}\kappa(p_{11}-p_{22})S_{\xi}(2\omega)\cos(\gamma)\cos 2\theta\sin 2\theta+\tfrac{1}{8}\kappa(p_{11}+p_{22})S_{\xi}(2\omega)\cos(\gamma)\sin 2\theta$$

with the constants given by

$$A_0=\tfrac{1}{32}\left\{4\kappa^2S(\omega^+)-\sum_{i=1}^{2}p_{ii}^2S(2\omega_i)\right\},$$

$$C_0=\tfrac{1}{32}\left\{4\kappa^2S(\omega^-)+\sum_{i=1}^{2}p_{ii}^2S(2\omega_i)\right\}$$

$$D_0=\tfrac{1}{2}(\Lambda_1+\Lambda_2)+\tfrac{1}{8}\kappa^2[S(\omega^+)-S(\omega^-)],$$

$$E_0=\tfrac{1}{16}\{p_{11}^2S(2\omega_1)-p_{22}^2S(2\omega_2)\}$$

$$F_0=\tfrac{1}{32}\left\{4\kappa^2S(\omega^+)+\sum_{i=1}^{2}p_{ii}^2S(2\omega_i)\right\},\quad \Lambda=\Lambda_1-\Lambda_2$$

$$\Lambda_i=-\delta_i+\tfrac{1}{8}p_{ii}^2S(2\omega_i),\quad i=1,2$$

where $\omega^+=2\omega_i=2\omega$ and $\omega^-=0$.

We note in passing that making terms with γ identically zero in the above equation recovers the equations in Sri Namachchivaya and van Roessel [3]. It is clear that the operator $\tilde{L}(p)$ obtained above is identical to the one in [8] using the method of stochastic averaging. Hence we are sure of the correctness of the results obtained in this paper. Except for some special cases, the general solution of (23) cannot be obtained explicitly for $g_2(p)$.

The domain for the diffusion process in θ and γ is

$$D=\left\{(\theta,\gamma):\theta\in\left[0,\frac{\pi}{2}\right],\gamma\in[0,2\pi]\right\}$$

It is obvious that ψ_0 is 2π—periodic in γ. Boundaries for the θ process are not physical, thus it is not clear what boundary conditions one should use to solve the eigenvalue problem (22). In general it is also possible, to have singularities in θ, *only the nonsingular cases will be considered here*. In order to obtain the boundary conditions at the boundaries $\theta=0$ and $\pi/2$, we investigate the asymptotic behavior of the diffusion process near these boundaries. It is obvious that

$$\sigma^2_{\theta,\theta}(\theta,\gamma)|_{\theta=\pi/2,0}=\tfrac{1}{8}\kappa^2S_{\xi}(2\omega)+\tfrac{1}{8}\kappa^2S_{\xi}(0)[1-\cos(2\gamma)]$$

It can easily be seen from the expressions for the drift term $\mu(\theta,\gamma)$,

$$\mu(\theta,\gamma)\sim\frac{1}{8}\kappa^2\{S_{\xi}(2\omega)+S_{\xi}(0)[1+\cos(2\gamma)]\}\frac{1}{2\theta},\quad \theta\to 0^+$$

$$\mu(\theta,\gamma)\sim\frac{1}{8}\kappa^2\{S_{\xi}(2\omega)+S_{\xi}(0)[1+\cos(2\gamma)]\}\frac{-1}{2\left(\frac{\pi}{2}-\theta\right)},$$

$$\theta\to\frac{\pi^-}{2}$$

Therefore, the diffusion process near the boundary is characterized by the drift term. It follows, therefore, that for $\kappa\neq 0$, the trajectories close to the boundary at $\theta=0$ are pushed away from the boundary due to the positive drift. Similarly, near the boundary at $\theta=\pi/2$, the trajectories near the boundary $\theta=\pi/2$ are pushed away from the boundary by the negative drift. Thus for the case when both the degrees of freedom are coupled, i.e., $p_{12}\neq 0$ and $p_{21}\neq 0$, the boundaries at $\theta=0$ and $\theta=\pi/2$ are not attainable states. Also, trajectories starting at either of these boundaries are pushed into the interior of the domain D. Moreover, this assertion can also be justified from transformation (4) that $\theta=0$ implies the amplitude of the first oscillator a_1, is zero and $\theta=\pi/2$ implies the amplitude of the second oscillator a_2 is zero. It is clear physically that unless the coupling coefficients p_{12} and p_{21} are both zero, it is not possible to have a solution with either a_1 or a_2 identically zero. For the remaining part of this study, we shall make the assumption that $\kappa\neq 0$. This assumption is valid and appropriate for most of the practical situations to which these results can be applied. The case $\kappa=0$ corresponds to a set of uncoupled oscillators and, therefore, is not of interest to us here. The results deduced above imply that the probability flux in a direction normal to the boundaries at $\theta=0$ and $\theta=\pi/2$ is identically equal to zero. From the above observation and based on our earlier work [3] for the noncommensurable case, we *assume a Neumann boundary condition* for ψ_0.

4 Eigenvalue Problem

As in [3,12,13], the solution of (22) can be calculated from an orthogonal expansion. The nature of the coefficients of the equation suggests that a double Fourier series is appropriate. Since ψ_0 is 2π-periodic in γ and because of the assumed Neumann boundary conditions

$$\frac{\partial\psi_0}{\partial\theta}(\gamma,0)=\frac{\partial\psi_0}{\partial\theta}\left(\gamma,\frac{\pi}{2}\right)=0 \tag{24}$$

we may express ψ_0 as follows:

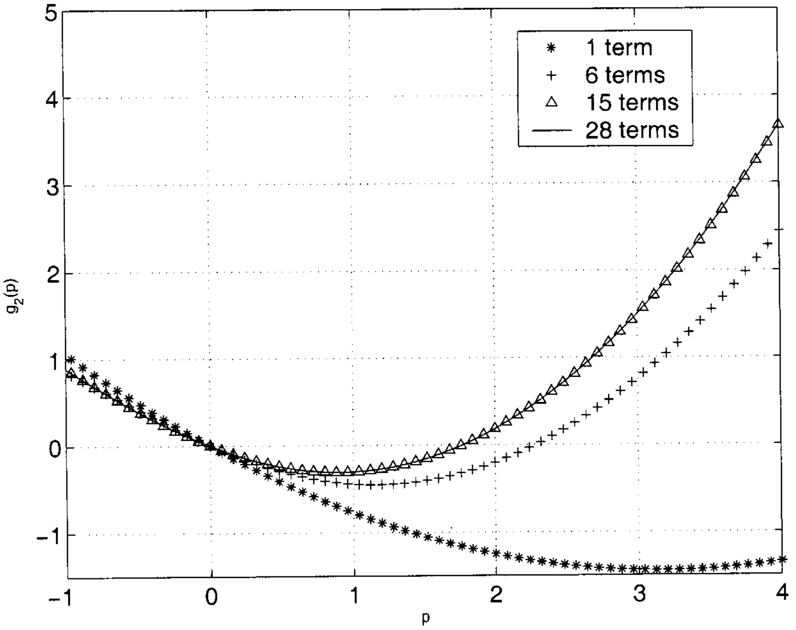


Fig. 1 Variation of moment Lyapunov exponent, $g_2(p)$ with p

$$\psi_0(\gamma,\theta)=a_{00}+\sum_{n=1}^{\infty} a_{0n}\cos 2n\theta+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}(a_{mn}\cos m\gamma + b_{mn}\sin m\gamma)\cos 2n\theta \qquad (25)$$

Insertion of (25) into (22) leads to the following equation for the coefficients:

$$\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\begin{pmatrix}\hat{C}_{rsmn} & \overline{C}_{rsmn}\\ \hat{D}_{rsmn} & \overline{D}_{rsmn}\end{pmatrix}\begin{pmatrix}a_{rs}\\ b_{rs}\end{pmatrix}=g(p)\begin{pmatrix}a_{mn}\\ b_{mn}\end{pmatrix} \qquad (26)$$

where

$$\hat{C}_{rsmn}\overset{\text{def}}{=}\int_0^{\pi/2}\int_0^{2\pi}L(p)(\cos r\gamma\cos 2s\theta)\times\cos m\gamma\cos 2n\theta d\gamma d\theta$$

$$\overline{C}_{rsmn}\overset{\text{def}}{=}\int_0^{\pi/2}\int_0^{2\pi}L(p)(\sin r\gamma\cos 2s\theta)\times\cos m\gamma\cos 2n\theta d\gamma d\theta$$

$$\hat{D}_{rsmn}\overset{\text{def}}{=}\int_0^{\pi/2}\int_0^{2\pi}L(p)(\cos r\gamma\cos 2s\theta)\times\sin m\gamma\cos 2n\theta d\gamma d\theta$$

$$\overline{D}_{rsmn}\overset{\text{def}}{=}\int_0^{\pi/2}\int_0^{2\pi}L(p)(\sin r\gamma\cos 2s\theta)\times\sin m\gamma\cos 2n\theta d\gamma d\theta$$

The existence of a nontrivial solution to (26) requires that the determinant of the coefficient matrix be zero. Thus to evaluate $g(p)$, the leading eigenvalue of the coefficient matrix, we construct a sequence of approximations by truncating the sums. Consider the truncated system

$$\sum_{r=0}^{M}\sum_{s=0}^{N}\begin{pmatrix}\hat{C}_{rsmn} & \overline{C}_{rsmn}\\ \hat{D}_{rsmn} & \overline{D}_{rsmn}\end{pmatrix}\begin{pmatrix}a_{rs}\\ b_{rs}\end{pmatrix}=g(p)\begin{pmatrix}a_{mn}\\ b_{mn}\end{pmatrix} \qquad (27)$$

We approximate the solution to (26) by numerically solving the truncated Eq. (27) for $g(p)$ with $M=N$

$$\sum_{r=0}^{N}\sum_{s=0}^{N}\begin{pmatrix}\hat{C}_{rsmn} & \overline{C}_{rsmn}\\ \hat{D}_{rsmn} & \overline{D}_{rsmn}\end{pmatrix}\begin{pmatrix}a_{rs}\\ b_{rs}\end{pmatrix}=g(p)\begin{pmatrix}a_{mn}\\ b_{mn}\end{pmatrix} \qquad (28)$$

Let **A** and **B** represent the following matrices:

$$\mathbf{A}=\begin{bmatrix}a_{00} & a_{01} & \cdots & a_{0N}\\ a_{10} & a_{11} & \cdots & a_{1N}\\ \vdots & \vdots & \ddots & \vdots\\ a_{N0} & a_{N1} & \cdots & a_{NN}\end{bmatrix}=\begin{bmatrix}\vdots & \vdots & \vdots & \vdots\\ A_0 & A_1 & \vdots & A_N\\ \vdots & \vdots & \vdots & \vdots\end{bmatrix}$$

$$\mathbf{B}=\begin{bmatrix}b_{10} & b_{11} & \cdots & b_{1N}\\ b_{20} & b_{21} & \cdots & b_{2N}\\ \vdots & \vdots & \ddots & \vdots\\ b_{N0} & b_{N1} & \cdots & b_{NN}\end{bmatrix}=\begin{bmatrix}\vdots & \vdots & \vdots & \vdots\\ B_0 & B_1 & \vdots & B_N\\ \vdots & \vdots & \vdots & \vdots\end{bmatrix}$$

Then (28) may be written as the following $(2N+1)(N+1)$ dimensional system:

$$\underline{A}\mathbf{r}=g(p)\mathbf{r} \qquad (29)$$

where

$$\mathbf{r}=[A_0,A_1,\ldots,A_N,B_0,B_1,\ldots,B_N]^T \qquad (30)$$

Thus to evaluate $g(p)$, the leading eigenvalue of $\underline{A}$, we construct a sequence of approximations by finding the eigenvalues of a sequence of submatrices. The set of approximate eigenvalues obtained by this procedure converges to the corresponding true eigenvalues as $N\rightarrow\infty$. However, the amount of calculation increases drastically with the increase in the number of terms considered.

5 Numerical Results

We now present some numerical results to illustrate the convergence of various orders of approximations. For this purpose, we consider the numerical values $p_{11}=1$, $p_{22}=2$, $\kappa=1$, $\delta_1=1$, $\delta_2=2$, $S_\xi(2\omega)=S_\xi(0)=1$, $T_\xi(2\omega)=0$. We obtain several sets of results corresponding to $N=0$, 1, 2, 3 giving rise, respectively, to 1, 6, 15, and 28 terms in the Fourier expansion. We compare the 6-, 15-, and 28-term expansions for the fourth-order eigenvalue approximations. From the results shown in Fig. 1, it is clear that the results for $N=2$ (15 terms) and $N=3$ (28 terms) are essentially the same. Thus, it suffices to use a 15-term expansion in order to obtain a sufficiently accurate expression for $g_2(p)$. For the parameter values given above, the system is almost-surely stable, but moment unstable beyond $p\approx 1.77$.

In conclusion, a method to compute the moment Lyapunov exponent of a 2-dof coupled linear system with commensurable eigenvalues, under random parametric excitation was developed. We derive the generator $L(p)$ for finite p, whose principal eigenvalue is the moment Lyapunov exponent, based on stochastic averaging. Except for some special cases the general solution of (22) cannot be obtained explicitly for $g_2(p)$. In this paper we have obtained approximate solutions based on Fourier analysis, and it is shown that a 15-term expansion is sufficiently accurate.

The results obtained above can be directly applied to analyze the stochastic stability of a number of nongyroscopic mechanical, structural, and diverse other physical systems with symmetry. These include for instance, surface wave excitations of fluid in a container with nearly square cross-section, transverse motion of a flexible spinning disc with time-dependent (random) variation in the spin rate, randomly loaded shallow arch structures, and beams with a rectangular cross-section and excited stochastically at the points of support along the longitudinal direction of the beam.

Acknowledgments

The authors would like to acknowledge the support of the National Science Foundation under Grant No. CMS 03-01412, and NSERC of Canada. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendix

In this appendix, the explicit expressions for the coefficients Q_{ψ_1}, $\hat{Q}_{\psi_1}$, μ_{ψ_1}, $\hat{\mu}_{\psi_1}$, ν_{ψ_1}, $\hat{\nu}_{\psi_1}$, $\sigma^2_{\theta\theta}$, $\sigma^2_{\gamma_1\gamma_1}$, $\sigma^2_{\theta\gamma_1}$ in (21) are evaluated under a specific resonance condition $\omega=\omega_1=\omega_2$.

$$\sigma_{\theta\gamma}\overset{\text{def}}{=}\tfrac{1}{16}\omega[\{p_{21}^{-}p_{21}^{+}S_\xi(0)[1+\cos^2(2\theta)]+(p_{21}^{+2}+p_{21}^{-2})S_\xi(0)\cos(2\theta)\}\csc(2\theta)-2(p_{11}-p_{22})S_\xi(0)[p_{21}^{+}+p_{21}^{-}\cos(2\theta)]]\sin\left(\frac{\gamma_1}{\omega}\right)$$

$$\begin{aligned}\sigma_{\gamma\gamma}\overset{\text{def}}{=}&\frac{1}{8}\omega^2\left\{S_\xi(0)+S_\xi(2\omega)+S_\xi(0)\cos\left(2\frac{\gamma_1}{\omega}\right)\right\}\cdot(p_{21}^{-2}+2p_{21}^{+}p_{21}^{-}\cos(2\theta)+p_{21}^{+2}\cos^2(2\theta))\csc^2(2\theta)-\frac{1}{4}\omega^2(p_{11}-p_{22})\{2S_\xi(0)\\&+S_\xi(2\omega)\}[p_{21}^{-}+p_{21}^{+}\cos(2\theta)]\cos\left(\frac{\gamma_1}{\omega}\right)\csc(2\theta)+\frac{1}{8}\omega^2\left[(p_{11}^2+p_{22}^2)S_\xi(2\omega)+2(p_{11}-p_{22})^2S_\xi(0)-2p_{11}p_{22}S_\xi(2\omega)\cos\left(2\frac{\gamma_1}{\omega}\right)\right]\end{aligned}$$

$$\begin{aligned}\sigma_{\theta\theta}\overset{\text{def}}{=}&\frac{1}{32}\left[(p_{21}^{+2}-p_{11}^2-p_{22}^2)S_\xi(2\omega)+p_{21}^{-2}S_\xi(0)-(p_{21}^{-2}S_\xi(0)-2p_{11}p_{22}S_\xi(2\omega))\cos\left(2\frac{\gamma_1}{\Omega}\right)\right]\cos^2(2\theta)+\frac{1}{16}p_{21}^{-}p_{21}^{+}\left\{[S_\xi(0)+S_\xi(2\omega)]\right.\\&\left.-S_\xi(0)\cos\left(2\frac{\gamma_1}{\Omega}\right)\right\}\cos(2\theta)+\frac{1}{32}\left[(p_{21}^{-2}+p_{11}^2+p_{22}^2)S_\xi(2\omega)+p_{21}^{+2}S_\xi(0)-(2p_{11}p_{22}S_\xi(2\omega)+p_{21}^{+2}S_\xi(0))\cos\left(2\frac{\gamma_1}{\Omega}\right)\right]\\&-\frac{1}{16}(p_{11}-p_{22})S_\xi(2\omega)\cos\left(\frac{\gamma_1}{\Omega}\right)[p_{21}^{+}\cos(2\theta)+p_{21}^{-}]\sin(2\theta)\end{aligned}$$

$$\begin{aligned}\nu_{\psi_1}\overset{\text{def}}{=}&-\frac{1}{16}\omega S_\xi(0)\{(2p_{21}^{+2}+p_{21}^{-2})+4p_{21}^{+}p_{21}^{-}\cos(2\theta)+p_{21}^{-2}\cos^2(2\theta)\}\sin\left(2\frac{\gamma_1}{\omega}\right)\csc^2(2\theta)+\frac{1}{8}\omega\left[(p_{11}+p_{22})[p_{21}^{+}\cos(2\theta)\right.\\&\left.+p_{21}^{-}]T_\xi(2\omega)\cos\left(\frac{\gamma_1}{\omega}\right)+\{(p_{11}-p_{22})S_\xi(0)[p_{21}^{+}\cos(2\theta)+p_{21}^{-}]-(p_{11}+p_{22})S_\xi(2\omega)[p_{21}^{+}+p_{21}^{-}\cos(2\theta)]\}\sin\left(\frac{\gamma_1}{\omega}\right)\right]\csc(2\theta)\\&+\frac{1}{16}\omega\left[p_{21}^{+2}S_\xi(0)\sin\left(2\frac{\gamma_1}{\omega}\right)-2(p_{11}^2-p_{22}^2)T_\xi(2\omega)\right]\end{aligned}$$

$$\begin{aligned}\hat{\nu}_{\psi_1}\overset{\text{def}}{=}&-\frac{1}{16}\omega\{2p_{21}^{-}(p_{11}-p_{22})S_\xi(0)\sin^2(2\theta)+(p_{11}-p_{22})S_\xi(2\omega)[p_{21}^{+}\cos(2\theta)+p_{21}^{-}]+(p_{11}+p_{22})S_\xi(2\omega)[p_{21}^{+}\\&+p_{21}^{-}\cos(2\theta)]\}\sin\left(\frac{\gamma_1}{\omega}\right)\csc(2\theta)+\frac{1}{16}\omega\{[p_{21}^{-2}S_\xi(0)-2p_{11}p_{22}S_\xi(2\omega)]+p_{21}^{+}p_{21}^{-}S_\xi(0)\cos(2\theta)\}\sin\left(2\frac{\gamma_1}{\omega}\right)\end{aligned}$$

$$\begin{aligned}\hat{\mu}_{\psi_1}\overset{\text{def}}{=}&\frac{1}{32}\left[(p_{21}^{+2}-p_{11}^2-p_{22}^2)S_\xi(2\omega)+p_{21}^{-2}S_\xi(0)+(2p_{11}p_{22}S_\xi(2\omega)-p_{21}^{-2}S_\xi(0))\cos\left(2\frac{\gamma_1}{\omega}\right)\right]\cos(2\theta)\sin(2\theta)+\frac{1}{32}\left[p_{21}^{+}p_{21}^{-}\left\{S_\xi(2\omega)\right.\right.\\&\left.\left.+S_\xi(0)-S_\xi(0)\cos\left(2\frac{\gamma_1}{\omega}\right)\right\}-S_\xi(2\omega)(p_{11}^2-p_{22}^2)\right]\sin(2\theta)+\frac{1}{32}S_\xi(2\omega)[2p_{21}^{+}(p_{11}-p_{22})\cos^2(2\theta)+\{p_{21}^{-}(p_{11}-p_{22})\\&+p_{21}^{+}(p_{11}+p_{22})\}\cos(2\theta)+p_{21}^{-}(p_{11}+p_{22})-p_{21}^{+}(p_{11}-p_{22})]\cos\left(\frac{\gamma_1}{\omega}\right)\end{aligned}$$

$$
\begin{aligned}
\mu_{\psi_1} \overset{\text{def}}{=} & \Bigg[\frac{1}{32}\left\{ (p_{21}^{+2} - p_{11}^2 - p_{22}^2) S_\xi(2\omega) + p_{21}^{-2} S_\xi(0) - (p_{21}^{-2} S_\xi(0) - 2 p_{11} p_{22} S_\xi(2\omega)) \cos\left(2\frac{\gamma_1}{\omega} \right) \right\} \cos^2(2\theta) + \frac{1}{16}\Bigg\{ p_{21}^{-} p_{21}^{+} [S_\xi(0) + S_\xi(2\omega)] \\
& - p_{21}^{-} p_{21}^{+} S_\xi(0) \cos\left(2\frac{\gamma_1}{\omega} \right) \Bigg\} \cos(2\theta) + \frac{1}{32}\left\{ (p_{11}^2 + p_{21}^{-2} + p_{22}^2) S_\xi(2\omega) + p_{21}^{+2} S_\xi(0) - (2 p_{11} p_{22} S_\xi(2\omega) + p_{21}^{+2} S_\xi(0)) \cos\left(2\frac{\gamma_1}{\omega} \right) \right\} \\
& - \frac{1}{16}(p_{11} - p_{22}) S_\xi(2\omega) \cos\left(\frac{\gamma_1}{\omega} \right) \sin(2\theta)(p_{21}^{+} \cos(2\theta) + p_{21}^{-}) \Bigg] \cot(2\theta) - \frac{1}{16}\left[(p_{11}^2 - p_{22}^2) S_\xi(2\omega) + p_{21}^{-} p_{21}^{+} S_\xi(0) \cos\left(2\frac{\gamma_1}{\omega} \right) \right] \sin(2\theta) \\
& + \frac{1}{16}[(p_{21}^{+2} + p_{21}^{-2}) S_\xi(0) \cos(2\theta) + 2 p_{21}^{-} p_{21}^{+} S_\xi(0)] \cos\left(2\frac{\gamma_1}{\omega} \right) \csc(2\theta) + \frac{1}{16}\Bigg[\{ p_{21}^{+}(p_{11} + p_{22}) S_\xi(2\omega) - p_{21}^{-}(p_{11} \\
& - p_{22}) S_\xi(0) \} \cos\left(\frac{\gamma_1}{\omega} \right) + p_{21}^{-}(p_{11} + p_{22}) T_\xi(2\omega) \sin\left(\frac{\gamma_1}{\omega} \right) \Bigg] \cos(2\theta) + \frac{1}{16}\Bigg[\{ -p_{21}^{+}(p_{11} - p_{22}) S_\xi(0) + p_{21}^{-}(p_{11} + p_{22}) S_\xi(2\omega) \} \\
& \times \cos\left(\frac{\gamma_1}{\omega} \right) + p_{21}^{+}(p_{11} + p_{22}) T_\xi(2\omega) \sin\left(\frac{\gamma_1}{\omega} \right) \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{Q}_{\psi_1} \overset{\text{def}}{=} & \frac{1}{32}\left[(p_{21}^{+2} - p_{11}^2 - p_{22}^2) S_\xi(2\omega) + p_{21}^{-2} S_\xi(0) - (p_{21}^{-2} S_\xi(0) - 2 p_{11} p_{22} S_\xi(2\omega)) \cos\left(2\frac{\gamma_1}{\omega} \right) \right] \cos^2(2\theta) + \frac{1}{16} p_{21}^{-} p_{21}^{+} \Bigg(S_\xi(0) + S_\xi(2\omega) \\
& - S_\xi(0) \cos\left(2\frac{\gamma_1}{\omega} \right) \Bigg) \cos(2\theta) + \frac{1}{32}[(p_{11}^2 + p_{21}^{-2} + p_{22}^2) S_\xi(2\omega) + p_{21}^{+2} S_\xi(0)] - \frac{1}{32}(2 p_{11} p_{22} S_\xi(2\omega) + p_{21}^{+2} S_\xi(0)) \cos\left(2\frac{\gamma_1}{\omega} \right) \\
& - \frac{1}{16}(p_{11} - p_{22}) S_\xi(2\omega) \cos\left(\frac{\gamma_1}{\omega} \right) [p_{21}^{+} \cos(2\theta) + p_{21}^{-}] \sin(2\theta) + \frac{1}{16}\left((p_{11}^2 - p_{22}^2) S_\xi(2\omega) + p_{21}^{-} p_{21}^{+} S_\xi(0) \cos\left(2\frac{\gamma_1}{\omega} \right) \right) \cos(2\theta) \\
& + \frac{1}{32}(2 p_{11}^2 + 2 p_{22}^2 + p_{21}^{+2} - p_{21}^{-2}) S_\xi(2\omega) - \frac{1}{32}\left[(p_{21}^{+2} - p_{21}^{-2}) S_\xi(0) - (p_{21}^{+2} + p_{21}^{-2}) S_\xi(0) \cos\left(2\frac{\gamma_1}{\omega} \right) \right] \\
& + \frac{1}{16}\left[\{ p_{21}^{+}(p_{11} + p_{22}) S_\xi(2\omega) - p_{21}^{-}(p_{11} - p_{22}) S_\xi(0) \} \cos\left(\frac{\gamma_1}{\omega} \right) + p_{21}^{-}(p_{11} + p_{22}) T_\xi(2\omega) \sin\left(\frac{\gamma_1}{\omega} \right) \right] \sin(2\theta)
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mathbf{Q}}_{\psi_1} \overset{\text{def}}{=} & \frac{1}{16} p_{21}^{+} S_\xi(2\omega)[(p_{11} - p_{22}) \cos(2\theta) + (p_{11} + p_{22})] \cos\left(\frac{\gamma_1}{\omega} \right) \sin(2\theta) + \frac{1}{32}\Bigg[(p_{21}^{-2} S_\xi(0) - 2 p_{11} p_{22} S_\xi(2\omega)) \cos\left(2\frac{\gamma_1}{\omega} \right) \\
& + (p_{11}^2 + p_{22}^2 - p_{21}^{+2}) S_\xi(2\omega) - p_{21}^{-2} S_\xi(0) \Bigg] \cos^2(2\theta) + \frac{1}{16} S_\xi(2\omega)(p_{11}^2 - p_{22}^2) \cos(2\theta) + \frac{1}{32}(2 p_{11} p_{22} S_\xi(2\omega) \\
& - p_{21}^{-2} S_\xi(0)) \cos\left(2\frac{\gamma_1}{\omega} \right) + \frac{1}{32}[p_{21}^{-2} S_\xi(0) + (p_{11}^2 + p_{22}^2 + p_{21}^{+2}) S_\xi(2\omega)]
\end{aligned}
$$

References

[1] Namachchivaya, N. Sri, and Van Roessel, H. J., 1993, ‘‘Maximal Lyapunov Exponent and Rotation Numbers for Two Coupled Oscillators Driven by Real Noise,’’ J. Stat. Phys., **71**(3/4), pp. 549–567.

[2] Namachchivaya, N. Sri, Van Roessel, H. J., and Doyle, M. M., 1996, ‘‘Moment Lyapunov Exponent for Two Coupled Oscillators Driven by Real Noise,’’ SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **56**, pp. 1400–1423.

[3] Namachchivaya, N. Sri, and Van Roessel, H. J., 2001, ‘‘Moment Lyapunov Exponent and Stochastic Stability of Two Coupled Oscillators Driven by Real Noise,’’ ASME J. Appl. Mech., **68**, pp. 1400–1412.

[4] Molčanov, S. A., 1978, ‘‘The Structure of Eigenfunctions of One-Dimensional Unordered Structures,’’ Math. USSR, Izv., **12**(1), pp. 69–101.

[5] Arnold, L., 1984, ‘‘A Formula Connecting Sample and Moment Stability of Linear Stochastic Systems,’’ SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **44**(4), pp. 793–802.

[6] Arnold, L., Oeljeklaus, E., and Pardoux, E., 1986, ‘‘Almost Sure and Moment Stability for Linear Itô Equations,’’ *Lyapunov Exponents, Lecture Notes in Mathematics*, Vol. 1186, Springer-Verlag, New York, pp. 129–159.

[7] Arnold, L., Kliemann, W., and Oeljeklaus, E., 1986, ‘‘Lyapunov Exponents of Linear Stochastic Systems,’’ *Lyapunov Exponents, Lecture Notes in Mathematics*, Vol. 1186 Springer-Verlag, New York, pp. 85–125.

[8] Namachchivaya, N. Sri, Ramakrishnan, N., Van Roessel, H. J., and Vedula, L., 2003, ‘‘Stochastic Stability of Two Coupled Oscillators in Resonance: Averaging Approach,’’ *Nonlinear Stochastic Dynamics*, N. Sri Namachchivaya and Y. K. Lin, (eds.), Solid Mechanics and Its Applications, Vol. 110, pp. 167–178, Kluwer Dordrecht.

[9] Arnold, L., Doyle, M. M., and Namachchivaya, N. Sri, 1997, ‘‘Small Noise Expansion of Moment Lyapunov Exponents for General Two Dimensional Systems,’’ Dyn. Stab. Syst., **12**(3), pp. 187–211.

[10] Khasminskii, R. Z., and Moshchuk, N., 1998, ‘‘Moment Lyapunov Exponent and Stability Index for Linear Conservative System With Small Random Perturbation,’’ SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **58**(1), pp. 245–256.

[11] Pardoux, E., and Wihstutz, V., 1988, ‘‘Lyapunov Exponent and Rotation Number of Two-Dimensional Linear Stochastic Systems With Small Diffusion,’’ SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **48**(2), pp. 442–457.

[12] Bolotin, V. V., 1964, *The Dynamic Stability of Elastic Systems*, Holden-Day, San Francisco.

[13] Wedig, W. V., 1988, ‘‘Lyapunov Exponents of Stochastic Systems and Related Bifurcation Problems,’’ *Stochastic Structural Dynamics: Progress in Theory and Applications*, S. T. Ariaratnam, G. I. Schuëller, and I. Elishakoff, eds., Elsevier Applied Science, London, pp. 315–327.

Alan Wineman
Department of Mechanical Engineering,
University of Michigan,
Ann Arbor, MI 48109
Fellow ASME

John Shaw
Department of Aerospace Engineering,
University of Michigan,
Ann Arbor, MI 48109
Mem. ASME

A Correspondence Principle for Scission-Induced Stress Relaxation in Elastomeric Components

A method is presented for calculating the stress relaxation due to scission in elastomeric components that operate at a fixed deformation while at an elevated temperature. A relationship is established between stresses at different temperatures that is called the correspondence principle for scission/healing materials. Two examples involving cylinders illustrate its use. The first example involves combined tension-torsion, for which an axial force-twisting moment relation is derived, that might be useful in experimental studies to assess the applicability of the correspondence principle. The second example provides a criterion for estimating the lifetime of an annular seal.
[DOI: 10.1115/1.1794701]

1 Introduction

In applications where the mechanical and thermal loads on elastomeric structural components are benign enough that no changes in microstructure occur, stresses and deformations can be calculated using the nonlinear theory of elasticity. However, when the temperature (or deformation) of an elastomeric structural component is sufficiently large, scission of molecular cross-links and possible in situ recross-linking (healing) can result in significant time-dependent softening of mechanical properties as well as permanent set [1,2]. In applications involving elastomeric seals or bushings in automotive or truck suspension systems, for example, these changes can impair performance and require eventual replacement. In such applications, where it is important to be able to predict the lifetime of the elastomeric component, the nonlinear theory of elasticity is no longer applicable, and a new means for determining stresses is required.

In this paper, a method is presented which can be used to calculate the stress relaxation due to scission in elastomeric components that operate at a fixed deformation while at elevated temperatures as, for example, could occur in seals or bushings. The method is based on a correspondence that is established between stresses in an elastomeric component at different temperatures. Because an analogous situation in the linear theory of viscoelasticity has proven to be very useful [3], the method presented here is referred to as the *correspondence principle for scission/healing materials*.

The proposed *correspondence principle* is restricted to conditions when the deformation is fixed and the temperature is spatially uniform. There are applications when these conditions should be at least approximately satisfied. For example, if the surface temperature of a seal is increased, the time required for the temperature field within the seal to become uniform may be small compared to the time for there to be significant scission. Also, a seal may be subjected to a large initial deformation and then small superposed deformations as, for example, would occur if the seal were used in a vibrating pump. Thus, the conditions for the application of the *correspondence principle* would be at least approximately satisfied and would lead to a useful first approximation for the relaxation of stresses in an elastomeric component undergoing scission.

The constitutive theory that accounts for scission-induced stress relaxation in an elastomer at an arbitrary fixed deformation is presented in Sec. 2. The *correspondence principle for scission/healing materials* is developed in Sec. 3. Two examples involving cylinders are presented in Sec. 4. The first involves combined tension-torsion. A result relating the axial force and twisting moment is derived which might be useful in experimental studies. The second example provides a criterion for estimating the lifetime of a seal.

2 Constitutive Equation

Tobolsky [1] described experiments in which a rubber strip at room temperature was subjected to a fixed uniaxial stretch and then held at a higher fixed temperature for a specified time interval. At temperatures above T_{cr} (say 100 °C), called the chemorheological temperature, the stress decreased with time. At the end of the specified time interval, the external force was reduced to zero and the specimen was returned to its original temperature. The specimen was observed to have a permanent stretch. Tests were carried out for different applied stretches, temperatures, and time intervals. The decrease in tensile stress with time and the permanent stretch were measured. Results of more recent experiments can be found in the article by Wineman, Jones, and Shaw [2]. Tobolsky analyzed the data assuming the elastomer to be instantaneously neo-Hookean, for which the relation between tensile (Cauchy) stress $\sigma(t)$ and uniaxial stretch ratio λ is

$$\sigma(t)=2n(t)kT\left(\lambda^{2}-\frac{1}{\lambda}\right) \tag{1}$$

where T is the absolute temperature, k is the Boltzmann constant, and $n(t)$ is the current cross-link density. The decrease in $\sigma(t)$ was attributed to scission of molecular network cross-links, resulting in a decrease in $n(t)$. The permanent stretch was attributed to a new network that formed in the stretched state (healing). At temperatures below T_{cr}, the stress-stretch relation for the system consisting of the two networks was assumed to be

---

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS. Manuscript received by the Applied Mechanics Division, March 26, 2002; final revision, October 30, 2003. Associate Editor: K. R. Rajagopal. 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.

$$\sigma = 2n_1 kT\left(\lambda^2 - \frac{1}{\lambda}\right) + 2n_2 kT\left[\left(\frac{\lambda}{\hat{\lambda}}\right)^2 - \left(\frac{\hat{\lambda}}{\lambda}\right)\right] \tag{2}$$

where $\hat{\lambda}$ is the stretch ratio of the original network while held at the high temperature, n_1 is the cross-link density of the original network at the end of the test, and n_2 is the cross-link density of the new network. Equation (2) expresses the assumptions that i) the total stress is the sum of the stresses in each network, ii) each network acts as an incompressible isotropic neo-Hookean elastic material, and iii) broken cross-links reform to produce a new network that is stress free when the stretch ratio of the original network is $\hat{\lambda}$ [1,4].

Tobolsky's data suggested that $n(t)$ in (1) is independent of the stretch ratio $\hat{\lambda}$ up to a value of about 4. This was supported by the results of Scanlan and Watson [5]. According to Eq. (1), $\sigma(t)/\sigma(0) = \beta(T,t)$ and $n(t) = \beta(T,t)n_o$ where n_o is the initial cross-link density. $\beta(T,t)$ is a material property function that can be obtained experimentally (see [1], Fig. V.4). Tobolsky ([1], p. 226) suggested that $\beta(T,t)$ can be represented in the form

$$\beta(T,t) = \phi(\alpha(T)t) \tag{3}$$

for a number of elastomers. For a particular natural rubber vulcanizate in the temperature range $100\,^\circ\mathrm{C} \leq T \leq 130\,^\circ\mathrm{C}$, Tobolsky showed that

$$\beta(T,t) = \exp(-\alpha(T)t), \tag{4}$$

with

$$\alpha(T) = \frac{k_B}{k_P} T \exp(-E_{act}/RT) \tag{5}$$

k_B is Boltzmann's constant, k_P is Planck's constant, E_{act} is an activation energy whose value was found to be 30.4 kcal/mol, and R is the gas constant.

Neubert and Saunders [6] carried out tests similar to those of Tobolsky, but for a pure shear deformation. They measured permanent biaxial stretch upon removal of stress and reduction of the temperature to its original value, and found that predictions based on the assumption of a neo-Hookean response led to inaccurate predictions of permanent set. They modified assumption ii) by modeling the rubber as a Mooney-Rivlin material, and showed that this model led to better agreement with the measured permanent biaxial stretch. Fong and Zapas [7] later proposed using the Rivlin-Saunders model to determine the permanent biaxial stretch.

These results are now used as a guide for the development of a constitutive framework for the three-dimensional response of a rubber undergoing scission while at a fixed homogeneous deformation and constant temperature history. For a detailed discussion of the constitutive equation, see Wineman and Shaw [8]. Consider a rubbery material in a stress-free reference configuration at a temperature T. There is a range of deformations and temperatures for which the material response can be regarded as incompressible, isotropic, and nonlinearly elastic. If $\mathbf{x}$ is the position at current time t of a particle located at $\mathbf{X}$ in the reference configuration, the deformation gradient is $\mathbf{F} = \partial\mathbf{x}/\partial\mathbf{X}$. The left Cauchy-Green tensor is $\mathbf{B} = \mathbf{F}\mathbf{F}^T$ and the Cauchy stress $\boldsymbol{\sigma}$ is given by

$$\boldsymbol{\sigma} = -p\mathbf{I} + 2\frac{\partial W}{\partial I_1}\mathbf{B} - 2\frac{\partial W}{\partial I_2}\mathbf{B}^{-1} \tag{6}$$

where p is an arbitrary hydrostatic pressure arising from the constraint that deformations are isochoric. I_1, I_2 are the first and second invariants of $\mathbf{B}$, respectively, and $W(I_1, I_2, T)$ is the strain energy density associated with the original material. In Eq. (6), $\boldsymbol{\sigma}$,$\mathbf{B}$, and T are evaluated at the current time t, which is omitted from the notation for brevity. For many proposed models of rubber elasticity, the strain energy density function is written as $W(I_1, I_2, T) = n_o kT W^o(I_1, I_2)$, that is, the dependence on temperature and deformation is separable. Note that the subscript or superscript o indicates the scission-independent part of a quantity. This is the case for the phantom model, affine model, constrained chain model, localization model, liquidlike model, and eight-chain model [9]. Accordingly, Eq. (6) can be restated as

$$\boldsymbol{\sigma} = -p\mathbf{I} + \boldsymbol{\sigma}^o(\mathbf{B}, T) \tag{7a}$$

where

$$\boldsymbol{\sigma}^o(\mathbf{B}, T) = 2n_o kT\mathbf{S}^o(\mathbf{B}), \quad \mathbf{S}^o(\mathbf{B}) = \frac{\partial W^o}{\partial I_1}\mathbf{B} - \frac{\partial W^o}{\partial I_2}\mathbf{B}^{-1}. \tag{7b}$$

For temperatures $T < T_{cr}$ and moderate deformations, no microstructural changes are assumed to occur, and the stress is given by Eq. (6), $(7a)$, or $(7b)$. If the material is held at a fixed homogeneous deformation and the temperature is increased to a fixed value $T \geq T_{cr}$ at time $t = 0$, scission of the original microstructural network is assumed to occur continuously in time. The volume fraction of the original network cross-link density at time t is denoted as $\beta(T,t)$, a monotonically decreasing function of t satisfying $\beta(T,0) = 1$.

The current stress is given by

$$\boldsymbol{\sigma} = -p\mathbf{I} + \beta\left[2\frac{\partial W}{\partial I_1}\mathbf{B} - 2\frac{\partial W}{\partial I_2}\mathbf{B}^{-1}\right], \tag{8a}$$

or, alternatively,

$$\boldsymbol{\sigma} = -p\mathbf{I} + 2\beta n_o kT\left[\frac{\partial W^o}{\partial I_1}\mathbf{B} - \frac{\partial W^o}{\partial I_2}\mathbf{B}^{-1}\right], \tag{8b}$$

where β and $\boldsymbol{\sigma}$ are evaluated at the current time t. This constitutive equation extends the ideas inherent in Eq. (1) to arbitrary homogeneous deformations. Neubert and Saunders [6] used it in their analysis of the permanent set due to new networks that formed during a pure shear deformation.

Several comments are in order regarding constitutive Eq. (8). First, in accordance with Tobolsky's experimental results, $\beta(T,t)$ is assumed to be independent of the deformation. This is strictly justified only for fixed uniaxial extensions with $\lambda < 4$. There is a lack of experimental evidence for other deformations. Second, although Tobolsky assumed the response of the original and newly formed networks to be neo-Hookean, Neubert, and Saunders [6] and Fong and Zapas [7] considered other possibilities. Thus, $W^o(I_1, I_2)$ is left unspecified. Third, consistent with assumption iii) above, new networks that result from cross-linking are formed in a stress-free state. Provided the deformation is held fixed, these new networks do not contribute to the stress and no further constitutive assumptions are required.

3 Correspondence Principle for Scission-Healing Materials

Consider an elastomeric body that has been subjected to a nonhomogeneous deformation and is in equilibrium at a spatially uniform temperature $T_o < T_{cr}$. Let its deformed configuration be denoted by κ. Surface tractions are specified on the portion of the deformed surface denoted as $\partial\kappa^{(\sigma)}$ and the current positions of particles are specified on the portion of the deformed surface denoted as $\partial\kappa^{(d)}$. Let $\hat{\mathbf{x}}$ denote the prescribed current particle positions on $\partial\kappa^{(d)}$, $\hat{\mathbf{T}}^{(\sigma)}$ denote the prescribed surface traction on $\partial\kappa^{(\sigma)}$, and $\hat{\mathbf{T}}^{(d)}$ denote the computed surface traction on $\partial\kappa^{(d)}$. The stress and deformation fields satisfy the following conditions:

$$\operatorname{div}\boldsymbol{\sigma} = \mathbf{0} \quad \text{in} \quad \kappa, \tag{9a}$$

$$\boldsymbol{\sigma}\mathbf{n} = \hat{\mathbf{T}}^{(\sigma)} \quad \text{on} \quad \partial\kappa^{(\sigma)}, \tag{9b}$$

$$\mathbf{x} = \hat{\mathbf{x}} \quad \text{on} \quad \partial\kappa^{(d)} \tag{9c}$$

where $\mathbf{n}$ denotes the unit outer normal at a point of the external surface. The constitutive equation is given by Eqs. $(7a)$ and $(7b)$, which when substituted into Equation $(9a)$, gives

$$-\operatorname{grad} p_o + 2n_o k T_o \operatorname{div} \mathbf{S}^o = \mathbf{0} \tag{10}$$

Boundary condition $(9b)$ with $(7b)$ can be written in the form

$$-p_o \mathbf{n} + 2n_o k T_o \mathbf{S}^o \mathbf{n} = \hat{\mathbf{T}}^{(\sigma)} \tag{11}$$

Equations (9*c*), (10), and (11) define a boundary value problem for the scalar field $p_o(\mathbf{X})$ and deformation $\mathbf{x}_o(\mathbf{X})$. The corresponding stresses are given by

$$\boldsymbol{\sigma}_o = -p_o \mathbf{I} + 2n_o k T_o \mathbf{S}^o(\mathbf{B}_o) \tag{12}$$

where $\mathbf{B}_o$ is calculated from $\mathbf{x}_o(\mathbf{X})$. The surface tractions $\hat{\mathbf{T}}^{(d)}$ on $\partial\kappa^{(d)}$ are

$$\hat{\mathbf{T}}^{(d)} = \boldsymbol{\sigma}_o \mathbf{n} = -p_o \mathbf{n} + 2n_o k T_o \mathbf{S}^o(\mathbf{B}_o)\mathbf{n} \tag{13}$$

Now, suppose the body is brought to a higher, spatially uniform temperature $T_1 > T_{cr}$ and is in equilibrium at a fixed deformation. According to various researchers (e.g., [1,5]), the volume changes associated with the temperature change of interest and the process of scission and subsequent reforming of cross-links are small enough to be neglected. It is thus assumed that the body has the same deformation as when at the lower temperature $T_o < T_{cr}$, that is $\mathbf{x} = \mathbf{x}_o(\mathbf{X})$. Constitutive Eq. $(8b)$ gives

$$\boldsymbol{\sigma}_1 = -p_1 \mathbf{I} + 2\beta(T_1, t) n_o k T_1 \mathbf{S}^o(\mathbf{B}_o) \tag{14}$$

Equilibrium condition $(9a)$ becomes, using Eq. (14),

$$\operatorname{div} \boldsymbol{\sigma}_1 = -\operatorname{grad} p_1 + 2\beta(T_1, t) n_o k T_1 \operatorname{div} \mathbf{S}^o(\mathbf{B}_o) = \mathbf{0} \tag{15}$$

Since $p_o(\mathbf{X})$ and $\mathbf{x}_o(\mathbf{X})$ satisfy Eq. (10), Eq. (15) becomes

$$\begin{aligned} \operatorname{div} \boldsymbol{\sigma}_1 &= -\operatorname{grad} p_1 + \beta(T_1, t) T_1 / T_o \operatorname{grad} p_o \\ &= \operatorname{grad}(-p_1 + p_o \beta(T_1, t) T_1 / T_o) = \mathbf{0} \end{aligned} \tag{16}$$

If we let $p_1 = p_o(\mathbf{X}) \beta(T_1, t) / T_1 / T_o$, the equilibrium equation is satisfied.

The corresponding stress is found from Eq. (14),

$$\boldsymbol{\sigma}_1 = \beta(T_1, t) T_1 / T_o [-p_o \mathbf{I} + 2n_o k T_o \mathbf{S}^o(\mathbf{B}_o)] \tag{17}$$

Boundary condition $(9c)$ is automatically satisfied because of the assumed deformation. The tractions on the deformed external surface are calculated using Eq. (17),

$$\boldsymbol{\sigma}_1 \mathbf{n} = \beta(T_1, t) T_1 / T_o [-p_o \mathbf{n} + 2n_o k T_o \mathbf{S}^o(\mathbf{B}_o)\mathbf{n}] \tag{18}$$

Evaluating Eq. (18) on $\partial\kappa^{(\sigma)}$ gives $\beta(T_1, t) T_1 / T_o \hat{\mathbf{T}}^{(\sigma)}$ and on $\partial\kappa^{(d)}$ gives $\beta(T_1, t) T_1 / T_o \hat{\mathbf{T}}^{(d)}$.

The results of this section establish the following:

Correspondence Principle for Scission/Healing Materials

Let $\mathbf{x} = \mathbf{x}_o(\mathbf{X})$ be an equilibrium deformation for an incompressible isotropic elastomeric body at a spatially uniform temperature $T_o < T_{cr}$ and let the corresponding stress field be denoted as $\boldsymbol{\sigma}_o(\mathbf{X})$. Then $\mathbf{x} = \mathbf{x}_o(\mathbf{X})$ is also an equilibrium deformation when the body is brought to a higher, spatially uniform temperature $T_1 > T_{cr}$, where it undergoes the scission-recross-linking process. The corresponding stresses are $\boldsymbol{\sigma}_1(\mathbf{X}, t) = \beta(T_1, t) T_1 / T_o \boldsymbol{\sigma}_o(\mathbf{X})$, where $\beta(T_1, t)$ is the material's scission response function, i.e., the ratio of the current to the original cross-link density for the original material. If the surface tractions are $\hat{\mathbf{T}}^{(\sigma)}$ on $\partial\kappa^{(\sigma)}$ and $\hat{\mathbf{T}}^{(d)}$ on $\partial\kappa^{(d)}$ at $T_o < T_{cr}$, then at $T_1 > T_{cr}$ the surface tractions are $\beta(T_1, t) T_1 / T_o \hat{\mathbf{T}}^{(\sigma)}$ and $\beta(T_1, t) T_1 / T_o \hat{\mathbf{T}}^{(d)}$, respectively.

4 Applications of the Correspondence Principle for Scission/Healing Materials

In this section, two examples are presented to examine the consequences of the *correspondence principle for scission/healing materials*. The first example discusses a nontrivial multiaxial deformation state that could lead to a nice experimental assessment of the validity of the proposed correspondence principle. The second example is a practical application of the correspondence principle, showing a method for determining the lifetime of an elastomeric seal at an elevated temperature.

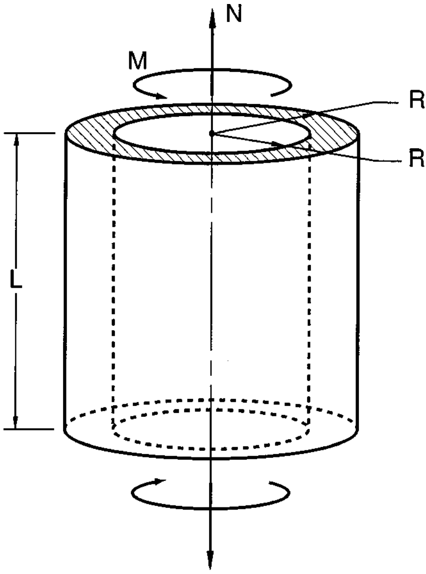


Fig. 1 Hollow cylinder subjected to axial force and torsion

Combined Tension-Torsion of a Circular Cylinder. Consider a circular cylinder at a uniform temperature $T_o < T_{cr}$ with length L, inner radius R_i, and outer radius R_o. The inner and outer cylindrical surfaces are traction free and axial force N and twisting moment M are applied to its end surfaces (see Fig. 1). The cylinder is in equilibrium under these applied loads.

The resulting deformation is assumed to be axially symmetric, in which plane cross sections remain plane, displace along and rotate about the axis of symmetry, and cylindrical surfaces deform into cylindrical surfaces. Let a cylindrical coordinate system be introduced that is coaxial with the cylinder and has its origin at one end. A material point at (R, Θ, Z) in the reference configuration deforms to (r, Θ, z) in the current configuration. The mapping describing this deformation has the form

$$\begin{aligned} r &= \left[\frac{1}{\lambda} R^2 + \gamma\right]^{1/2} \\ \theta &= \Theta + \psi\lambda Z \\ z &= \lambda Z \end{aligned} \tag{19}$$

λ and ψ are constants that represent the uniform axial stretch ratio and uniform cross-sectional rotation per current length, respectively. If r_i and r_o are the radii of the deformed inner and outer surfaces, then

$$\frac{1}{\lambda} = \frac{r_o^2 - r_i^2}{R_o^2 - R_i^2}, \qquad \gamma = \frac{R_o^2 r_i^2 - R_i^2 r_o^2}{R_o^2 - R_i^2} \tag{20}$$

Consider a possible experiment in which λ and ψ are specified and the cylindrical surfaces are traction free in the current configuration. Using the notation of Sec. 3, the inner and outer cylindrical surfaces form $\partial\kappa^{(\sigma)}$ and $\hat{\mathbf{T}}^{(\sigma)} = \mathbf{0}$. The ends of the cylinder, $z = 0$ and $z = \lambda L$, form the surface $\partial\kappa^{(d)}$. $\hat{\mathbf{x}}$ is obtained by evaluating the mapping in Eq. (19) at $Z = 0$ and $Z = L$.

An analysis of the combined torsion and tension of a circular cylinder can be found in [10]. A scalar field $p_o(r)$ can be found so that the equilibrium equation is met. The radii r_i and r_o of the deformed cylindrical surfaces are determined from the first of Eq. (20) and the boundary condition that $\hat{\mathbf{T}}^{(\sigma)} = \mathbf{0}$ on $\partial\kappa^{(\sigma)}$. The constant γ is then known. Expressions for the stress are presented in

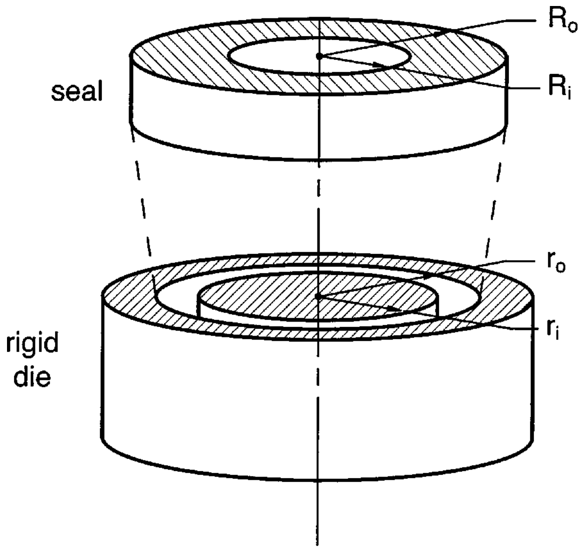


Fig. 2 Elastomeric seal and rigid die

[9] but are omitted here for brevity. Of particular interest here are the expressions for the axial force and twisting moment applied the ends of the cylinder,

$$N=2\pi\int_{r_i}^{r_o}\sigma_{zz}r\,dr \tag{21a}$$

$$M=2\pi\int_{r_i}^{r_o}\sigma_{z\theta}r^2\,dr \tag{21b}$$

Suppose that the temperature of the cylinder is increased to $T_1>T_{cr}$. According to the correspondence principle, the deformation $\mathbf{x}=\mathbf{x}_o(\mathbf{X})$ is given by Eq. (19). The traction on $\partial\kappa^{(\sigma)}$ is $\beta(T_1,t)T_1/T_o\hat{\mathbf{T}}^{(\sigma)}=\mathbf{0}$. On $\partial\kappa^{(d)}$, $\sigma_{zz}|_{T_1}=\beta(T_1,t)T_1/T_o\sigma_{zz}|_{T_0}$, and $\sigma_{z\theta}|_{T_1}=\beta(T_1,t)T_1/T_o\sigma_{z\theta}|_{T_o}$. Since Eq. (21) applies at all temperatures, it follows that

$$\begin{aligned}N|_{T_1}&=2\pi\int_{r_i}^{r_o}\sigma_{zz}|_{T_1}r\,dr=2\pi\int_{r_i}^{r_o}\beta(T_1,t)T_1/T_o\sigma_{zz}|_{T_1}r\,dr\\&=\beta(T_1,t)T_1/T_oN|_{T_o}\end{aligned} \tag{22a}$$

In a similar manner it can be shown that

$$M|_{T_1}=\beta(T_1,t)T_1/T_oM|_{T_o} \tag{22b}$$

It also follows that

$$\frac{N|_{T_1}}{M|_{T_1}}=\frac{N|_{T_o}}{M|_{T_o}} \tag{23}$$

is a constant, independent of time. Thus, the axial force and twisting moment have the same decrease with time. This is a result that could potentially be assessed experimentally by measuring the time dependence of the normal force and twisting moment at high temperatures to determine the validity of the proposed correspondence principle.

Lifetime Prediction for Elastomeric Seals. Consider the cylinder of the previous example once again at a uniform temperature $T_o<T_{cr}$. Let it now be force fit and thus seal the annular space between an inner rigid cylindrical of radius $r_i>R_i$ and outer rigid cylinder of radius $r_o<R_o$. Let the ends of the cylinder be traction free (see Fig. 2).

The resulting deformation is assumed to be axially symmetric. The end surfaces form the portion of the boundary $\partial\kappa^{(\sigma)}$ and the cylindrical surfaces form the portion of the boundary $\partial\kappa^{(d)}$. Suppose it is assumed that the deformation is given by Eq. (19) with $\psi=0$. It can then be shown that $M=0$. Since r_i and r_o are specified, λ and γ are determined from Eq. (20). In general, the boundary condition $\hat{\mathbf{T}}^{(\sigma)}=\mathbf{0}$ cannot be satisfied at each point on $\partial\kappa^{(\sigma)}$. Instead, let the relaxed boundary condition $\mathbf{N}=0$ be imposed. The scalar field $p_o(r)$ and the stress components are then completely determined. σ_{rr}, the radial stress, is the only nonzero stress component acting on $\partial\kappa^{(d)}$, and the pressure between the seal and a rigid cylinder is $|\sigma_{rr}|$. Suppose that the pressure between the seal and a cylindrical surface must be at least p^* if a leak is to be avoided. When $T_o<T_{cr}$, r_i and r_o can be chosen so that $|\sigma_{rr}(r_o)|_{T_o}>p^*$ and $|\sigma_{rr}(r_i)|_{T_o}>p^*$.

Suppose that the temperature of the seal is increased to $T_1>T_{cr}$ at $t=0$. According to the correspondence principle, the deformation is unchanged. Using the discussion from the first example, it can be shown that $N|_{T_1}=0$. The pressures for $t>0$ between the seal and the inner and outer rigid cylinders are, respectively,

$$\begin{aligned}p_{inner}(t)&=|\sigma_{rr}(r_i)|_{T_1}=\beta(T_1,t)T_1/T_o|\sigma_{rr}(r_i)|_{T_o},\\p_{outer}(t)&=|\sigma_{rr}(r_o)|_{T_1}=\beta(T_1,t)T_1/T_o|\sigma_{rr}(r_o)|_{T_o}.\end{aligned} \tag{24}$$

Because of scission, these pressures will relax with time. Leakage is predicted to occur at the smallest time t^* when

$$\min\{p_{inner}(t^*),p_{outer}(t^*)\}=p^* \tag{25}$$

is reached, thereby giving an estimate of the seal's useful life at temperature T_1.

5 Summary and Conclusions

A correspondence principle has been introduced which can be used to calculate the stress relaxation due to scission in an elastomeric component at an elevated temperature T_1 in terms of the stresses in the component at a lower temperature T_o where there is no scission. The application of the principle assumes knowledge of two items: 1) the stress distribution at temperature T_o, determined by either analytical or numerical methods and 2) a material property $\beta(T,t)$ that can be determined from uniaxial stress relaxation experiments at different constant temperatures. The correspondence principle requires that the deformations be the same at T_1 and T_o and that the temperature fields be homogeneous. When these conditions are at least approximately satisfied, the correspondence principle can give a useful first approximation to the actual stresses during scission. Two examples illustrate the application of the correspondence principle. In the first, a tension-torsion experiment can be used to assess the validity of the principle. In the second example, the usable lifetime of a seal at high temperature can be predicted.

Acknowledgment

This material is based on work supported by NSF Grant No. CTS 9908925.

References

[1] Tobolsky, A. V., 1960, *Properties and Structures of Polymers*, Wiley, New York, Chap. V, pp. 223–265.
[2] Wineman, A. S., Jones, A., and Shaw, J. A., 2002, "Life-Cycle and Durability Predictions of Elastomeric Components," *Modeling and Simulation-Based Life Cycle Engineering*, K. P., Chong, S., Saigal, S., Thynell, and H. S. Morgan, eds., Spon, New York, pp. 155–169.
[3] Wineman, A. S., and Rajagopal, K. R., 2000, *Mechanical Response of Polymers*, Cambridge University Press, New York.
[4] Tobolsky, A. V., Prettyman, I. B., and Dillon, J. H., 1944, "Stress Relaxation of Natural and Synthetic Rubber Stocks," J. Appl. Phys., **15**, pp. 380–395.
[5] Scanlan, J., and Watson, W. F., 1958, "The Interpretation of Stress-Relaxation Measurements Made on Rubber During Aging," Trans. Faraday Soc., **54**, pp. 740–750.
[6] Neubert, D., and Saunders, D. W., 1958, "Some Observations of the Permanent Set of Cross-linked Natural Rubber Samples After Heating in a State of Pure Shear," Rheol. Acta, **1**, pp. 151–157.

[7] Fong, J. T., and Zapas, L. J., 1976, ''Chemorheology and the Mechanical Behavior of Vulcanized Rubber,'' Trans. Soc. Rheol., **20**, pp. 319–338.
[8] Wineman, A. S., and Shaw, J. A., 2002, ''Scission and Healing in a Spinning Elastomeric Cylinder at Elevated Temperature,'' ASME J. Appl. Mech., **69**, pp. 602–609.
[9] Han, W. H., Horkay, F., and McKenna, G. B., 1999, ''Mechanical and Swelling Behaviors of Rubber: A Comparison of Some Molecular Models With Experiment,'' Math. Mech. Solids, **4**, pp. 139–167.
[10] Ogden, R. W., 1997, *Non-Linear Elastic Deformations*, Dover, Mineola, NY.

L. Z. Sun[1]
Department of Civil and Environmental Engineering,
The University of Iowa,
Iowa City, IA 52242-1527
e-mail: lizhi-sun@uiowa.edu
Mem. ASME

J. W. Ju
Department of Civil and Environmental Engineering,
University of California,
Los Angeles, CA 90095-1593
Fellow ASME

Elastoplastic Modeling of Metal Matrix Composites Containing Randomly Located and Oriented Spheroidal Particles

Micromechanics-based effective elastic and plastic formulations of metal matrix composites (MMCs) containing randomly located and randomly oriented particles are developed. The averaging process over all orientations upon three elastic governing equations for aligned particle-reinforced MMCs is performed to obtain the explicit formulation of effective elastic stiffness of MMCs with randomly oriented particles. The effects of volume fraction of particles and particle shape on the overall elastic constants are studied. Comparisons with the Hashin-Shtrikman bounds and Ponte Castaneda-Willis bounds show that the present effective elastic formulation does not violate the variational bounds. Good agreement with experimental elastic stiffness data is also illustrated. Furthermore, the orientational averaging procedure is employed to derive the overall elastoplastic yield function for the MMCs. Elastoplastic constitutive relations for the composites are constructed on the basis of the derived composite yield function. The stress-strain responses of MMCs under the axisymmetric loading are also investigated in detail. Finally, elastoplastic comparisons with the experimental data for SiCp/Al composites are performed to illustrate the capability of the proposed formulation. [DOI: 10.1115/1.1794699]

1 Introduction

Metal matrix composites (MMCs) have evolved over the past 40 years, and have been applied as structural materials in aerospace and automotive industries, due to their high-performance specific mechanical properties in service. The stress-strain response of MMCs generally represents their basic mechanical behavior; i.e., the elastic stiffness, yield stress, and plastic flow [1]. The prediction of effective (overall) elastic and plastic behavior of MMCs are of considerable interest to researchers and engineers in many science and engineering disciplines.

Although the constitutive relations of discontinuously reinforced MMCs were widely studied in recent years, most of the research efforts have been directed to unidirectionally aligned particle-reinforced composites. For MMCs containing *randomly oriented* particles, the first elastic study seems attributed to Cox [2] who considered the effects of random orientation of short fibers on the stiffness and strength of paper and other fibrous materials. However, Cox's results did not take into account interactions between fiber and matrix phases. Christensen and coworkers [3,4] proposed a geometric averaging method for determining the effective isotropic elastic properties of randomly oriented fiber composites. Christensen's averaging method is directly based on the stiffness-type mechanical properties of aligned continuous fiber composites so that the end effects of short fibers are neglected. Since the 1980s, more investigators were involved in predicting the effective elastic behavior of composites based on micromechanical approaches [5–20]. For example, Chou and Nomura [5] derived the elastic formulation of randomly oriented composites based on the results from both the mathematical bounds and the self-consistent method for each orientation. Takao et al. [6], Tandon and Weng [7], Benveniste [8], and Ferrari and Johnson [9] showed the effect of misoriented fiber on the overall elastic constants of composites, based on the Mori-Tanaka mean-field theory [21], coupled with different fiber distribution functions. Ponte Castaneda and Willis [15] developed the refined variational bounds for randomly oriented composites. Riccardi and Montheillet [18] proposed a generalized self-consistent method to predict the effective elastic moduli of composites.

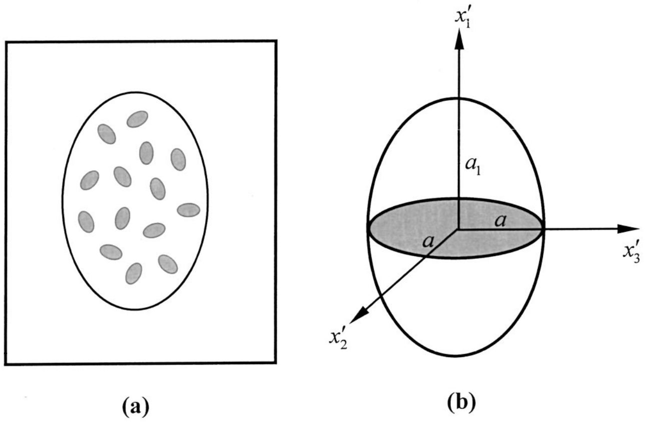
For the elastoplastic behavior of MMCs with randomly oriented particles, Qiu and Weng [22,23] considered the overall composite behavior, using a secant moduli approach coupling with either the Mori-Tanaka method [7,21,24] or the energy method proposed by the authors earlier [25]. Li et al. [26] derived the elastoplastic response of the composites, based on the variational procedure developed by Ponte Castaneda [27,28]. Dunn and Ledbetter [29] also proposed a micromechanical model to predict the elastoplastic behavior of MMCs with orientation-preferred (textured) reinforcements. From a numerical point of view, Bao et al. [30] calculated particle orientation effects on discontinuously reinforced MMCs by considering randomly oriented needle-type elastic reinforcements and disc-type elastic particles embedded in an elastoplastic matrix. Moreover, Sorensen et al. [31] and Dong et al. [32] employed micromechanical finite element methods to simulate the effects of reinforcement misalignment on the tensile elastoplastic response of MMCs.

The objective of the present paper is to develop a micromechanics-based effective elastoplastic model of MMCs containing randomly located and randomly oriented spheroidal particles. The averaging process over all orientations upon three governing equations for aligned spheroidal particle-reinforced composites is performed to obtain the constitutive relations and isotropic elastic stiffness of MMCs with randomly oriented spheroidal particles. The effects of the volume fraction of particles and the particle aspect ratio on the overall elastic constants of the composites are discussed. The comparisons with the Hashin-Shtrikman bounds [33], Ponte Castaneda-Willis bounds [15], and the available experimental data are also illustrated. Furthermore,

---

[1] Author 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, August 20, 2002; final revision; December 29, 2003. Associate Editor: A Needleman. 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.



(a) (b)

Fig. 1 (*a*) **A composite containing randomly dispersed and randomly oriented spheroidal particles, and** (*b*) **the sketch of a spheroid**

the orientational average is utilized to derive the overall yield function for the composites based on the previous results [34,35] for the aligned particle-reinforced MMCs. As a demonstration, the overall elastoplastic stress-strain curves under uniaxial loading are studied in detail. Comparisons between the present theoretical predictions and experimental data are performed to illustrate the capability of the proposed model. Finally, initial yield surfaces and effects of stress ratio on the MMCs under axisymmetric loading are investigated.

2 Effective Elasticity of MMCs With Randomly Oriented Particles

Let us start by considering a two-phase composite consisting of a linearly elastic isotropic matrix (phase 0, with the stiffness $\mathbf{C}^{(0)}$) and linearly elastic isotropic spheroidal particles (phase 1, with the stiffness $\mathbf{C}^{(1)}$), as shown in Fig. 1(*a*). The randomly oriented particles are also randomly located in the matrix. The aspect ratio of the spheroidal particles is defined as $\alpha = a_1/a$ where a_1 and a are the semi-axes of the spheroids [Fig. 1(*b*)]. It is assumed that the two phases are perfectly bonded at the interfaces.

The transformation between local coordinates of particles and global coordinates of composites is demonstrated first. As shown in Fig. 2, the local axes of a spheroidal inclusion are denoted by the primed coordinate system (x_1', x_2', x_3') with axis x_1' being the symmetric axis of a spheroid. The global or fixed axes are represented by the unprimed coordinate system (x_1, x_2, x_3). Each component of the coordinate transformation matrix $[Q_{ij}]$ is simply the cosine between the *i*th primed and *j*th unprimed axes. Thus, we have (c.f. [7])

$$\{x_i'\} = [Q_{ij}]\{x_j\} \tag{1}$$

where the transformation matrix has the following form

$$[Q_{ij}] = \begin{bmatrix} \cos\theta & \sin\theta\cos\gamma & \sin\theta\sin\gamma \\ -\sin\theta & \cos\theta\cos\gamma & \cos\theta\sin\gamma \\ 0 & -\sin\gamma & \cos\gamma \end{bmatrix} \tag{2}$$

with θ $(0 \leqslant \theta \leqslant \pi/2)$ denoting the angle between x_1 and x_1', and γ $(0 \leqslant \gamma \leqslant 2\pi)$ representing the angle between x_3 and x_3'. Therefore, any second-rank tensor, e.g., the stress tensor, can be transformed between the global and the local as

$$\sigma_{ij}' = Q_{ik}Q_{jl}\sigma_{kl} \tag{3}$$

or

$$\sigma_{ij} = Q_{ki}Q_{lj}\sigma_{kl}' \tag{4}$$

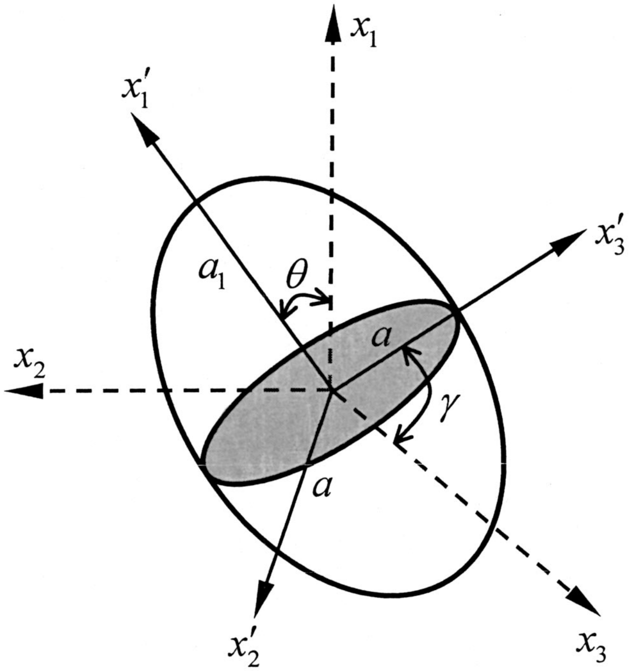


Fig. 2 Transformation between local coordinate system x_i' and global coordinate system x_j

2.1 Orientation-Averaged Constitutive Equations. To obtain effective constitutive equations of random heterogeneous composites, one typically performs the ensemble-volume and orientational averaging procedures (homogenizations) within a mesoscopic representative volume element (RVE). In particular, Ju and Chen [36] performed the ensemble-volume homogenization and derived the governing stress-strain equations for particle-reinforced composites under the applied far-field stress σ^0. In the case of the first-order approximation (where the near-field strong interparticle interactions are neglected), the three governing constitutive equations take the form [36]:

$$\begin{aligned} \overline{\sigma} &= \mathbf{C}^{(0)}:(\overline{\varepsilon} - \phi\overline{\varepsilon^*}) \\ \overline{\varepsilon} &= \varepsilon^0 + \phi\mathbf{S}:\overline{\varepsilon^*} \\ \overline{\varepsilon^*} &= -(\mathbf{S}+\mathbf{A})^{-1}:\varepsilon^0 \end{aligned} \tag{5}$$

where $\overline{\sigma}$, $\overline{\varepsilon}$, and $\overline{\varepsilon^*}$ are the ensemble-volume averaged stress, strain, and eigenstrain tensors of composites, respectively. The corresponding far-field strain is $\varepsilon^0 = \mathbf{C}^{(0)^{-1}}:\sigma^0$ and the elastic mismatch tensor reads $\mathbf{A} = [\mathbf{C}^{(1)} - \mathbf{C}^{(0)}]^{-1}\cdot\mathbf{C}^{(0)}$. The operation symbol ":" denotes the contraction between a fourth-rank tensor and a second-rank tensor while the symbol "·" represents the multiplication between two fourth-rank tensors. In addition, ϕ defines the volume fraction of particles, and $\mathbf{S}$ is the fourth-rank Eshelby's tensor of an ellipsoid. For spheroidal particles, the Eshelby's tensor $\mathbf{S}$ can be expressed as [37]

$$S_{ijkl} = S_{IK}^{(1)}\delta_{ij}\delta_{kl} + S_{IJ}^{(2)}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{6}$$

where the second-rank tensors $S_{IK}^{(1)}$ and $S_{IJ}^{(2)}$ are given in Appendix A. It is noted that Mura's [38] tensorial indicial notation is followed in the above expressions; i.e., upper-case indices take on the same numbers as the corresponding lower-case ones but are not summed up. The effective stiffness tensor of aligned particle-reinforced composites can be easily obtained from the foregoing three governing equations. The explicit expression of effective stiffness is of the transversely isotropic form as shown in Ref. [34].

When all of the reinforcements of composites are randomly located and randomly oriented in the three-dimensional (3D) space, the orientation-averaging process is further applied to Eq. (5). For example, the orientational average of $\overline{\sigma}$ is defined as

$$\langle\overline{\sigma}\rangle \triangleq \int_0^{2\pi}\int_0^{\pi/2}\overline{\sigma}P(\theta,\gamma)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\gamma \tag{7}$$

where the range of the two Eulerian angles θ and γ (see Fig. 2) are $0\leqslant\theta\leqslant\pi/2$ and $0\leqslant\gamma\leqslant 2\pi$, respectively. Further, the function $P(\theta,\gamma)$ represents the probability density function of the random orientation. In what follows, the random orientation of particles is assumed to be uniform so that $P(\theta,\gamma)=1/2\pi$ is a constant. The three governing equations of composites together with the uniformly random orientation of particles can be formulated as

$$\begin{aligned}\langle\overline{\sigma}\rangle&=\mathbf{C}^{(0)}:(\langle\overline{\varepsilon}\rangle-\phi\langle\overline{\varepsilon^*}\rangle)\\ \langle\overline{\varepsilon}\rangle&=\Lambda:\varepsilon^0\\ \langle\overline{\varepsilon^*}\rangle&=-\Omega:\varepsilon^0\end{aligned} \tag{8}$$

where the fourth-rank tensors Λ and Ω read

$$\Lambda_{ijkl}=I_{ijkl}-\frac{\phi}{2\pi}\int_0^{2\pi}\int_0^{\pi/2}Q_{mi}Q_{nj}S_{mnpq}(A_{pqst}+S_{pqst})^{-1}Q_{sk}Q_{tl}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\gamma \tag{9}$$

and

$$\Omega_{ijkl}=\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\pi/2}Q_{mi}Q_{nj}(A_{mnpq}+S_{mnpq})^{-1}Q_{pk}Q_{ql}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\gamma \tag{10}$$

It is noted that **I** signifies the fourth-rank identity tensor. Emanating from Eq. (8), the effective elastic stiffness tensor for randomly oriented particle-reinforced composites can be derived as

$$\langle\overline{\mathbf{C}}\rangle=\mathbf{C}^{(0)}\cdot(\mathbf{I}+\phi\Omega\cdot\Lambda^{-1}) \tag{11}$$

2.2 Explicit Effective Stiffness. Before we explicitly derive the isotropic elastic constants for composites with randomly oriented spheroidal particles, we need to consider the following formula. For any transversely isotropic fourth-rank tensor, say L, which is expressed as

$$L_{ijkl}=L_{IK}^{(1)}\delta_{ij}\delta_{kl}+L_{IJ}^{(2)}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) \tag{12}$$

and satisfying $L_{12}^{(1)}=L_{13}^{(1)}$, $L_{21}^{(1)}=L_{31}^{(1)}$, $L_{22}^{(1)}=L_{23}^{(1)}=L_{32}^{(1)}=L_{33}^{(1)}$, $L_{12}^{(2)}=L_{21}^{(2)}=L_{13}^{(2)}=L_{31}^{(2)}$, and $L_{22}^{(2)}=L_{23}^{(2)}=L_{32}^{(2)}=L_{33}^{(2)}$, the following orientational average formulation can be obtained after lengthy but straightforward manipulations:

$$\begin{aligned}\langle L_{ijkl}\rangle&=\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\pi/2}Q_{mi}Q_{nj}L_{mnpq}Q_{pk}Q_{ql}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\gamma\\ &=\xi\delta_{ij}\delta_{kl}+\eta(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk})\end{aligned} \tag{13}$$

where

$$\begin{aligned}\xi&=\tfrac{1}{15}[L_{11}^{(1)}+4L_{12}^{(1)}+4L_{21}^{(1)}+6L_{22}^{(1)}+2L_{11}^{(2)}-4L_{12}^{(2)}+2L_{22}^{(2)}]\\ \eta&=\tfrac{1}{15}[L_{11}^{(1)}-L_{12}^{(1)}-L_{21}^{(1)}+L_{22}^{(1)}+2L_{11}^{(2)}+6L_{12}^{(2)}+7L_{22}^{(2)}]\end{aligned} \tag{14}$$

The above formulation shows that any transversely isotropic fourth-rank tensor, after the 3D orientational average, becomes an isotropic fourth-rank tensor, as expected.

Based on Eq. (13), the transforming tensor Λ between the macroscopic strain $\langle\overline{\varepsilon}\rangle$ and the far-field strain ε^0 can be constructed as

$$\Lambda_{ijkl}=-\phi\Lambda_1\delta_{ij}\delta_{kl}+(1-2\phi\Lambda_2)(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) \tag{15}$$

where

$$\begin{aligned}\Lambda_1=&\frac{[S_{11}^{(1)}+4S_{21}^{(1)}+2S_{11}^{(2)}][1-\Gamma_{11}-4\Gamma_{12}]+10S_{21}^{(1)}\Gamma_{12}}{30[Z_2+S_{11}^{(2)}]}\\ &-\frac{2S_{12}^{(2)}}{15[Z_2+S_{12}^{(2)}]}\\ &+\frac{[3S_{22}^{(1)}+2S_{12}^{(1)}+3S_{22}^{(2)}][3-4\Gamma_{21}-6\Gamma_{22}]-6S_{22}^{(2)}+5S_{12}^{(1)}\Gamma_{21}}{45[Z_2+S_{22}^{(2)}]}\end{aligned} \tag{16}$$

and

$$\begin{aligned}\Lambda_2=&\frac{[S_{11}^{(1)}-S_{21}^{(1)}+2S_{11}^{(2)}][1-\Gamma_{11}+\Gamma_{12}]}{30[Z_2+S_{11}^{(2)}]}+\frac{S_{12}^{(2)}}{5[Z_2+S_{12}^{(2)}]}\\ &+\frac{[S_{22}^{(1)}-S_{12}^{(1)}+S_{22}^{(2)}][1+2\Gamma_{21}-2\Gamma_{22}]+6S_{22}^{(2)}}{30[Z_2+S_{22}^{(2)}]}\end{aligned} \tag{17}$$

with

$$\Gamma_{I1}=\frac{[Z_1+Z_2+S_{22}^{(1)}+S_{22}^{(2)}][Z_1+S_{I1}^{(1)}]-[Z_1+S_{21}^{(1)}][Z_1+S_{I2}^{(1)}]}{[Z_1+Z_2+S_{22}^{(1)}+S_{22}^{(2)}][Z_1+2Z_2+S_{11}^{(1)}+2S_{11}^{(2)}]-[Z_1+S_{12}^{(1)}][Z_1+S_{21}^{(1)}]}$$

$$\Gamma_{I2}=\Gamma_{I3}=\frac{[Z_1+2Z_2+S_{11}^{(1)}+2S_{11}^{(2)}][Z_1+S_{I2}^{(1)}]-[Z_1+S_{12}^{(1)}][Z_1+S_{I1}^{(1)}]}{2[Z_1+Z_2+S_{22}^{(1)}+S_{22}^{(2)}][Z_1+2Z_2+S_{11}^{(1)}+2S_{11}^{(2)}]-2[Z_1+S_{12}^{(1)}][Z_1+S_{21}^{(1)}]}$$

$$\begin{aligned}Z_1&=\frac{\lambda_0\mu_1-\lambda_1\mu_0}{(\mu_1-\mu_0)[2(\mu_1-\mu_0)+3(\lambda_1-\lambda_0)]}\\ Z_2&=\frac{\mu_0}{2(\mu_1-\mu_0)}\end{aligned} \tag{18}$$

It is noted that λ_β and μ_β are the Lame constants of the β-phase (β=0,1). Furthermore, an inverse formula for the transversely isotropic fourth-rank tensor is utilized; see, Appendix B.

Similarly, the transforming tensor Ω between the orientational eigenstrain $\langle\overline{\varepsilon^*}\rangle$ and the far-field strain ε^0 can be constructed as

$$\Omega_{ijkl}=\Omega_1\delta_{ij}\delta_{kl}+\Omega_2(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) \tag{19}$$

in which

$$\Omega_1=\frac{1-\Gamma_{11}-4\Gamma_{12}}{30[Z_2+S_{11}^{(2)}]}-\frac{1}{15[Z_2+S_{12}^{(2)}]}+\frac{1-4\Gamma_{21}-6\Gamma_{22}}{30[Z_2+S_{22}^{(2)}]} \tag{20}$$

and

$$\Omega_2=\frac{1-\Gamma_{11}+\Gamma_{12}}{30[Z_2+S_{11}^{(2)}]}-\frac{1}{10[Z_2+S_{12}^{(2)}]}+\frac{7+\Gamma_{21}-\Gamma_{22}}{60[Z_2+S_{22}^{(2)}]} \tag{21}$$

Therefore, from Eq. (11), the overall bulk modulus κ^{comp} and shear modulus μ^{comp} of the randomly oriented composites can be derived as

$$\begin{aligned}\kappa^{\text{comp}}&=\frac{\kappa_0}{1-2\phi\Lambda_2}\bigg[1-2\phi\Lambda_2+2\phi\Omega_2\\&\quad+\frac{3\phi(\Omega_1+2\phi\Lambda_1\Omega_2-2\phi\Lambda_2\Omega_1)}{1-3\phi\Lambda_1-2\phi\Lambda_2}\bigg]\\\mu^{\text{comp}}&=\frac{\mu_0(1-2\phi\Lambda_2+2\phi\Omega_2)}{1-2\phi\Lambda_2}\end{aligned}\tag{22}$$

where κ_0 and μ_0 are the bulk modulus and shear modulus of the matrix, respectively.

Alternatively, the overall Young's modulus E^{comp} and Poisson ratio ν^{comp} of the randomly oriented composites can be obtained as

$$\begin{aligned}E^{\text{comp}}&=\frac{9\kappa^{\text{comp}}\mu^{\text{comp}}}{3\kappa^{\text{comp}}+\mu^{\text{comp}}}\\\nu^{\text{comp}}&=\frac{3\kappa^{\text{comp}}-2\mu^{\text{comp}}}{6\kappa^{\text{comp}}+2\mu^{\text{comp}}}\end{aligned}\tag{23}$$

To illustrate the capability of the proposed effective elasticity model, let us compare our analytical results [Eq. (22)] with the variational bounds for the isotropic elastic stiffness of composites. The current stiffness bounds for randomly oriented isotropic composites are available from Hashin and Shtrikman [33] and Ponte Castaneda and Willis [15]. The celebrated Hashin-Shtrikman (H-S) upper and lower bounds for the bulk modulus and shear modulus of isotropic composites were developed by Hashin and Shtrikman [33], and generalized by Hill [39] and Walpole [40,41]. By utilizing the theorems of minimum energy and minimum complementary energy, the H-S upper and lower bounds for the effective bulk modulus κ^{comp} and shear modulus μ^{comp} of a two-phase composite are described as [39–42]

$$\begin{aligned}\frac{\phi}{1+\dfrac{(1-\phi)(\kappa_1-\kappa_0)}{(\kappa_0+\kappa_l)}}&\leqslant\frac{\kappa^{\text{comp}}-\kappa_0}{\kappa_1-\kappa_0}\leqslant\frac{\phi}{1+\dfrac{(1-\phi)(\kappa_1-\kappa_0)}{(\kappa_0+\kappa_u)}}\\\frac{\phi}{1+\dfrac{(1-\phi)(\mu_1-\mu_0)}{(\mu_0+\mu_l)}}&\leqslant\frac{\mu^{\text{comp}}-\mu_0}{\mu_1-\mu_0}\leqslant\frac{\phi}{1+\dfrac{(1-\phi)(\mu_1-\mu_0)}{(\mu_0+\mu_u)}}\end{aligned}\tag{24}$$

where for $(\mu_1-\mu_0)(\kappa_1-\kappa_0)\geqslant 0$,

$$\begin{aligned}\kappa_l&=\tfrac{4}{3}\mu_0\\\kappa_u&=\tfrac{4}{3}\mu_1\\\mu_l&=\frac{3}{2}\left[\frac{1}{\mu_0}+\frac{10}{9\kappa_0+8\mu_0}\right]^{-1}\\\mu_u&=\frac{3}{2}\left[\frac{1}{\mu_1}+\frac{10}{9\kappa_1+8\mu_1}\right]^{-1}\end{aligned}\tag{25}$$

and for $(\mu_1-\mu_0)(\kappa_1-\kappa_0)\leqslant 0$,

$$\begin{aligned}\kappa_l&=\tfrac{4}{3}\mu_1\\\kappa_u&=\tfrac{4}{3}\mu_0\\\mu_l&=\frac{3}{2}\left[\frac{1}{\mu_0}+\frac{10}{9\kappa_1+8\mu_0}\right]^{-1}\\\mu_u&=\frac{3}{2}\left[\frac{1}{\mu_1}+\frac{10}{9\kappa_0+8\mu_1}\right]^{-1}\end{aligned}\tag{26}$$

Figure 3 shows the comparisons between our analytical predictions [Fig. 3(*a*) for the bulk modulus and Fig. 3(*b*) for the shear modulus] and the H-S bounds for the SiC hard particle (Young's modulus $E_p=450$ GPa and Poisson ratio $\nu_p=0.2$) reinforced Al (Young's modulus $E_m=70$ GPa and Poisson ratio $\nu_m=\nu_0=0.3$) metal matrix composites. Similar comparisons are also made for the soft particle ($E_p=20$ GPa and $\nu_p=0.25$) filled Al composites, as illustrated in Figs. 4(*a*) and (*b*). These comparisons demonstrate that the present analytical results always lie within the H-S bounds. In addition, for the spherical particle (with the aspect ratio $\alpha=1$) reinforced composite, our elastic prediction is identical to the H-S *lower* bound if the reinforcement is harder than the matrix, and coincides with H-S *upper* bound if the reinforcement is softer. Moreover, it is observed from Figs. 3(*a*) and (*b*) that the composites with *oblate* aspect ratios of particles are more effective in enhancing κ^{comp} and μ^{comp}, whereas the spherical particles are least effective. These results are consistent with those of Wu [43] and Tandon and Weng [7]. By contrast, if the matrix is the harder phase, the trend of the overall elastic moduli of composites is completely reversed, as shown in Figs. 4(*a*) and (*b*). It is also noted that there is a continuous transition of elastic properties between the upper and lower bounds with relative changes in constituent moduli. Furthermore, care must be exercised to prevent the spheroidal reinforcements from overlapping with each other. Such cases would occur if the aspect ratios of the spheroids were extremely high or extremely low.

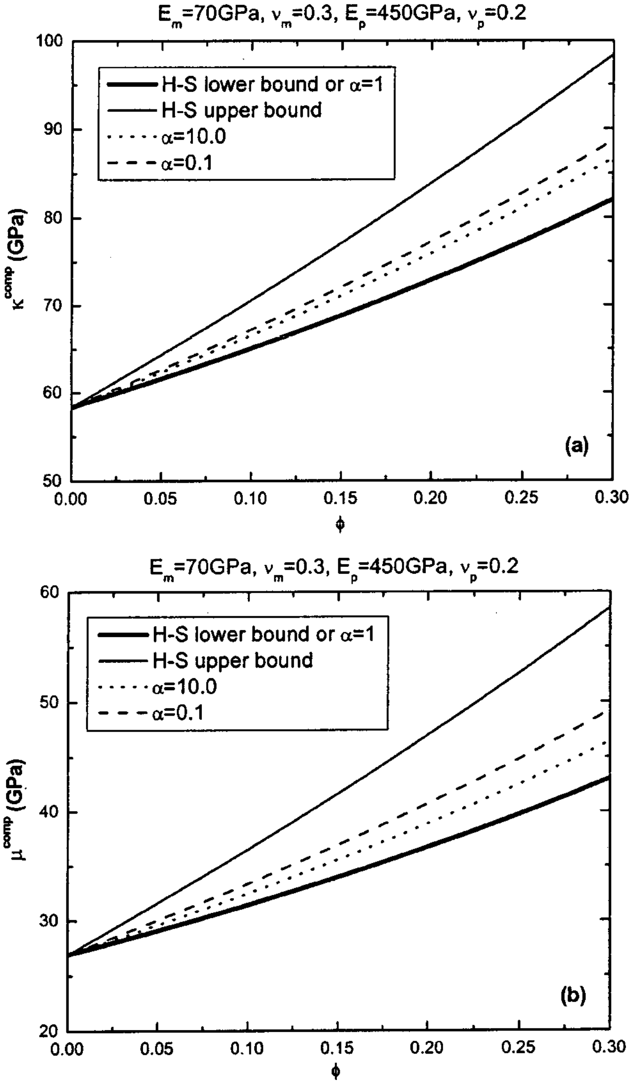


Fig. 3 The overall (*a*) bulk modulus and (*b*) shear modulus of composites with randomly located and oriented harder particles as a function of the volume fraction and aspect ratio of particles

From Eq. (24), it is clear that the H-S bounds are related to the

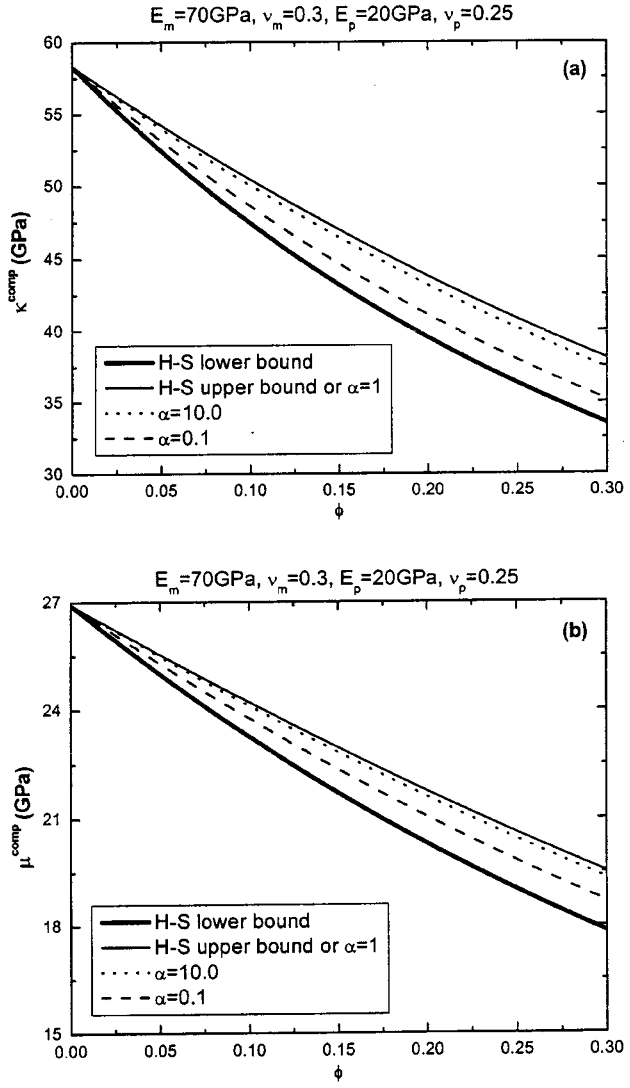


Fig. 4 The overall (*a*) bulk modulus and (*b*) shear modulus of composites with randomly located and oriented softer particles as a function of the volume fraction and aspect ratio of particles

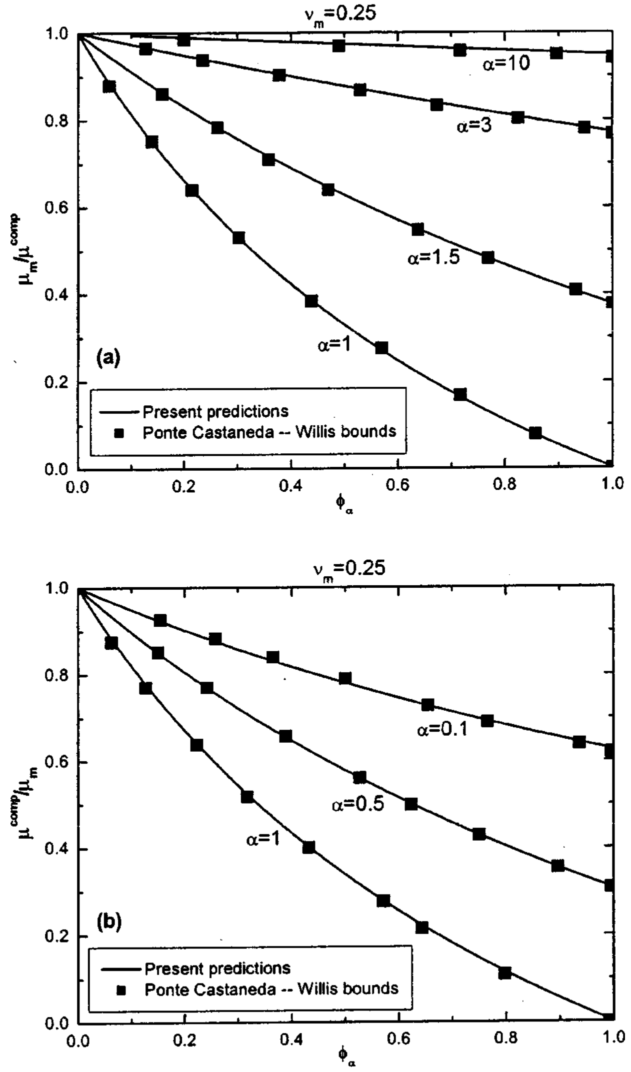


Fig. 5 The comparisons between the present predictions and the Ponte Castaneda-Willis bounds [15] for (*a*) effective shear modulus of composites with randomly oriented spheroidal rigid particles and (*b*) effective shear modulus of porous materials with randomly oriented spheroidal voids

volume fraction of reinforcements and elastic constants of the constituent phases. Nevertheless, the H-S bounds are independent of the shapes, spatial locations, or spatial orientations of particles. Based on Hashin and Shtrikman [33], Walpole [40,41], Willis [44], and Ponte Castaneda and Willis [15] further refined the variational methodology of the overall stiffness bounds for randomly oriented isotropic composites, which directly depend on the aspect ratio of misoriented spheroidal particles. Two cases were explicitly provided (as shown in Fig. 8 in [15]): effective shear modulus lower bound for isotropic composites with randomly oriented rigid spheroids, and effective shear modulus bound of isotropic materials with randomly oriented spheroidal voids. These two cases can be directly compared with the present predictions resulting from Eq. (22). It is shown from Figs. 5(*a*) and (*b*) that predictions from our proposed model are mostly identical with the bounds for various aspect ratios of spheroids. The effective shear moduli μ^{comp} are plotted as a function of spheroid density parameter ϕ_α introduced by Ref. [15]. The density parameter ϕ_α is equal to $\phi\alpha^2$ for prolate spheroids and is equal to ϕ/α for oblate spheroids.

We further compare our analytical predictions with the experimental data reported by Yang et al. [45], as displayed in Fig. 6. The composites are made of SiC particulates dispersed in the Al matrix. The input data are taken as $E_m = 75\,\mathrm{GPa}$, $\nu_m = 0.33$, $E_p = 420\,\mathrm{GPa}$, and $\nu_p = 0.17$, in accordance with the constituents. The mean aspect ratio of 2.0 for the SiC particulates is converted to 3.0 based on the volume equivalence between particulates and spheroidal particles [46]. The present elastic formulation performs very well for the volume fraction of particles up to about 25%. When the volume fraction of particles becomes larger, the present predictions somewhat underestimate the effective Young's moduli as compared with the experimental data. This is expected since the present formulation neglects the near-field direct interactions among the spheroidal particles in the composites.

3 Effective Elastoplasticity of MMCs With Randomly Oriented Particles

3.1 Orientation-Averaged Constitutive Equations. Let us consider two-phase MMCs containing randomly located and oriented elastic spheroidal particles embedded in the elastoplastic matrix. For simplicity, the commonly used von Mises J_2-yield

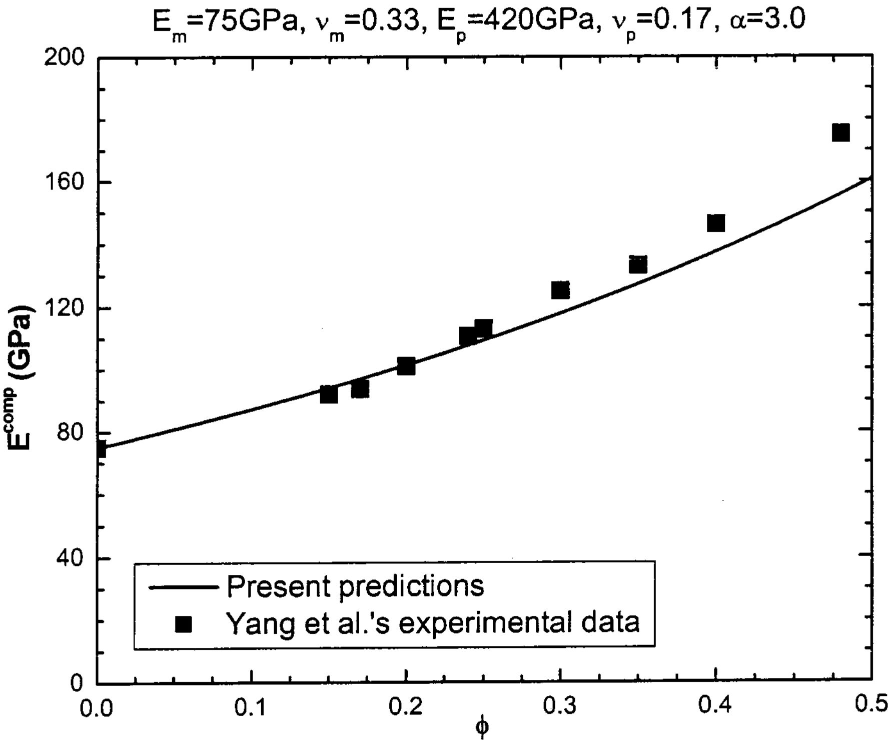


Fig. 6 Comparison between the present predictions and the experimental Young's moduli of SiCp/Al composites (Yang et al., [44])

criterion with an isotropic hardening law is assumed for the matrix material; i.e., at any matrix material point **x**, the local stress $\sigma(x)$ must satisfy the following yield criterion:

$$F(\sigma,e_m^p)=\sqrt{H(\sigma)}-\sqrt{\tfrac{2}{3}}[\sigma_y+h(e_m^p)^q]\leqslant 0 \qquad (27)$$

where σ_y, h, q, and e_m^p signify the initial yield strength, the linear hardening parameter, the exponential hardening parameter, and the effective equivalent plastic strain of the matrix material, respectively. Furthermore, $H(\sigma)$ represents the square of the stress norm at matrix point x, defined by $H=\sigma:\mathbf{I}_d:\sigma$, with $\mathbf{I}_d$ being the deviatoric part of the fourth-rank identity tensor.

The overall yield function of MMCs with randomly located and oriented spheroids can be derived by employing the aforementioned orientational averaging process upon the analytical formulation for the randomly located, yet unidirectionally aligned, spheroidal particle-reinforced MMCs, which was previously developed by Ju and Sun [34]. Following Ju and Sun [34], the ensemble-volume-averaged stress norm $\bar{H}$ reads

$$\bar{H}=\sigma^0:\mathbf{T}:\sigma^0 \qquad (28)$$

where the transversely isotropic fourth-rank tensor **T** takes the form:

$$T_{ijkl}=T_{IK}^{(1)}\delta_{ij}\delta_{kl}+T_{IJ}^{(2)}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) \qquad (29)$$

with the second-rank tensors $T_{IK}^{(1)}$ and $T_{IJ}^{(2)}$ given in Appendix A.

Since the far-field stress σ^0 is independent of the local orientations of particles, the orientational average of $\bar{H}$ can be simply obtained as

$$\langle\bar{H}\rangle=\sigma^0:\langle\bar{\mathbf{T}}\rangle:\sigma^0 \qquad (30)$$

where the fourth-rank tensor $\langle\bar{\mathbf{T}}\rangle$ reads

$$\langle T_{ijkl}\rangle=T_1\delta_{ij}\delta_{kl}+T_2(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) \qquad (31)$$

Note that $\langle\bar{\mathbf{T}}\rangle$ is an isotropic fourth-rank tensor, and its two components are calculated as

$$\begin{aligned}T_1&=\tfrac{1}{15}[T_{11}^{(1)}+4T_{12}^{(1)}+4T_{21}^{(1)}+6T_{22}^{(1)}+2T_{11}^{(2)}-4T_{12}^{(2)}+2T_{22}^{(2)}]\\ T_2&=\tfrac{1}{15}[T_{11}^{(1)}-T_{12}^{(1)}-T_{21}^{(1)}+T_{22}^{(1)}+2T_{11}^{(2)}+6T_{12}^{(2)}+7T_{22}^{(2)}]\end{aligned} \qquad (32)$$

In Eq. (30), $\langle\bar{H}\rangle$ is expressed in terms of the far-field stress σ^0. Alternatively, the orientation-averaged square of the stress norm can also be formulated via the macroscopic (orientation-ensemble-volume-averaged) stress $\langle\bar{\sigma}\rangle$. The relationship between the far-field stress σ^0 and the orientation-averaged stress $\langle\bar{\sigma}\rangle$ can be established through Eq. (8) as follows

$$\sigma^0=\langle\mathbf{P}\rangle:\langle\bar{\sigma}\rangle \qquad (33)$$

where the isotropic fourth-rank tensor $\langle\mathbf{P}\rangle$ reads

$$\langle P_{ijkl}\rangle=P_1\delta_{ij}\delta_{kl}+P_2(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) \qquad (34)$$

with

$$\begin{aligned}P_1&=\frac{\phi(\Lambda_1-\Omega_1)}{[1+2\phi(\Omega_2-\Lambda_2)][1+\phi(3\Omega_1+2\Omega_2-3\Lambda_1-2\Lambda_2)]}\\ P_2&=\frac{1}{2+4\phi(\Omega_2-\Lambda_2)}\end{aligned} \qquad (35)$$

As a result, Eq. (30) can be rephrased as

$$\langle\bar{H}\rangle=\langle\bar{\sigma}\rangle:\langle\bar{\mathbf{T}}\rangle:\langle\bar{\sigma}\rangle \qquad (36)$$

where

$$\langle\bar{T}_{ijkl}\rangle=\bar{T}_1\delta_{ij}\delta_{kl}+\bar{T}_2(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) \qquad (37)$$

with

$$\overline{T}_1=(3P_1+2P_2)^2T_1+2P_1(3P_1+4P_2)T_2$$
$$\overline{T}_2=4P_2^2T_2 \tag{38}$$

Therefore, the effective (orientation-ensemble-volume-averaged) yield function of randomly oriented particle-reinforced MMCs can be proposed as

$$\langle\overline{F}\rangle=(1-\phi)\sqrt{\langle\overline{\sigma}\rangle:\langle\overline{\mathbf{T}}\rangle:\langle\overline{\sigma}\rangle}-\sqrt{\frac{2}{3}}[\sigma_y+h(\langle\overline{e^p}\rangle)^q]\leqslant 0 \tag{39}$$

where $\langle\overline{e^p}\rangle$ denotes effective equivalent plastic strain of the composites. It is noted that the effective yield function is pressure dependent and not of the von Mises type anymore.

The total macroscopic strain $\langle\overline{\varepsilon}\rangle$ is composed of an elastic part and plastic part:

$$\langle\overline{\varepsilon}\rangle=\langle\overline{\varepsilon^e}\rangle+\langle\overline{\varepsilon^p}\rangle \tag{40}$$

where $\langle\overline{\varepsilon^e}\rangle$ and $\langle\overline{\varepsilon^p}\rangle$ are the macroscopic (orientation-ensemble-volume-averaged) elastic and plastic strains of the composites, respectively. The elastic constitutive relation reads

$$\langle\overline{\sigma}\rangle=\langle\overline{\mathbf{C}}\rangle:\langle\overline{\varepsilon^e}\rangle \tag{41}$$

where the effective elastic stiffness tensor $\langle\overline{\mathbf{C}}\rangle$ has previously been given in Eq. (11). The effective plastic constitutive law is assumed to satisfy the associated flow rule:

$$\langle\dot{\overline{\varepsilon}}{}^p\rangle=\dot{\lambda}\frac{\partial\langle\overline{F}\rangle}{\partial\langle\overline{\sigma}\rangle}=(1-\phi)\dot{\lambda}\frac{\langle\overline{\mathbf{T}}\rangle:\langle\overline{\sigma}\rangle}{\sqrt{\langle\overline{\sigma}\rangle:\langle\overline{\mathbf{T}}\rangle:\langle\overline{\sigma}\rangle}} \tag{42}$$

where $\dot{\lambda}$ is the plastic consistency parameter to be determined from the plastic consistency condition. It is noted that, when deriving Eq. (42), only the first term of the load function $\langle\overline{F}\rangle$ in Eq. (39) should be differentiated with respect to $\langle\overline{\sigma}\rangle$.

Consequently, the overall elastoplastic constitutive model for randomly oriented particle-reinforced MMCs is established as above. In the following subsection, a specific example is provided to illustrate the capability of the proposed effective elastoplastic formulation.

3.2 Axisymmetric Loading. For axisymmetric loading, the symmetrical axis is assumed parallel to the x_1 axis of the global coordinates. The overall stresses $\langle\overline{\sigma}\rangle$ satisfy

$$\langle\overline{\sigma}_{11}\rangle>0,\quad \langle\overline{\sigma}_{22}\rangle=\langle\overline{\sigma}_{33}\rangle=R\langle\overline{\sigma}_{11}\rangle,\quad \langle\overline{\sigma}_{12}\rangle=\langle\overline{\sigma}_{23}\rangle=\langle\overline{\sigma}_{31}\rangle=0 \tag{43}$$

where the stress ratio R is a function of the loading history. For simplicity, only a constant R is considered here. If $R=0$, the uniaxial tension loading is recovered. On the other hand, $R=1$ recovers the pure hydrostatic loading. Under the axisymmetric loading, the effective elastic strains can be expressed as

$$\langle\overline{\varepsilon}{}^e_{ij}\rangle=\frac{\langle\overline{\sigma}_{11}\rangle}{E^{\text{comp}}}\begin{bmatrix} D_1 & 0 & 0\\ 0 & D_2 & 0\\ 0 & 0 & D_2\end{bmatrix} \tag{44}$$

where

$$D_1=1-2\nu^{\text{comp}}R$$
$$D_2=-\nu^{\text{comp}}+(1-\nu^{\text{comp}})R \tag{45}$$

On the basis of Eq. (39), the effective yield function for the axisymmetric loading becomes

$$\langle\overline{F}\rangle=(1-\phi)\Phi(R)\langle\overline{\sigma}_{11}\rangle-\sqrt{\tfrac{2}{3}}[\sigma_y+h(\langle\overline{e^p}\rangle)^q]\leqslant 0 \tag{46}$$

where

$$\Phi(R)=\sqrt{\overline{T}_1+2\overline{T}_2+4R\overline{T}_1+4R^2(\overline{T}_1+\overline{T}_2)} \tag{47}$$

The initial yield surface of the composites can be secured from Eq. (46) with $h=0$ and $q=0$. By introducing the volumetric stress $\langle\overline{\sigma}_v\rangle\triangleq\langle\overline{\sigma}_{kk}\rangle/3=(\langle\overline{\sigma}_{11}\rangle+2\langle\overline{\sigma}_{22}\rangle)/3$ and the effective stress $\langle\overline{\sigma}_e\rangle\triangleq\sqrt{3\langle s_{ij}\rangle\langle s_{ij}\rangle/2}=\langle\overline{\sigma}_{11}\rangle-\langle\overline{\sigma}_{22}\rangle$ (where $\langle\overline{\mathbf{s}}\rangle$ is the deviatoric part of the macroscopic stress $\langle\overline{\sigma}\rangle$), the initial yield surface of the composites can be written as

$$9(3\overline{T}_1+2\overline{T}_2)\left(\frac{\overline{\sigma}_v}{\sigma_y}\right)^2+4\overline{T}_2\left(\frac{\overline{\sigma}_e}{\sigma_y}\right)^2=\frac{2}{(1-\phi)^2} \tag{48}$$

The overall elastoplastic stress-strain relationship is thus expressed as

$$\langle\overline{\varepsilon}_{ij}\rangle=\frac{\langle\overline{\sigma}_{11}\rangle}{E^{\text{comp}}}\begin{bmatrix} D_1 & 0 & 0\\ 0 & D_2 & 0\\ 0 & 0 & D_2\end{bmatrix}+\frac{(1-\phi)\lambda}{\Phi(R)}\begin{bmatrix} L_1 & 0 & 0\\ 0 & L_2 & 0\\ 0 & 0 & L_2\end{bmatrix} \tag{49}$$

where the first term on the right-hand side of the above equation represents the elastic part and the second term is attributed to the plastic effect. In addition, the matrix components L_1 and L_2 are

$$L_1=\overline{T}_1+2R\overline{T}_1+2\overline{T}_2$$
$$L_2=\overline{T}_1+2R\overline{T}_1+2R\overline{T}_2 \tag{50}$$

The cumulative plastic consistency parameter λ is determined as

$$\lambda=\frac{1}{\sqrt{2/3}(1-\phi)}\left[\frac{\sqrt{3/2}(1-\phi)\Phi(R)\langle\overline{\sigma}_{11}\rangle-\sigma_y}{h}\right]^{1/q} \tag{51}$$

In all subsequent numerical simulations in this subsection, unless otherwise noted, the matrix is assumed to be an aluminum alloy and the Young's modulus E_m, the Poisson ratio ν_m, the initial yield stress σ_y, and the linear and exponential isotropic hardening parameters h and q are taken as 70 GPa, 0.3, 300 MPa, 1000 MPa, and 0.5, respectively. The Young's modulus E_p and Poisson ratio ν_p of reinforcements are 450 GPa and 0.2, respectively, similar to those of SiC reinforcements.

The mechanical behavior of MMCs in practice often involves the monotonic uniaxial stress-strain relationship. To demonstrate the present effective elastoplastic micromechanics-based model for randomly oriented particle-reinforced MMCs, it is of interest to consider the special case of uniaxial tensile stress loading with $R=0$. Figure 7 provides the illustration of uniaxial overall stress-strain responses of the composites. From Fig. 7(*a*), it is clearly observed that, with increasing reinforcement concentration, the effective yield strength and plastic hardening modulus increases. However, the strengthening effects are not as significant as those of the composites with unidirectionally aligned particles; see, Sun and Ju [35]. In addition, Fig. 7(*b*) exhibits the effects of particle shape on the stress-strain behavior of MMCs. Unlike the aligned particle-reinforced MMCs, the aspect ratio of particles does not result in a significant difference in the elastoplastic behavior of the randomly oriented MMCs. However, the composites with randomly oriented *oblate* particles correspond to stiffer (higher) responses than those with prolate particles. Figure 7(*c*) also indicates that the composites with higher Young's modulus of reinforcements result in higher (stiffer) stress-strain responses.

As a step further, the present effective elastoplastic uniaxial predictions of MMCs with randomly oriented particles are compared with the experimental data reported by Yang et al. [45]; see Fig. 8. Here, we adopt the elastoplastic parameters $E_m=75\,\text{GPa}$, $\nu_m=0.33$, $\sigma_y=50\,\text{MPa}$, $h=320\,\text{MPa}$, and $q=0.265$, in accordance with the aluminum matrix material. Moreover, elastic constants for the SiC particles are taken as $E_p=420\,\text{GPa}$, and $\nu_p=0.17$. Again, the mean aspect ratio of 2.0 for randomly oriented particulates is converted to 3.0 as previously discussed. As depicted in Fig. 8, the overall elastoplastic behavior of the composite is well modeled by the present formulation, for particle volume fractions up to 48%. Since the effect of aspect ratio of particles

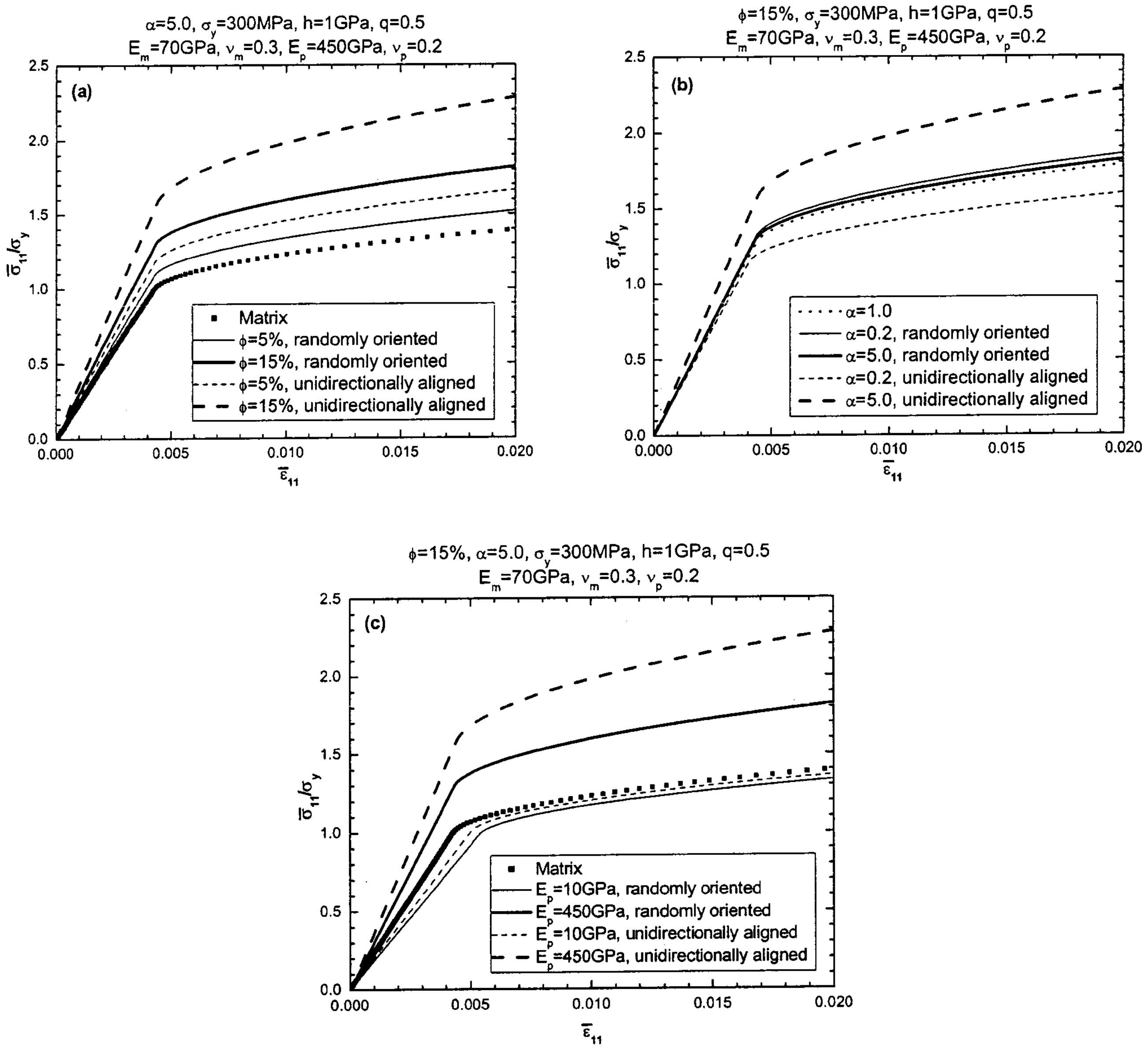


Fig. 7 Effects of (*a*) volume fraction, (*b*) aspect ratio, and (*c*) Young's modulus of particles on the uniaxial elastoplastic behavior of randomly oriented particle-reinforced MMCs

seems not to be significant for uniaxially elastoplastic responses of randomly oriented particle-reinforced MMCs (see Fig. 7(*b*)), the predictions for spherical particle-reinforced MMCs ($\alpha=1$) are also plotted in Fig. 8 for comparison purposes. It is observed that the simulation from $\alpha=1$ underestimates the elastoplastic responses but its difference is not significant compared with the prediction from $\alpha=3$. It is further noted that, rigorously speaking, since the present micromechanics-based model does not consider the near-field direct interactions among particles, our effective elastoplastic formulation should only be valid for moderate concentrations of particles. Based on these preliminary validations, it seems that the proposed framework along with the assumption of associated plastic flow rule is adequate and satisfactory.

The overall initial yield surfaces of the composites under axisymmetric loading are displayed in Figs. 9(*a*) and (*b*). It is observed that all yield surfaces are pressure (mean stress $\langle \overline{\sigma_v} \rangle$) dependent. Furthermore, Fig. 9(*a*) illustrates that, for the composites with prolate spheroidal particles, the initial yielding point for the effective stress $\langle \overline{\sigma_e} \rangle$ increases with increasing volume fraction of particles, whereas the yielding point for volumetric stress $\langle \overline{\sigma_v} \rangle$ decreases. On the other hand, Fig. 9(*b*) demonstrates that the composites with smaller aspect ratios exhibit more strengthening effect in terms of the volumetric stress $\langle \overline{\sigma_v} \rangle$. However, the particle shape does not have a significant influence on $\langle \overline{\sigma_e} \rangle$. In general, it is difficult to find a certain aspect ratio or volume fraction of particles in the composite that would produce the strongest composite under all loading conditions. This observation suggests that the selection of MMCs for the most effective strengthening application depends on the specific applied service load conditions. While only initial yield surfaces are demonstrated, subsequent loading surfaces with nonzero equivalent plastic strain would be expected to be different, but following the isotropic hardening rule.

Finally, effects of the stress ratio R on the overall stress-strain behavior of MMCs with randomly oriented particles are shown in Fig. 10. The loading combination has a considerable effect on the overall response of the composites. As R increases from zero, the stress-strain relationship of the symmetrical x_1-axis direction of spheroidal inclusions tends to render more strengthening effect until the opposite plastic strain effect ("bendover") appears for

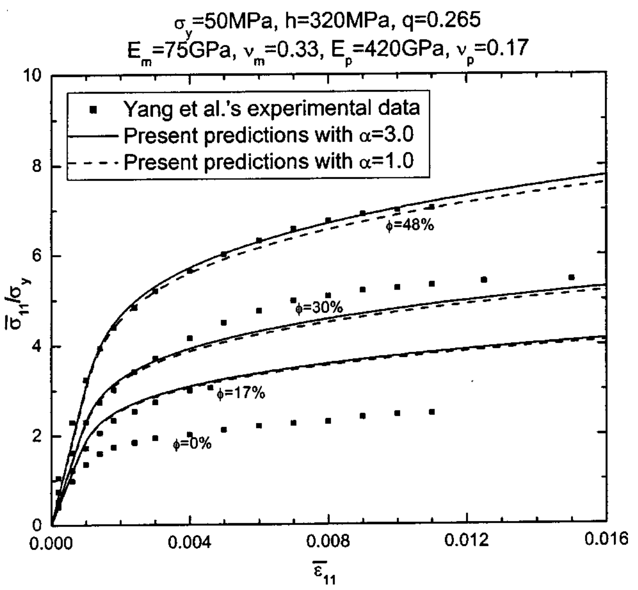


Fig. 8 Comparison between the model predictions and the experimental data of uniaxial elastoplastic behavior of randomly oriented particle-reinforced MMCs

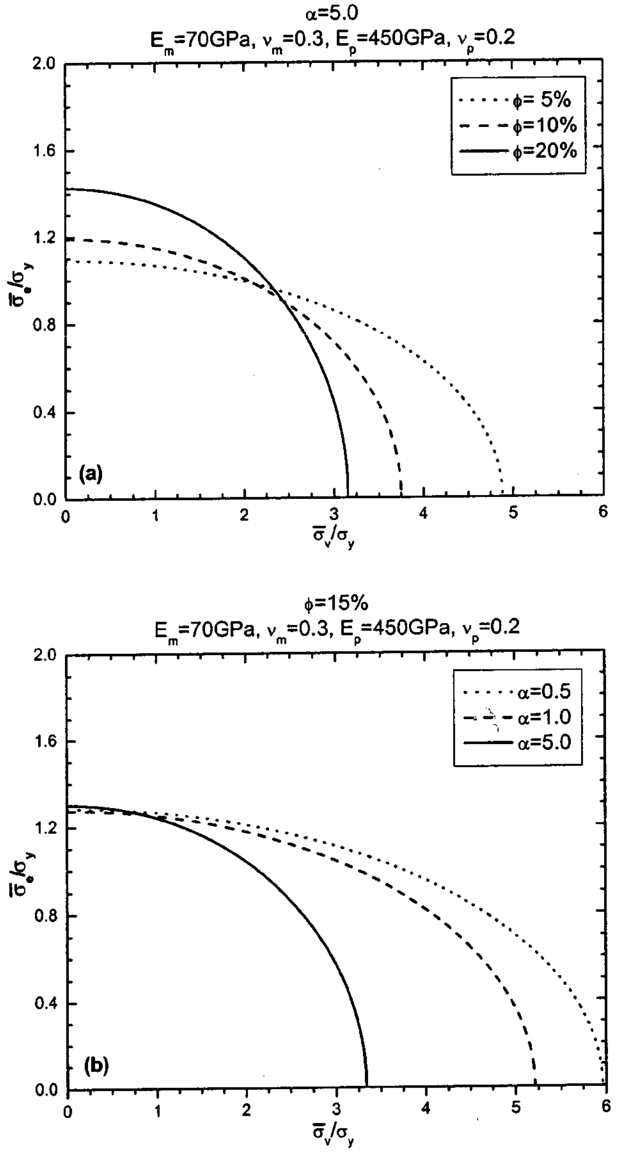


Fig. 9 Effects of (*a*) volume fraction and (*b*) aspect ratio of particles on the overall initial yield surfaces of randomly oriented particle-reinforced MMCs

$R>1$, as indicated in Fig. 10(*a*). The reason for the "bend-over" behavior is that the plastic strain $\langle \overline{\varepsilon}^p_{11} \rangle$ may become negative due to the fact that the magnitude of combined negative lateral plastic strains from the 22- and 33-directions gradually becomes greater than that from the 11-direction. It is also noted that effective stresses are high when the stress ratio R is close to 1, indicating that the local stress level in randomly oriented particles may exceed the fracture strength of particles and/or interfacial bonding between the matrix and particle. Correspondingly, particle cracking and/or interfacial debonding may occur under such circumstance. Future research is focused on the detailed interfacial debonding processes.

From Fig. 10(*b*), it is observed that the effect of the aspect ratio of particles is no longer insignificant when stress ratio R increases. The aspect ratio of randomly oriented particles plays a significant role on the elastoplastic responses of MMCs under purely hydrostatic loading.

4 Concluding Remarks

The effective elastic and elastoplastic models of MMCs with randomly located and oriented reinforcements are developed. The averaging process over all orientations upon three elastic governing equations for aligned particle-reinforced composites is performed to obtain the effective constitutive relations and isotropic elastic stiffness of MMCs with randomly oriented particles. The effects of volume fraction of particles and particle shape on the overall elastic constants of the composite are presented. Comparisons with the elasticity bounds show that the present analytical formulation does not violate the variational bounds. Good agreement with experimental data is also illustrated.

Furthermore, the orientational averaging procedure is utilized to derive the overall yield function for the underlying MMCs. Elastoplastic constitutive relations for the MMCs are systematically established. The overall elastoplastic stress-strain behaviors under monotonic uniaxial tensile loading are studied in detail. Furthermore, comparisons between the present analytical predictions and experimental data are performed to illustrate the capability of the proposed formulation. Finally, the initial yield surfaces and effects of stress ratio R upon the MMCs under axisymmetric loading are also discussed.

The proposed micromechanics-based homogenization framework allows the overall elastic and plastic behavior of randomly oriented particle-reinforced MMCs to be analytically and explicitly predicted in terms of the 3D far-field loading histories and microstructural information, such as the mechanical properties of the constituent phases and the spatial concentration and geometry of reinforcements. Good agreement between the theoretical predictions and experimental data are observed and provide preliminary assessment of the capability of the proposed formulation in quantitatively predicting the overall elastoplastic behavior of the MMCs.

In the present study, with an assumption of J_2 von Mises plastic flow in the metal matrix, the overall elastoplastic yield function of MMCs is micromechanically derived to be quadratic and pressure-dependent, which is not of the J_2 type anymore. However, in general, this form may not be universal for all MMCs. For example, Dvorak and co-workers [47,48] suggested that the overall yield function of anisotropic composites be constructed from piecewise smooth sections and not from a single smooth surface. We would need to consider more complicated local yield mecha-

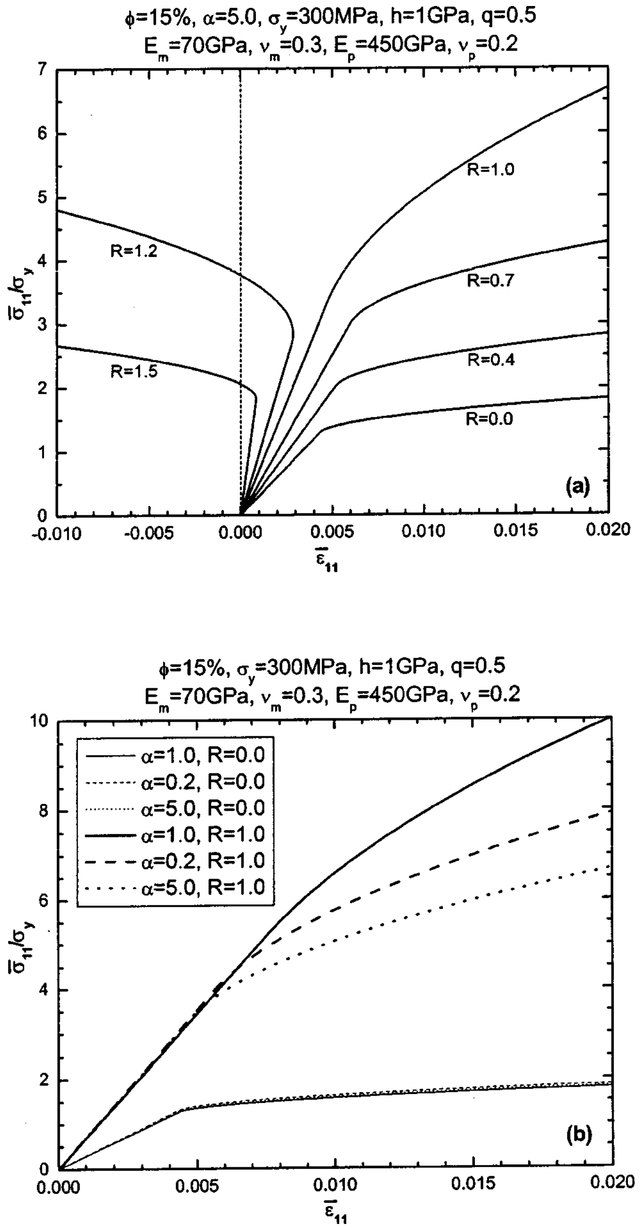


Fig. 10 Effects of stress ratio *R* on the overall elastoplastic behavior of randomly oriented particle-reinforced MMCs under axisymmetric loading

nisms in the matrix and connections between the matrix and reinforcements to render a sophisticated multisurface piecewise smooth elastoplasticity model. Furthermore, since isotropic hardening is postulated in the present model, elastoplastic responses of MMCs under complex loading such as cyclic loading should not render the Bauschinger effect which could be an essential component for fatigue behavior of MMCs. More general isotropic/kinematic hardening laws and an alternative nonassociated flow rule can be considered within the proposed framework based on reliable experimental data and evidence; see, e.g., Dvorak et al. [47] for kinematic hardening law and a nonassociated flow rule for boron-aluminum composites. These issues can be further investigated within the proposed context in the future, but with considerably more effort. It is noted, however, that the present paper does not purport to include all continuum plasticity aspects or features. Instead, we apply the micromechanical and ensemble-volume-orientational averaging methodology to the simple J_2-type plastic yield function with the power-law isotropic hardening rule and the associated flow rule as an illustration of the proposed concept and framework. In addition, we have not considered the effect of interfacial particle/matrix debonding upon the overall elastoplastic behavior. Future work is needed to improve the proposed method.

Acknowledgments

This work is sponsored by the NASA Iowa Space Grant Consortium, the University of Iowa Central Investment Fund for Research Enhancement (UI-CIFRE), the National Science Foundation PYI Grant No. MSS-9157238, the AFOSR Grant No. F49620-96-1-0384, and the Faculty Research Grant of the Academic Senate of the University of California—Los Angeles (Fund No. 4-592565-19914). These supports are greatly acknowledged. In addition, the authors are grateful for the helpful comments provided by the reviewers.

Appendix A

The two second-rank tensors $S^{(1)}_{IK}$ and $S^{(2)}_{IJ}$ in Eq. (6) are expressed as

$$S^{(1)}_{11}=\left[4\nu_0+\frac{2}{\alpha^2-1}\right]g(\alpha)+4\nu_0+\frac{4}{3(\alpha^2-1)}$$

$$S^{(1)}_{12}=S^{(1)}_{13}=\left[4\nu_0-\frac{2\alpha^2+1}{\alpha^2-1}\right]g(\alpha)+4\nu_0-\frac{2\alpha^2}{\alpha^2-1}$$

$$S^{(1)}_{21}=S^{(1)}_{31}=\left[-2\nu_0-\frac{2\alpha^2+1}{\alpha^2-1}\right]g(\alpha)-\frac{2\alpha^2}{\alpha^2-1}$$

$$S^{(1)}_{22}=S^{(1)}_{23}=S^{(1)}_{32}=S^{(1)}_{33}=\left[-2\nu_0+\frac{4\alpha^2-1}{4(\alpha^2-1)}\right]g(\alpha)+\frac{\alpha^2}{2(\alpha^2-1)}$$

$$S^{(2)}_{11}=\left[-4\nu_0+\frac{4\alpha^2-2}{\alpha^2-1}\right]g(\alpha)-4\nu_0+\frac{12\alpha^2-8}{3(\alpha^2-1)}$$

$$S^{(2)}_{12}=S^{(2)}_{13}=S^{(2)}_{21}=S^{(2)}_{31}=\left[-\nu_0-\frac{\alpha^2+2}{\alpha^2-1}\right]g(\alpha)-2\nu_0-\frac{2}{\alpha^2-1}$$

$$S^{(2)}_{22}=S^{(2)}_{23}=S^{(2)}_{32}=S^{(2)}_{33}=\left[2\nu_0-\frac{4\alpha^2-7}{4(\alpha^2-1)}\right]g(\alpha)+\frac{\alpha^2}{2(\alpha^2-1)}$$

with α denoting the aspect ratio of particles, ν_0 denoting the Poisson ratio of the matrix, and

$$g(\alpha)=\begin{cases}\dfrac{\alpha}{(1-\alpha^2)^{3/2}}[\alpha(1-\alpha^2)^{1/2}-\cos^{-1}\alpha], & \alpha<1\\[2ex] \dfrac{\alpha}{(\alpha^2-1)^{3/2}}[\cosh^{-1}\alpha-\alpha(\alpha^2-1)^{1/2}], & \alpha>1\end{cases}$$

The two second-rank tensors $T^{(1)}_{IK}$ and $T^{(2)}_{IJ}$ in Eq. (29) are expressed as

$$\begin{aligned}T^{(1)}_{IK}=&-\frac{1}{3}+\frac{2\phi}{4725(1-\nu_0)^2B_{II}B_{KK}}[1575(1-2\nu_0)^2\Gamma_{II}\Gamma_{KK}\\&+21(25\nu_0-23)(1-2\nu_0)(\Gamma_{II}\Delta_K+\Gamma_{KK}\Delta_I)+21(25\nu_0-2)\\&\times(1-2\nu_0)(\Gamma_{II}+\Gamma_{KK})+3(35\nu_0^2-70\nu_0+36)\Delta_{IK}+7(50\nu_0^2\\&-59\nu_0+8)(\Delta_I+\Delta_K)-2(175\nu_0^2-343\nu_0+103)]\end{aligned}$$

$$T_{IJ}^{(2)}=\frac{1}{2}+\frac{\phi}{1575(1-\nu_0)^2B_{IJ}B_{IJ}}\Bigg[(70\nu_0^2-140\nu_0-72)\Delta_{IJ}$$

$$-(175\nu_0^2-266\nu_0+75)\,\frac{\Delta_I+\Delta_J}{2}+350\nu_0^2-476\nu_0+164\Bigg]$$

where

$$B_{IJ}=2[Z_2+S_{IJ}^{(2)}]$$

$$\Delta_1=\frac{3[1-\alpha^4f(\alpha^2)]}{1-\alpha^4},\quad \Delta_2=\Delta_3=\frac{1}{2}(3-\Delta_1)$$

$$\Delta_{11}=\frac{5[2+\alpha^4-3\alpha^4f(\alpha^2)]}{2(1-\alpha^4)^2}$$

$$\Delta_{12}=\Delta_{21}=\Delta_{13}=\Delta_{31}=\frac{15\alpha^4[-3+(1+2\alpha^4)f(\alpha^2)]}{4(1-\alpha^4)^2}$$

$$\Delta_{22}=\Delta_{23}=\Delta_{32}=\Delta_{33}=\frac{15\alpha^4[1+2\alpha^4+(1-4\alpha^4)f(\alpha^2)]}{16(1-\alpha^4)^2}$$

with

$$f(\alpha)=\begin{cases}\dfrac{\cos^{-1}\alpha}{\alpha\sqrt{1-\alpha^2}}, & \alpha<1\\[2ex] \dfrac{\cosh^{-1}\alpha}{\alpha\sqrt{\alpha^2-1}}, & \alpha>1\end{cases}$$

Appendix B

Consider a transversely isotropic fourth-rank tensor **Q** of the following type:

$$Q_{ijkl}=Q_{IK}^{(1)}\delta_{ij}\delta_{kl}+Q_{IJ}^{(2)}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk})$$

where $Q_{IK}^{(1)}$ and $Q_{IJ}^{(2)}$ are the second-rank tensors and $Q_{IJ}^{(2)}$ is symmetric. The inverse of the tensor **Q** takes the form from Sun [49]:

$$Q_{ijkl}^{-1}=-\frac{Q_{IK}^{(3)}}{2Q_{II}^{(2)}}\delta_{ij}\delta_{kl}+\frac{1}{4Q_{IJ}^{(2)}}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk})$$

where the second-rank tensor $Q_{IK}^{(3)}$ can be calculated from

$$\begin{Bmatrix}Q_{I1}^{(3)}\\Q_{I2}^{(3)}\\Q_{I3}^{(3)}\end{Bmatrix}=\begin{bmatrix}Q_{11}^{(1)}+2Q_{11}^{(2)} & Q_{21}^{(1)} & Q_{31}^{(1)}\\ Q_{12}^{(1)} & Q_{22}^{(1)}+2Q_{22}^{(2)} & Q_{32}^{(1)}\\ Q_{13}^{(1)} & Q_{23}^{(1)} & Q_{33}^{(1)}+2Q_{33}^{(2)}\end{bmatrix}^{-1}$$

$$\times\begin{Bmatrix}Q_{I1}^{(1)}\\Q_{I2}^{(1)}\\Q_{I3}^{(1)}\end{Bmatrix}\quad (I=1,2,3)$$

References

[1] Taya, M., and Arsenault, R. J., 1989, *Metal Matrix Composites, Thermomechanical Behavior*, Pergamon Press, New York.

[2] Cox, H. L., 1952, ''The Elasticity and Strength of Paper and Other Fibrous Materials,'' British J. Appl. Phys., **3**, pp. 72–79.

[3] Christensen, R. M., and Waals, F. M., 1972, ''Effective Stiffness of Randomly Oriented Fiber Composites,'' J. Compos. Mater., **6**, pp. 518–532.

[4] Christensen, R. M., 1976, ''Asymptotic Modulus Results for Composites Containing Randomly Oriented Fibers,'' Int. J. Solids Struct., **12**, pp. 537–544.

[5] Chou, T. W., and Nomura, S., 1981, ''Fiber Orientation Effects on the Thermoelastic Properties of Short-Fiber Composites,'' Fibre Sci. Technol., **14**, pp. 279–291.

[6] Takao, Y., Chou, T. W., and Taya, M., 1982, ''Effective Longitudinal Young's Modulus of Misoriented Short Fiber Composites,'' ASME J. Appl. Mech., **49**, pp. 536–540.

[7] Tandon, G. P., and Weng, G. J., 1986, ''Average Stress in the Matrix and Effective Moduli of Randomly Oriented Composites,'' Compos. Sci. Technol., **27**, pp. 111–132.

[8] Benveniste, Y., 1987, ''A New Approach to the Application of Mori-Tanaka's Theory in Composite Materials,'' Mech. Mater., **6**, pp. 147–157.

[9] Ferrari, M., and Johnson, G. C., 1989, ''Effective Elasticities of Short-Fiber Composites With Arbitrary Orientation Distribution,'' Mech. Mater., **8**, pp. 67–73.

[10] Haddad, Y. M., 1992, ''On the Deformation Theory of a Class of Randomly Structured Composite Systems,'' ASME J. Energy Resour. Technol., **114**, pp. 110–116.

[11] Chen, T., Dvorak, G. J., and Benveniste, Y., 1992, ''Mori-Tanaka Estimates of the Overall Elastic Moduli of Certain Composite Materials,'' ASME J. Appl. Mech., **59**, pp. 539–546.

[12] Banerjee, P. K., and Henry, D. P., 1992, ''Elastic Analysis of Three-Dimensional Solids With Fiber Inclusions by BEM,'' Int. J. Solids Struct., **29**, pp. 2423–2440.

[13] Sayers, C. M., 1992, ''Elastic Anisotropy of Short-Fiber Reinforced Composites,'' Int. J. Solids Struct., **29**, pp. 2933–2944.

[14] Papathanasiou, T. D., Ingber, M. S., Mondy, L. A., and Graham, A. L., 1994, ''The Effective Elastic Modulus of Fiber-Reinforced Composites,'' J. Compos. Mater., **28**, pp. 288–304.

[15] Ponte Castaneda, P., and Willis, J. R., 1995, ''The Effect of Spatial Distribution on the Effective Behavior of Composite Materials and Cracked Media,'' J. Mech. Phys. Solids, **43**, pp. 1919–1951.

[16] Dunn, M. L., Ledbetter, H., Heyliger, P. R., and Choi, C. S., 1996, ''Elastic Constants of Textured Short-Fiber Composites,'' J. Mech. Phys. Solids, **44**, pp. 1509–1541.

[17] Luo, J., and Stevens, R., 1996, ''Micromechanics of Randomly Oriented Ellipsoidal Inclusion Composites, Part I: Stress, Strain and Thermal Expansion,'' J. Appl. Phys., **79**, pp. 9047–9056.

[18] Luo, J., and Stevens, R., 1996, ''Micromechanics of Randomly Oriented Ellipsoidal Inclusion Composites, Part II: Elastic Moduli,'' J. Appl. Phys., **79**, pp. 9057–9063.

[19] Riccardi, A., and Montheillet, F., 1999, ''A Generalized Self-Consistent Method for Solids Containing Randomly Oriented Spheroidal Inclusions,'' Acta Mech., **133**, pp. 39–56.

[20] Huang, J. H., 2001, ''Some Closed-Form Solutions for Effective Moduli of Composites Containing Randomly Oriented Short Fibers,'' Mater. Sci. Eng., A, **A315**, pp. 11–20.

[21] Mori, T., and Tanaka, K., 1973, ''Average Stress in Matrix and Average Elastic Energy of Materials With Misfitting Inclusions,'' Acta Metall., **21**, pp. 571–574.

[22] Qiu, Y. P., and Weng, G. J., 1991, ''The Influence of Inclusion Shape on the Overall Elastoplastic Behavior of a Two-Phase Isotropic Composite,'' Int. J. Solids Struct., **27**, pp. 1537–1550.

[23] Qiu, Y. P., and Weng, G. J., 1993, ''Plastic Potential and Yield Function of Porous Materials With Aligned and Randomly Oriented Spheroidal Voids,'' Int. J. Plast., **9**, pp. 271–290.

[24] Tandon, G. P., and Weng, G. J., 1988, ''A Theory of Particle-Reinforced Plasticity,'' ASME J. Appl. Mech., **55**, pp. 126–135.

[25] Qiu, Y. P., and Weng, G. J., 1992, ''A Theory of Plasticity for Porous Materials and Particle-Reinforced Composites,'' ASME J. Appl. Mech., **59**, pp. 261–268.

[26] Li, G., Ponte Castaneda, P., and Douglas, A. S., 1993, ''Constitutive Models for Ductile Solids Reinforced by Rigid Spheroidal Inclusions,'' Mech. Mater., **15**, pp. 279–300.

[27] Ponte Castaneda, P., 1991, ''The Effective Mechanical Properties of Nonlinear Isotropic Composites,'' J. Mech. Phys. Solids, **39**, pp. 45–71.

[28] Ponte Castaneda, P., 1992, ''New Variational Principles in Plasticity and Their Application to Composite Materials,'' J. Mech. Phys. Solids, **40**, pp. 1757–1788.

[29] Dunn, M. L., and Ledbetter, H., 1997, ''Elastic-Plastic Behavior of Textured Short-Fiber Composites,'' Acta Mater., **45**, pp. 3327–3340.

[30] Bao, G., Hutchinson, J. W., and McMeeking, R. M., 1991, ''Particle Reinforcement of Ductile Matrices Against Plastic Flow and Creep,'' Acta Metall. Mater., **39**, pp. 1871–1882.

[31] Sorensen, N. J., Suresh, S., Tvergaard, V., and Needleman, A., 1995, ''Effects of Reinforcement Orientation on the Tensile Response of Metal-Matrix Composites,'' Mater. Sci. Eng., A, **197**, pp. 1–10.

[32] Dong, M., Schmauder, S., Bidlingmaier, T., and Wanner, A., 1997, ''Prediction of the Mechanical Behavior of Short Fiber Reinforced MMCs by Combined Cell Models,'' Comput. Mater. Sci., **9**, pp. 121–133.

[33] Hashin, Z., and Shtrikman, S., 1963, ''A Variational Approach to the Theory of the Elastic Behavior Multiphase Materials,'' J. Mech. Phys. Solids, **11**, pp. 127–140.

[34] Ju, J. W., and Sun, L. Z., 2001, ''Effective Elastoplastic Behavior of Metal Matrix Composites Containing Randomly Located Aligned Spheroidal Inhomogeneities, Part I: Micromechanics-Based Formulation,'' Int. J. Solids Struct., **38**, pp. 183–201.

[35] Sun, L. Z., and Ju, J. W., 2001, ''Effective Elastoplastic Behavior of Metal Matrix Composites Containing Randomly Located Aligned Spheroidal Inhomogeneities, Part II: Applications,'' Int. J. Solids Struct., **38**, pp. 203–225.

[36] Ju, J. W., and Chen, T. M., 1994, ''Micromechanics and Effective Moduli of Elastic Composites Containing Randomly Dispersed Ellipsoidal Inhomogeneities,'' Acta Mech., **103**, pp. 103–121.

[37] Ju, J. W., and Sun, L. Z., 1999, ''A Novel Formulation for the Exterior-Point Eshelby's Tensor of an Ellipsoidal Inclusion,'' ASME J. Appl. Mech., **66**, pp. 570–574.

[38] Mura, T., 1987, *Micromechanics of Defects in Solids*, Second Edition, Martinus Nijhoff, The Netherlands.

[39] Hill, R., 1963, ''Elastic Properties of Reinforced Solids: Some Theoretical Principles,'' J. Mech. Phys. Solids, **11**, pp. 357–372.

[40] Walpole, L. J., 1966, ''On Bounds for the Overall Elastic Moduli of Inhomogeneous Systems—I,'' J. Mech. Phys. Solids, **14**, pp. 151–162.

[41] Walpole, L. J., 1966, ''On Bounds for the Overall Elastic Moduli of Inhomogeneous Systems—II,'' J. Mech. Phys. Solids, **14**, pp. 289–301.

[42] Christensen, R. M., 1991, *Mechanics of Composite Materials*, Krieger Publishing, Melbourne, Florida.

[43] Wu, T. T., 1966, ''The Effect of Inclusion Shape on the Elastic Moduli of a Two-Phase Material,'' Int. J. Solids Struct., **2**, pp. 1–8.

[44] Willis, J. R., 1977, ''Bounds and Self-Consistent Estimates for the Overall Moduli of Anisotropic Composites,'' J. Mech. Phys. Solids, **25**, pp. 185–202.

[45] Yang, J., Pickard, S. M., Cady, C., Evans, A. G., and Mehrabian, R., 1991, ''The Stress/Strain Behavior of Aluminum Matrix Composites With Discontinuous Reinforcements,'' Acta Metall. Mater., **39**, pp. 1863–1869.

[46] Li, G., and Ponte Castaneda, P., 1994, ''Variational Estimates for the Elastoplastic Response of Particle-Reinforced Metal-Matrix Composites,'' Appl. Mech. Rev., **47**, pp. S77–S94.

[47] Dvorak, G. J., Bahei-El-Din, Y. A., Macheret, Y., and Liu, C. H., 1988, ''An Experimental Study of Elastic-Plastic Behavior of a Fibrous Boron-Aluminum Composite,'' J. Mech. Phys. Solids, **36**, pp. 655–687.

[48] Dvorak, G. J., and Bahei-El-Din, A., 1997, ''Inelastic Composite Materials: Tranformation Analysis and Experiments,'' *Continuum Micromechanics*, P. Suquet, ed., CISM Courses and Lectures No. 377, Springer-Verlag, Berlin, pp. 1–59.

[49] Sun, L. Z., 1998, ''Micromechanics and Overall Elastoplasticity of Discontinuously Reinforced Metal Matrix Composites,'' Ph.D. thesis, University of California—Los Angeles.

Amnon J. Meir

Paul G. Schmidt

Department of Mathematics,
Auburn University,
232 Parker Hall,
Auburn University, AL 36849-5310

Sayavur I. Bakhtiyarov

Ruel A. Overfelt

Department of Mechanical Engineering,
Auburn University,
202 Ross Hall,
Auburn University, AL 36849-5341

Numerical Simulation of Steady Liquid-Metal Flow in the Presence of a Static Magnetic Field

We describe a novel approach to the mathematical modeling and computational simulation of fully three-dimensional, electromagnetically and thermally driven, steady liquid-metal flow. The phenomenon is governed by the Navier-Stokes equations, Maxwell's equations, Ohm's law, and the heat equation, all nonlinearly coupled via Lorentz and electromotive forces, buoyancy forces, and convective and dissipative heat transfer. Employing the electric current density rather than the magnetic field as the primary electromagnetic variable, it is possible to avoid artificial or highly idealized boundary conditions for electric and magnetic fields and to account exactly for the electromagnetic interaction of the fluid with the surrounding media. A finite element method based on this approach was used to simulate the flow of a metallic melt in a cylindrical container, rotating steadily in a uniform magnetic field perpendicular to the cylinder axis. Velocity, pressure, current, and potential distributions were computed and compared to theoretical predictions. [DOI: 10.1115/1.1796450]

1 Introduction

Numerous industrial processes involve the flow of a liquid metal in the presence of an applied magnetic field. *Static* magnetic fields are used, for example, to dampen undesirable melt flow, such as buoyancy-driven convection, in metal casting and crystal-growth processes [1,2]. When applied in conjunction with DC currents, static magnetic fields can also be used to activate or drive liquid-metal flows; this is exploited, for example, in MHD pumps and reactor cooling blankets [3,4]. *Rotating* magnetic fields have long been applied in industrial devices, such as induction furnaces and electromagnetic stirrers [5,6]. More recently, both static and rotating fields have garnered much attention as a means of controlling the melt flow in crystal-growth processes [7–11].

In all of these and many other processes, the flow of a metallic melt is modified by means of Lorentz forces, resulting from the interaction of an applied magnetic field with impressed or induced currents in the melt. Qualitative and quantitative understanding of the ensuing flow patterns is of critical importance in achieving optimal operating conditions. Efficient and accurate experimental and computational techniques are needed to measure and to predict velocity, current, and temperature distributions in the melt. Despite considerable research efforts over the past three decades, the methodology is still in need of development. This is due to the complexity of the underlying flow phenomena, which are, in general, fully three-dimensional, highly nonlinear, and characterized by the interaction of multiple physical effects.

Much of the pertinent literature on analytical and numerical issues is based on simplifying assumptions that lead to spatially two-dimensional (2D) problems and at least partial decoupling of the underlying PDEs, allowing, in particular, the separation of fluid flow and electromagnetic computations. Induced magnetic fields and buoyancy effects due to temperature fluctuations in the melt are frequently neglected, and so is the electromagnetic interaction of the melt with other conductors in the vicinity.

There is, for example, a substantial body of analytical and numerical work devoted to the electromagnetic stirring of round strands of liquid metal by the rotating-field method; see Moffatt [12], Spitzer et al. [6], or Davidson and Hunt [13] for seminal studies. In this method, a rotating magnetic field is generated in an inductor, surrounding a stationary column of liquid metal. The traveling magnetic field induces eddy currents in the melt, and it is the interaction of the magnetic field with these currents that brings about the stirring effect. If one assumes a circular-cylindrical geometry and averages the electromagnetic body force azimuthally, the problem becomes axisymmetric and thus, spatially two-dimensional. Its numerical solution then amounts to solving the stationary Navier-Stokes equations in two dimensions, under a known distribution of time-averaged Lorentz forces. The latter are determined from an asymptotic solution of the magnetic induction equation under appropriate boundary conditions for the magnetic field. Witkowski et al. [10] investigated the validity of azimuthal averaging and found the deviation from axisymmetry to be small under suitable assumptions on the flow parameters. However, secondary effects due to induced magnetic fields or buoyancy forces were neglected. Fully three-dimensional computational models, accounting for magnetic induction and/or thermal effects, have been proposed only recently [14,15].

Although based on the same general principle, the effect of rotating the melt in a stationary, transverse magnetic field is quite different from that of rotational stirring. In particular, the resulting flow field cannot be expected to be rotationally symmetric; in fact, the induced flow is symmetric with respect to the given axis of the magnetic field rather than the axis of rotation (see the discussion at the end of Section 4). As a consequence, the flow is fully three-dimensional, even in the simplest experimental configurations (cylindrical crucible of circular cross section, uniform magnetic field perpendicular to the cylinder axis, uniform temperature distribution). Moreover, careful analysis shows that, in general, secondary effects due to induced magnetic fields and temperature fluctuations in the melt are not negligible and of comparable magnitude (see the discussion at the end of Section 2).

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, December 27, 2002; final revision, April 24, 2004. Associate Editor: D. A. Siginer. 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.

While the importance of thermal effects has long been recognized [2,8,9], induced magnetic fields are rarely accounted for. The main problem, in this context, is the fact that electric and magnetic fields transcend the fluid region and must be determined in all of space (unless one imposes idealized or artificial boundary conditions on the surface of the fluid region). Accounting for both, magnetic induction and thermal effects leads to a fully coupled system of equations; fluid flow, heat flow, and electromagnetic computations cannot be separated, and the magnetic fields both inside and outside the fluid must be determined simultaneously with the fluid velocity and the current and temperature distributions in the melt.

The objective of this paper is to address these issues by presenting a new approach to the modeling and simulation of steady liquid-metal flow in the presence of a static magnetic field. Our approach differs from earlier work in that it is based on the direct numerical simulation of the full, nonlinear, three-dimensional, electromagnetically and thermally coupled flow problem. The mathematical foundation is a mixed variational formulation and finite element discretization of the (stationary) Navier-Stokes equations, Maxwell's equations, Ohm's law, and the heat equation, coupled via Lorentz and electromotive forces, buoyancy forces, and convective and dissipative heat transfer. One key feature is the use of the electric current density (rather than the magnetic field) as the primary electromagnetic variable. This approach, in which the induced magnetic field is computed from the current density via the Biot-Savart integral, renders idealized or artificial electromagnetic boundary conditions obsolete and allows us to account for both interior and exterior fields, while effectively confining computations to the fluid region itself. A similar ''integral-closure'' approach was developed by Natarajan and El-Kaddah [15–17], in the context of electromagnetic stirring and separation processes. We refer to Meir and Schmidt [18,19] for mathematical details regarding our method and to Meir and Schmidt [20,21] for prior applications to electromagnetically driven flows. While the present numerical implementation is limited to the simulation of *stationary* flow problems, our approach can be easily extended to the time-dependent case (see [22]). This extension, as well as the incorporation of a turbulence model, is the subject of ongoing research.

The paper is organized as follows. Section 2 contains a detailed description of the mathematical model and concludes with a careful order-of-magnitude analysis. Section 3 is concerned with a mixed variational formulation of the problem, its finite element discretization, and implementation issues. In Section 4 we describe preliminary computational experiments, thus far limited to the laminar flow regime and not accounting for temperature fluctuations, where our method is applied to simulate the flow of a metallic melt in a circular-cylindrical crucible, rotating steadily in a uniform magnetic field perpendicular to the cylinder axis. A corresponding experimental apparatus and measurement technique are described in a companion paper by Bakhtiyarov et al. [23].

2 Mathematical Model

We are concerned with the steady flow of a viscous, incompressible, electrically and thermally conducting fluid, confined to a vessel with solid walls, in the presence of gravity, an applied static magnetic field, and a radiative heat source; see Bakhtiyarov et al. [23] for a detailed description of the experimental apparatus that motivated this investigation. We assume the presence of a lid on top of the fluid, thus avoiding the complication of a free surface; also neglected is the thickness of the vessel walls. The problem is governed by balance equations for momentum, mass, and energy, along with Maxwell's equations and Ohm's law; see Hughes and Young [24] for the physical background.

The momentum balance is given by the stationary Navier-Stokes equations,

$$-\eta\nabla^2\mathbf{V}+\rho(\mathbf{V}\cdot\nabla)\mathbf{V}+\nabla p=\mathbf{F}\quad\text{(in the fluid)}\tag{1}$$

where $\mathbf{V}$ is the flow velocity, p the hydrodynamic pressure, $\mathbf{F}$ the sum of all body forces, including buoyancy and Lorentz forces. The dynamic viscosity η is assumed to be constant, while the density ρ is allowed, at least for now, to vary with temperature. Conservation of mass is enforced through the continuity equation,

$$\nabla\cdot(\rho\mathbf{V})=0\quad\text{(in the fluid)}\tag{2}$$

Equations (1) and (2) must be supplemented with a boundary condition for the fluid velocity. If the vessel is at rest, the usual no-slip condition for viscous fluids requires $\mathbf{V}$ to vanish at the walls. Allowing for a possible rigid motion of the vessel, we impose a more general boundary condition of the form

$$\mathbf{V}=\mathbf{V}_0\quad\text{(at the walls)}\tag{3}$$

where $\mathbf{V}_0$ denotes the velocity field associated with the rigid motion of the vessel.

The balance of energy can be written as a scalar convection-diffusion equation in terms of the temperature T,

$$-\kappa\nabla^2T+\rho c(\mathbf{V}\cdot\nabla)T=H\quad\text{(in the fluid)}\tag{4}$$

where H is the sum of all heat sources, including dissipative and radiative heating. The thermal conductivity κ and specific heat (at constant pressure) c are assumed to be constant in the relevant temperature range. As a boundary condition, we assume that the heat flux across the walls is proportional to the temperature difference between the fluid and the exterior of the vessel,

$$-\kappa(\nabla T)\cdot\mathbf{n}=\gamma(T-T_{\text{ext}})\quad\text{(at the walls)}\tag{5}$$

Here, $\mathbf{n}$ denotes the unit outward normal vector field on the surface of the fluid region; γ is the (constant) heat transfer coefficient of the walls, and T_{ext} stands for the given ambient temperature.

The electric current distribution is determined by Ohm's law,

$$\mathbf{J}=\sigma(-\nabla\phi+\mathbf{V}\times\mathbf{B})\quad\text{(in the fluid)}\tag{6}$$

along with the continuity equation,

$$\nabla\cdot\mathbf{J}=0\quad\text{(in the fluid)}\tag{7}$$

where $\mathbf{J}$ is the electric current density, ϕ a scalar electric potential, $\mathbf{B}$ the magnetic flux density, and σ the (constant) electric conductivity of the fluid. Since the exterior of the vessel is assumed to be nonconducting, the obvious boundary condition for $\mathbf{J}$ is that

$$\mathbf{J}\cdot\mathbf{n}=0\quad\text{(at the walls)}\tag{8}$$

The magnetic field can be decomposed as

$$\mathbf{B}=\mathbf{B}_a+\mathbf{B}_i$$

where $\mathbf{B}_a$ and $\mathbf{B}_i$ represent, respectively, the applied field and the field induced by the current $\mathbf{J}$. The latter satisfies Maxwell's equations,

$$\begin{aligned}\nabla\times\mathbf{B}_i&=\mu\mathbf{J}\quad\text{(in the fluid)}\\ \nabla\times\mathbf{B}_i&=0\quad\text{(in the exterior)}\\ \nabla\cdot\mathbf{B}_i&=0\quad\text{(throughout space)}\end{aligned}\tag{9}$$

Since the fluid is heated beyond the Curie point, μ is the magnetic permeability of free space. In addition, $\mathbf{B}_i$ must be continuous across the vessel walls (constant permeability, no surface currents) and vanish at infinity (finite source):

$$\begin{aligned}[\mathbf{B}_i]&=0\quad\text{(across the walls)}\\ \mathbf{B}_i&=0\quad\text{(at infinity)}\end{aligned}\tag{10}$$

For any reasonably regular current distribution $\mathbf{J}$, Eqs. (9) and (10) admit a unique solution $\mathbf{B}_i=\mathcal{B}(\mathbf{J})$, which is given by the Biot-Savart formula (a volume integral over the fluid region):

$$\mathcal{B}(\mathbf{J})(\mathbf{r})=-\frac{\mu}{4\pi}\int\frac{\mathbf{r}-\mathbf{s}}{|\mathbf{r}-\mathbf{s}|^3}\times\mathbf{J}(\mathbf{s})d^3\mathbf{s}\tag{11}$$

The body force $\mathbf{F}$ on the right-hand side of Eq. (1) includes the Lorentz force, $\mathbf{J}\times\mathbf{B}$ and the force of gravity $\rho\mathbf{g}$, where $\mathbf{g}$ is gravitational acceleration. We employ the Boussinesq approximation to account for buoyancy forces due to temperature gradients; that is, we assume that T fluctuates in a narrow range about a reference temperature T_{ref} and that the density, in this temperature range, decreases linearly with T:

$$\frac{\rho}{\rho_{\mathrm{ref}}}=1-\beta(T-T_{\mathrm{ref}})$$

Here, ρ_{ref} and β denote the density and thermal expansion coefficient of the fluid at the reference temperature. The force of gravity is then given by

$$\rho\mathbf{g}=\nabla p_{\mathrm{ref}}-\beta\rho_{\mathrm{ref}}(T-T_{\mathrm{ref}})\mathbf{g}$$

where $\nabla p_{\mathrm{ref}}=\rho_{\mathrm{ref}}\,\mathbf{g}$ is the hydrostatic pressure gradient at the reference temperature, while $\beta\rho_{\mathrm{ref}}(T-T_{\mathrm{ref}})\mathbf{g}$ represents the buoyancy force. Summarizing, we have

$$\mathbf{F}=\mathbf{J}\times[\mathbf{B}_a+\mathcal{B}(\mathbf{J})]+\nabla p_{\mathrm{ref}}-\beta\rho_{\mathrm{ref}}(T-T_{\mathrm{ref}})\mathbf{g} \tag{12}$$

Following standard practice, we now introduce a reduced pressure $p'=p-p_{\mathrm{ref}}$ and a reduced temperature $T'=T-T_{\mathrm{ref}}$, and we replace the density ρ in Eqs. (1), (2), and (4) by ρ_{ref}. Also, for notational convenience, we drop the primes in p' and T' and the subscript in ρ_{ref}; that is, from now on p, T, and ρ will denote the reduced pressure, reduced temperature, and reference density, respectively. (Note that Eqs. (4) and (5) remain unchanged after this reduction, except that T_{ext} must be replaced by $T_{\mathrm{ext}}-T_{\mathrm{ref}}$.)

The source term H in Eq. (4) comprises dissipative heating due to electric currents and viscous drag as well as radiative heating due to the presence of heating elements. Thus,

$$H=\sigma^{-1}|\mathbf{J}|^2+\tfrac{1}{2}\,\eta|\nabla\mathbf{V}+(\nabla\mathbf{V})^{\mathrm{tr}}|^2+h \tag{13}$$

where h is a given function of position, depending on the characteristics of the heating elements. For the apparatus described in Bakhtiyarov et al. [23], the following ansatz seems reasonable:

$$h(\mathbf{r})=I_0\chi(\mathbf{r})\exp[-\alpha d(\mathbf{r})] \tag{14}$$

Here, I_0 is the output intensity of the quartz lamps, $\chi(\mathbf{r})$ an empirical function describing the intensity distribution in the light cone, α the absorption coefficient of the fluid (in the relevant frequency range, at the reference temperature), and $d(\mathbf{r})$ the penetration depth. For any pair of reasonably regular vector fields $\mathbf{V}$ and $\mathbf{J}$, Eqs. (4) and (5), with H given by (13) and (14), constitute a linear, uniformly elliptic boundary-value problem, which admits a unique solution $T=\mathcal{T}(\mathbf{V},\mathbf{J})$.

We emphasize that, given any sufficiently regular velocity field $\mathbf{V}$ and current distribution $\mathbf{J}$, both the reduced temperature $\mathcal{T}(\mathbf{V},\mathbf{J})$ (the unique solution of the linear, elliptic boundary-value problem (4)–(5)) and the induced magnetic field $\mathcal{B}(\mathbf{J})$ (the unique solution of the linear div-curl system (9)–(10)) are relatively easy to compute in a variety of ways and to any desired numerical accuracy. In what follows, we will, therefore, focus our attention on the remaining coupled, nonlinear boundary-value problem for the velocity $\mathbf{V}$, current density $\mathbf{J}$, reduced pressure p, and electric potential ϕ, as obtained from Eqs. (1)–(3) (along with (12)) and (6)–(8):

$$\begin{aligned}
&-\eta\nabla^2\mathbf{V}+\rho(\mathbf{V}\cdot\nabla)\mathbf{V}+\nabla p\\
&\qquad=\mathbf{J}\times[\mathbf{B}_a+\mathcal{B}(\mathbf{J})]-\beta\rho\mathcal{T}(\mathbf{V},\mathbf{J})\mathbf{g}\quad\text{(in the fluid)}\\
&\sigma^{-1}\mathbf{J}+\nabla\phi=\mathbf{V}\times[\mathbf{B}_a+\mathcal{B}(\mathbf{J})]\quad\text{(in the fluid)}\\
&\nabla\cdot\mathbf{V}=0\quad\text{and}\quad\nabla\cdot\mathbf{J}=0\quad\text{(in the fluid)}\\
&\mathbf{V}=\mathbf{V}_0\quad\text{and}\quad\mathbf{J}\cdot\mathbf{n}=0\quad\text{(at the walls)}
\end{aligned} \tag{15}$$

Note that the above, due to the presence of the operators $\mathcal{B}$ and $\mathcal{T}$, is in fact a system of integro-differential equations ($\mathcal{B}$ is a linear, first-order integral operator, $\mathcal{T}$ a nonlinear, second-order integral operator). As a consequence, some care must be exercised in choosing linearization/iteration schemes for the numerical solution of the discretized equations, in order to avoid the occurrence of dense matrices (see Section 3).

To assess the relative importance of the various terms in the equations, it is useful to normalize the dependent and independent variables and to combine the physical parameters of the problem into nondimensional groups. To this end, let R be a characteristic length, for example, the radius of the vessel holding the fluid, and let V_0 be a characteristic speed, for example, the magnitude of the given velocity $\mathbf{V}_0$ of the walls ($V_0=\omega R$ in the case of a cylindrical vessel of radius R, rotating with angular velocity ω). Given a characteristic magnitude B_0 of the magnetic field $\mathbf{B}$, Ohm's law (6) yields $J_0=\sigma V_0 B_0$ as a characteristic magnitude of the current density $\mathbf{J}$. Moreover, Ampère's law, the first equation in (9), implies that the *induced* field $\mathbf{B}_i=\mathcal{B}(\mathbf{J})$ is of order $\mu R J_0=\mu\sigma R V_0 B_0$. The number $\mu\sigma R V_0$ (the magnetic Reynolds number) is small under laboratory conditions (see below); consequently, the induced field $\mathbf{B}_i$ is small compared to the total field, $\mathbf{B}=\mathbf{B}_a+\mathbf{B}_i$. We may, therefore, choose B_0 to be the magnitude of the *applied* field $\mathbf{B}_a$. A characteristic value T_0 of the temperature fluctuation in the fluid (that is, a typical deviation from the reference temperature T_{ref}) must be determined experimentally. In liquid-metal experiments, the metal is usually heated to just above its melting-point, which may then serve as the reference temperature T_{ref}.

Normalizing the spatial variables by R, the principal unknowns $\mathbf{V}$, $\mathbf{J}$, $\mathbf{B}$, and T by their respective characteristic values V_0, $J_0=\sigma V_0 B_0$, B_0, and T_0, the auxiliary unknown p and ϕ by ρV_0^2 and $\mu R^2 V_0 J_0$, and the source terms $\mathbf{g}$ and h by $g=|\mathbf{g}|$ and $\rho c R^{-1} V_0 T_0$, we obtain the following nondimensional versions of the momentum balance, Ohm's law, Ampère's law, and the heat equation:

$$-\mathrm{Re}^{-1}\nabla^2\mathbf{V}+(\mathbf{V}\cdot\nabla)\mathbf{V}+\nabla p=\mathrm{Re}^{-1}\,\mathrm{Ha}^2\mathbf{J}\times\mathbf{B}-\mathrm{Re}^{-2}\,\mathrm{Gr}T\mathbf{g}$$

$$\mathrm{Rm}^{-1}\mathbf{J}+\nabla\phi=\mathrm{Rm}^{-1}\mathbf{V}\times\mathbf{B}$$

$$\nabla\times\mathbf{B}_i=\mathrm{Rm}\mathbf{J}$$

$$-\mathrm{Pe}^{-1}\nabla^2 T+(\mathbf{V}\cdot\nabla)T=\mathrm{Re}^{-1}\,\mathrm{Ec}(\mathrm{Ha}^2|\mathbf{J}|^2+\tfrac{1}{2}|\nabla\mathbf{V}+(\nabla\mathbf{V})^{\mathrm{tr}}|^2)+h$$

Here, Re, Rm, and Pe denote the Reynolds number, magnetic Reynolds number, and Péclet number, respectively:

$$\mathrm{Re}=\eta^{-1}\rho R V_0,\quad \mathrm{Rm}=\sigma\mu R V_0,\quad \mathrm{Pe}=\kappa^{-1}\rho c R V_0$$

The remaining three nondimensional groups, Ha^2, Gr, and Ec, are the square of the Hartmann number, the Grasshoff number, and the Eckart number:

$$\mathrm{Ha}^2=\eta^{-1}\sigma R^2 B_0^2,\quad \mathrm{Gr}=\eta^{-2}\beta g\rho^2 R^3 T_0,\quad \mathrm{Ec}=c^{-1}T_0^{-1}V_0^2$$

The relative importance of the various terms in the equations is gauged by the ratios in the first and second columns of Table 1. The numerical values in the third column are for liquid aluminum at temperatures just above the melting-point (see Table 2); the orders of magnitude in the fourth column are estimates for the experimental apparatus described in Bakhtiyarov et al. [23], where $R\approx 1$ cm, $V_0\approx 20$ cm/s, $B_0\approx 0.1$ T, and $T_0\approx 10$ K. Conclusions, as pertaining to the laboratory conditions in [23], can be summarized as follows.

a. Lorentz forces and inertial forces are of comparable magnitude; both are large compared to buoyancy and viscous forces, which are comparable.
b. The induced magnetic field is small compared to the applied field (but less so if the vessel is rotating at high speed).
c. Convective heat transfer dominates diffusion (but less so if the vessel is rotating slowly). Viscous heating is small compared to Joule heating; both are negligible compared to diffusion and convection.

Table 1 Order-of-magnitude analysis (for liquid aluminum under laboratory conditions)

$\mathrm{Ha}^2/\mathrm{Re}$	Lorentz force to inertia	$2.14\times10^3\ RB_0^2/V_0$	$O(1)$
$\mathrm{Re}^2/\mathrm{Gr}$	inertia to buoyancy force	$8.79\times10^2\ V_0^2/(RT_0)$	$O(10^2)$
Gr/Re	buoyancy to viscous force	$9.34\times10^2\ R^2T_0/V_0$	$O(1)$
1/Rm	total magnetic field to induced field	$1.56\times10^{-1}/(RV_0)$	$O(10^2)$
Pe	thermal convection to diffusion	$2.50\times10^4\ RV_0$	$O(10)$
$\mathrm{Re}/(\mathrm{PeEcHa}^2)$	thermal diffusion to Joule heating	$2.02\times10^{-5}\ T_0/(R^2V_0^2B_0^2)$	$O(10^3)$
Ha^2	Joule heating to viscous heating	$1.76\times10^9\ R^2B_0^2$	$O(10^3)$

It is, therefore, feasible to neglect Joule and viscous heating when computing the temperature distribution. Neglecting temperature fluctuations (and thereby, buoyancy forces) altogether, that is, setting $\mathcal{T}(\mathbf{V},\mathbf{J})=0$, is viable only as a first approximation. The same can be said with regard to the induced magnetic field $\mathcal{B}(\mathbf{J})$.

3 Finite Element Discretization

The numerical solution of the boundary-value problem (15) is based on a mixed variational formulation in the spirit of the well-known Babuska-Brezzi theory; see, for example, Brezzi and Fortin [26]. This formulation is obtained by multiplying the four PDEs by suitable test functions $\mathbf{W}$ (for the velocity), $\mathbf{K}$ (for the current density), q (for the pressure), and ψ (for the electric potential) and integrating the equations over the fluid region. The two identities resulting from the momentum balance and Ohm's law are added together, and so are the identities resulting from the two continuity equations. After some algebra and several integrations by parts (using the boundary conditions), one arrives at two equations of the form

$$a_0((\mathbf{V},\mathbf{J}),(\mathbf{W},\mathbf{K}))+a_1((\mathbf{V},\mathbf{J}),(\mathbf{V},\mathbf{J}),(\mathbf{W},\mathbf{K}))+b((\mathbf{W},\mathbf{K}),(p,\phi))$$
$$=f((\mathbf{V},\mathbf{J}),(\mathbf{W},\mathbf{K})) \tag{16}$$

$$b((\mathbf{V},\mathbf{J}),(q,\psi))=0 \tag{17}$$

where a_0 (a bilinear form), a_1 (a trilinear form), and b (a bilinear form) are given by

$$a_0((\mathbf{W}_1,\mathbf{K}_1),(\mathbf{W}_2,\mathbf{K}_2))=\eta\int(\nabla\mathbf{W}_1)\cdot(\nabla\mathbf{W}_2)+\sigma^{-1}\int\mathbf{K}_1\cdot\mathbf{K}_2+\int((\mathbf{K}_2\times\mathbf{B}_a)\cdot\mathbf{W}_1-(\mathbf{K}_1\times\mathbf{B}_a)\cdot\mathbf{W}_2)$$

$$a_1((\mathbf{W}_1,\mathbf{K}_1),(\mathbf{W}_2,\mathbf{K}_2),(\mathbf{W}_3,\mathbf{K}_3))=\frac{\rho}{2}\int[((\mathbf{W}_1\cdot\nabla)\mathbf{W}_2)\cdot\mathbf{W}_3-((\mathbf{W}_1\cdot\nabla)\mathbf{W}_3)\cdot\mathbf{W}_2]+\int\{[\mathbf{K}_3\times\mathcal{B}(\mathbf{K}_1)]\cdot\mathbf{W}_2-[\mathbf{K}_2\times\mathcal{B}(\mathbf{K}_1)]\cdot\mathbf{W}_3\}$$

$$b((\mathbf{W},\mathbf{K}),(q,\psi))=-\int(\nabla\cdot\mathbf{W})q+\int\mathbf{K}\cdot(\nabla\psi)$$

while f (a nonlinear forcing term) is given by

$$f((\mathbf{W}_1,\mathbf{K}_1),(\mathbf{W}_2,\mathbf{K}_2))=-\beta\rho\int\mathcal{T}(\mathbf{W}_1,\mathbf{K}_1)\mathbf{g}\cdot\mathbf{W}_2$$

Under mild assumptions on the data, the original boundary-value problem (15) is equivalent to the following *weak problem*: Find vector fields $\mathbf{V}$ and $\mathbf{J}$ and scalar fields p and ϕ such that $\mathbf{V}=\mathbf{V}_0$ on the boundary of the fluid region and Eqs. (16) and (17) are satisfied for all relevant test functions $\mathbf{W}$, $\mathbf{K}$, q, and ψ. The boundary condition for $\mathbf{V}$ is an *essential* boundary condition and must be enforced explicitly; as a consequence, the velocity test functions must vanish on the boundary. The boundary condition for $\mathbf{J}$ is a *natural* boundary condition and can be recovered from Eq. (17). We refer to Meir and Schmidt [18,19] for mathematical details, including the choice of appropriate test function spaces and a rigorous proof that the weak problem, as stated above, is equivalent to the original boundary-value problem (15).

Note that Eq. (16) incorporates the first two PDEs in (15), that is, the momentum balance and Ohm's law, while Eq. (17) incorporates the divergence constraints on $\mathbf{V}$ and $\mathbf{J}$. In the weak formulation of the problem, the pressure p and potential ϕ play the role of Lagrange multipliers associated with these divergence constraints. Numerically, both will be determined simultaneously with $\mathbf{V}$ and $\mathbf{J}$, just as the pressure is determined simultaneously with the velocity field in the standard mixed variational formulation of the Navier-Stokes equations (see, for example, [27]).

A finite element discretization of the weak problem is obtained by requiring that Eqs. (16) and (17) be satisfied for only a finite number of test functions, namely, the basis functions of suitably chosen finite element spaces. Also, the essential boundary condition for $\mathbf{V}$ must be approximately satisfied in an appropriate sense, for example, by requiring that $\mathbf{V}=\mathbf{V}_0$ at the boundary nodes of the finite element grid (assuming the use of Lagrangian elements). This leads to a finite-dimensional system of nonlinear, algebraic equations that can be solved by way of linearization and iteration.

The nonlinear equation (16) is linearized by replacing the first arguments of the forms a_1 and f by initial guesses or previously computed values $\mathbf{V}_{\mathrm{old}}$ and $\mathbf{J}_{\mathrm{old}}$ for the velocity field and current density. In terms of the original PDEs, this amounts to lagging the first velocity in the inertia term $(\mathbf{V}\cdot\nabla)\mathbf{V}$, the induced magnetic field $\mathcal{B}(\mathbf{J})$, and the temperature distribution $\mathcal{T}(\mathbf{V},\mathbf{J})$. Lagging magnetic field and temperature also prevents the occurrence of dense matrices despite the presence of the integral operators $\mathcal{B}$ and $\mathcal{T}$. Given an initial guess or previously computed pair $(\mathbf{V}_{\mathrm{old}},\mathbf{J}_{\mathrm{old}})$, the linearized equations are solved to update $(\mathbf{V},\mathbf{J})$ and to compute (p,ϕ). The solution is unique only up to additive constants in p and ϕ; but this is easily dealt with by setting p and ϕ equal to zero at one node each of the finite element grid. The process is iterated until the change in $(\mathbf{V},\mathbf{J})$, as measured in a suitable norm, drops below a given tolerance. Of course, at the beginning of each iteration, the induced magnetic field $\mathcal{B}(\mathbf{J}_{\mathrm{old}})$ and the temperature distribution $\mathcal{T}(\mathbf{V}_{\mathrm{old}},\mathbf{J}_{\mathrm{old}})$ must be determined. The field $\mathcal{B}(\mathbf{J}_{\mathrm{old}})$ can be computed by evaluating the Biot-Savart integral (11) or by solving the linear div-curl system (9) and (10) in any other way. The temperature $\mathcal{T}(\mathbf{V}_{\mathrm{old}},\mathbf{J}_{\mathrm{old}})$ is obtained by solving the linear, elliptic boundary-value problem (4) and (5), with H given by (13) and (14), for example, using a standard finite element discretiza-

Table 2 Universal constants and properties of liquid aluminum at temperatures near the melting point, 933.4 K (quoted from Meyer et al. [25])

g	gravitational acceleration (sea level)	$9.81\ \mathrm{m/s^2}$
μ	magnetic permeability (free space)	1.26×10^{-6} H/m
σ	electric conductivity	5.10×10^6 mho/m
ρ	mass density	$2.38\times10^3\ \mathrm{kg/m^3}$
η	dynamic viscosity	2.90×10^{-3} kg/m-s
c	specific heat (at constant pressure)	1.08×10^3 J/kg-K
κ	thermal conductivity	1.03×10^2 W/m-K
β	thermal expansion coefficient	1.16×10^{-4}/K

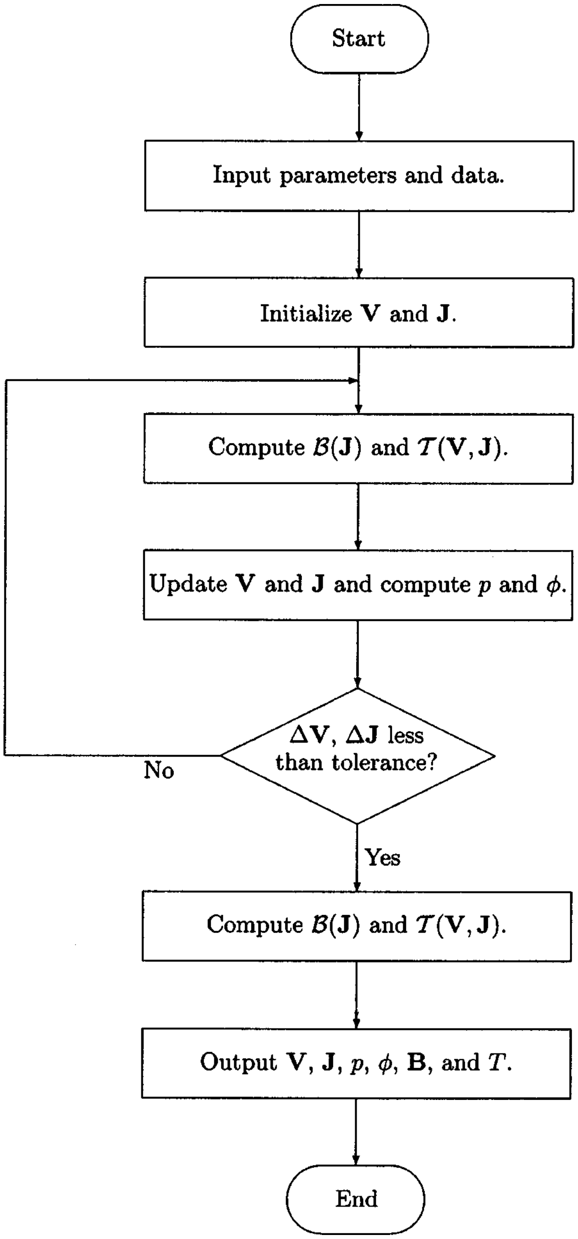


Fig. 1 Simplified flow diagram for the iterative solution of the discretized problem

tion. Figure 1 shows a simplified flow diagram for the iteration scheme. Note that only **V** and **J** need to be initialized.

In order to guarantee stability and convergence of the algorithm, some care must be exercised in choosing finite element spaces. The main restriction is that the elements used to approximate velocity and pressure as well as those used for electric current density and potential must satisfy so-called inf-sup conditions or LBB (Ladyzhenskaya-Babuska-Brezzi) conditions. In the present implementation of the method, we construct a Lipschitz-continuous coordinate transformation that maps the physical domain (a circular cylinder) onto a circumscribed square cylinder (the computational domain), allowing all computations to be performed on a logically rectangular grid. The square cylinder is decomposed into rectangular parallelepipeds of equal size. This allows us to use standard Taylor-Hood elements for velocity and pressure, namely, continuous piecewise triquadratics for the velocity and continuous piecewise trilinears for the pressure. These elements are known to satisfy the LBB condition. It is natural to use continuous piecewise triquadratics for the electric potential as well, but the LBB condition then requires a somewhat nonstandard finite element space for the current density, this space must contain the gradients of continuous piecewise triquadratics. Thus, the elements used to approximate the ith component of the current density are piecewise linear and generally discontinuous in the ith variable, but continuous and piecewise quadratic in the remaining two variables.

Basis functions for the above finite element spaces are constructed using standard 27-node Lagrange elements for the velocity and potential, standard 8-node Lagrange elements for the pressure. For the ith component of the current density, we use Hermite elements with nine nodes, namely, those nodes of the standard 27-node Lagrange element that are not on faces perpendicular to the ith coordinate axis; two degrees-of-freedom are associated with each of these nodes, namely, the value of the function and the value of its derivative with respect to the ith variable. (Instead of 9-node Hermite elements, we could of course use 18-node Lagrange elements, but we would then be unable to utilize the same nodes as for velocity, pressure, and potential.) Since pressure and potential can be determined only up to additive constants, both are set equal to zero at one node each of the finite element grid. The essential boundary condition for the velocity field is enforced by setting $\mathbf{V}=\mathbf{V}_0$ at all boundary nodes.

With these choices, and in light of general results of finite element theory, the discretization error is expected to decrease quadratically with the grid size. For further details, including rigorous error estimates and a numerical validation of the predicted quadratic rate of convergence, we refer to Meir and Schmidt [18].

4 Numerical Experiments and Discussion

The method described in the previous section was implemented and tested in a series of computer experiments, simulating the laboratory conditions described in Bakhtiyarov et al. [23]. Although these experiments are of a preliminary nature, they demonstrate the feasibility of the approach.

While the present implementation allows the computation of the induced magnetic field $\mathcal{B}(\mathbf{J})$, via evaluation of the Biot-Savart integral (11), it does not yet incorporate the effect of temperature fluctuations; that is, $\mathcal{T}(\mathbf{V},\mathbf{J})$ is assumed to be zero. According to the remarks at the end of Section 2, this is viable only as a first approximation. Adding a subroutine for the computation of $\mathcal{T}(\mathbf{V},\mathbf{J})$ poses no problem in principle, but requires laboratory experiments to validate the proposed model (14) for the radiative heat source (including measurements of the intensity distribution in the light cone and infrared absorption properties of the metal samples).

Besides neglecting temperature fluctuations, the computations have thus far been limited to the laminar flow regime with small angular velocities and modest magnetic fields. All computations were done on a workstation, which, in conjunction with the sheer size of this fully three-dimensional problem, precluded the use of all but very coarse finite element grids. The simulation of more realistic flow conditions would require much higher spatial resolution (that is, much finer grids) and ultimately, the incorporation of a turbulence model.

We used the discretization and iteration scheme described in Section 3 on a grid of 432 elements with a total of 4225 nodes. The ensuing sparse linear systems, with roughly 30,000 unknowns (not counting the induced magnetic field), were solved directly, using a standard linear-algebra package. Stiffness matrices and load vectors were computed with a high-order Gaussian quadrature rule on the reference element. The induced magnetic field was determined by evaluating the Biot-Savart integral via Gaussian quadrature.

Given below are the results of a simulation of one of the experiments described in Bakhtiyarov et al. [23], where a cylindrical column of liquid aluminum is steadily rotated in a uniform mag-

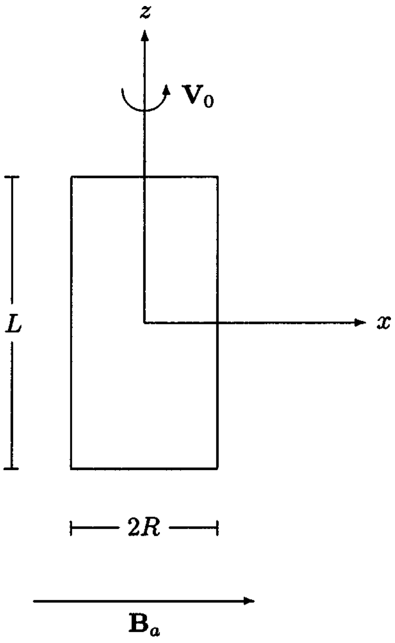


Fig. 2 Geometry of the configuration (not to scale; aspect ratio $L/2R$ is 4)

netic field perpendicular to the cylinder axis; see Fig. 2 for the geometry of the configuration and Table 3 for the data and parameter values used.

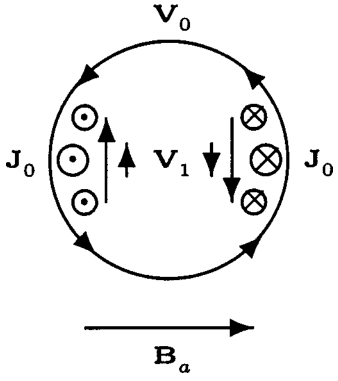


Fig. 3 Induced current and flow field

Table 3 Data and parameter values for computer experiment

R	cylinder radius	1.27×10^{-2} m
L	cylinder height	10.16×10^{-2} m
ω	angular frequency of rotation	1.05/s (10 rpm)
B_0	magnitude of applied magnetic field	0.1 T
ρ	mass density	2.38×10^{3} kg/m^3
η	viscosity	1.8×10^{-3} kg/m-s
σ	electric conductivity	4.1×10^{6} mho/m
μ	magnetic permeability	1.26×10^{-6} H/m

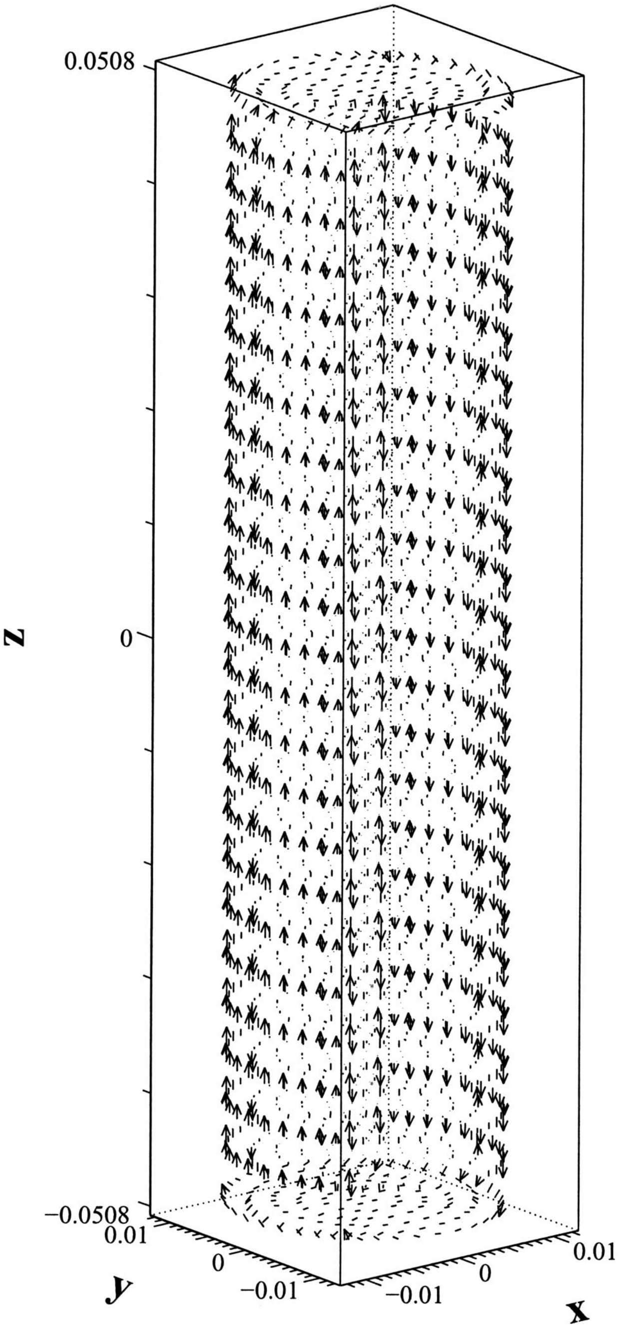


Fig. 4 Current density

Figures 4–7 show the computed current, potential, velocity, and pressure distributions. Clearly visible is a closed current loop parallel to the plane $y=0$ (which contains both the axis of the magnetic field and the axis of rotation). Significant potential gradients arise only near the top and bottom of the column. The velocity field is almost purely horizontal, the pressure almost purely radial. Figure 8, depicting the flow pattern in the plane $z=0$, reveals significant counter-rotation associated with two kidney-shaped vortices, centered on the x-axis (the axis of the magnetic field) and equidistant from the center. The same general pattern is found in horizontal cross sections along much of the cylinder axis. Due to the no-slip boundary condition, a transition to rigid rotation occurs near the top and bottom, but the transition layers are fairly thin (compare Figs. 8 and 9).

More quantitative information can be inferred from azimuthal velocity profiles, parallel or perpendicular to the magnetic field. Figure 10 shows the y-velocity along the x-axis, Fig. 11 the negative of the x-velocity along the y-axis, both in the plane $z=0$. Figures 12 and 13 give essentially the same information, but with respect to the rest frame of the cylinder; that is, they show the azimuthal components of the *induced* velocity $\mathbf{V}-\mathbf{V}_0$ (where $\mathbf{V}_0$ is the velocity field associated with the rigid rotation of the cylinder). The induced velocity is generally antiparallel to $\mathbf{V}_0$ and of

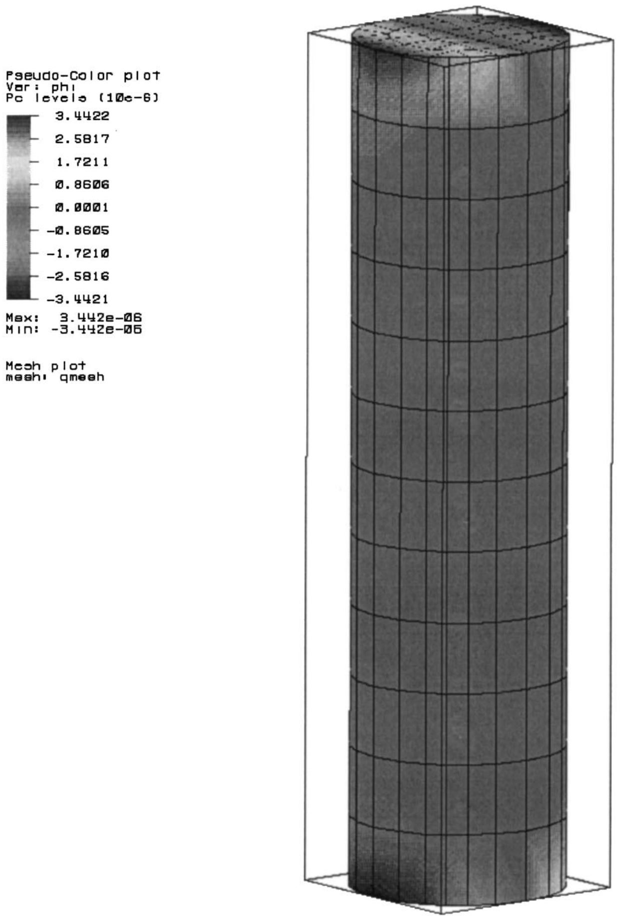


Fig. 5 Electric potential

the same order of magnitude. As a consequence, the fluid is virtually at rest in the region between the two vortices seen in Figs. 8 and 9. Figures 14 and 15 support the observation that the induced velocity does not appreciably decrease along much of the cylinder axis (it must, of course, go to zero in the transition layers at the top and bottom).

The results of the simulation are easily explained, at least qualitatively, by inspection of the leading-order terms in the Navier-Stokes equations and Ohm's law. The velocity field associated with the rigid rotation of the cylinder is $\mathbf{V}_0=\omega(-y\mathbf{i}+x\mathbf{j})$. This is a solution of the Navier-Stokes equations with zero body forces. No viscous shear is associated with $\mathbf{V}_0$, and inertia is balanced by a radial pressure gradient: $\rho(\mathbf{V}_0\cdot\nabla)\mathbf{V}_0=-\nabla p_0$ with $p_0=1/2\rho\omega^2(x^2+y^2)$. Due to the presence of the applied magnetic field $\mathbf{B}_a=B_0\mathbf{i}$, a current $\mathbf{J}_0=\sigma\mathbf{V}_0\times\mathbf{B}_a=-\sigma\omega B_0 x\mathbf{k}$ is induced. This current is parallel to the *z*-axis and concentrated in two near-wall regions centered on the *x*-axis (see Fig. 3); it is not accompanied by a potential gradient (since $\mathbf{V}_0\times\mathbf{B}_a$ is solenoidal). The finite length of the cylinder forces return currents to flow parallel to the *x*-axis in boundary layers (Hartmann layers) near the top and bottom of the cylinder; those are associated with potential gradients.

The current $\mathbf{J}_0$ and applied field $\mathbf{B}_a$ generate a Lorentz force $\mathbf{J}_0\times\mathbf{B}_a=-\sigma\omega B_0^2 x\mathbf{j}$. This force is rotational and cannot be balanced by a pressure gradient; it thus accelerates the fluid, resulting in a secondary velocity $\mathbf{V}_1$, antiparallel to $\mathbf{V}_0$ (see Fig. 3). This explains the general flow pattern (and, in particular, the kidney-shaped vortices) seen in Figs. 8 and 9. According to the remarks at the end of Section 2, it must be the inertial force, $\rho(\mathbf{V}_1\cdot\nabla)\mathbf{V}_1$, which balances $\mathbf{J}_0\times\mathbf{B}_a$. A characteristic magnitude of $\mathbf{V}_1$ is thus given by $V_1=(\sigma\omega/\rho)^{1/2}RB_0$, and the ratio $V_1/V_0=(\sigma/\rho\omega)^{1/2}B_0$ is roughly of order one. (This scaling argument applies as long as the hydrodynamic Reynolds number is large compared to unity, which is the case even for fairly small *ω*, but of course not in the limit *ω*→0.)

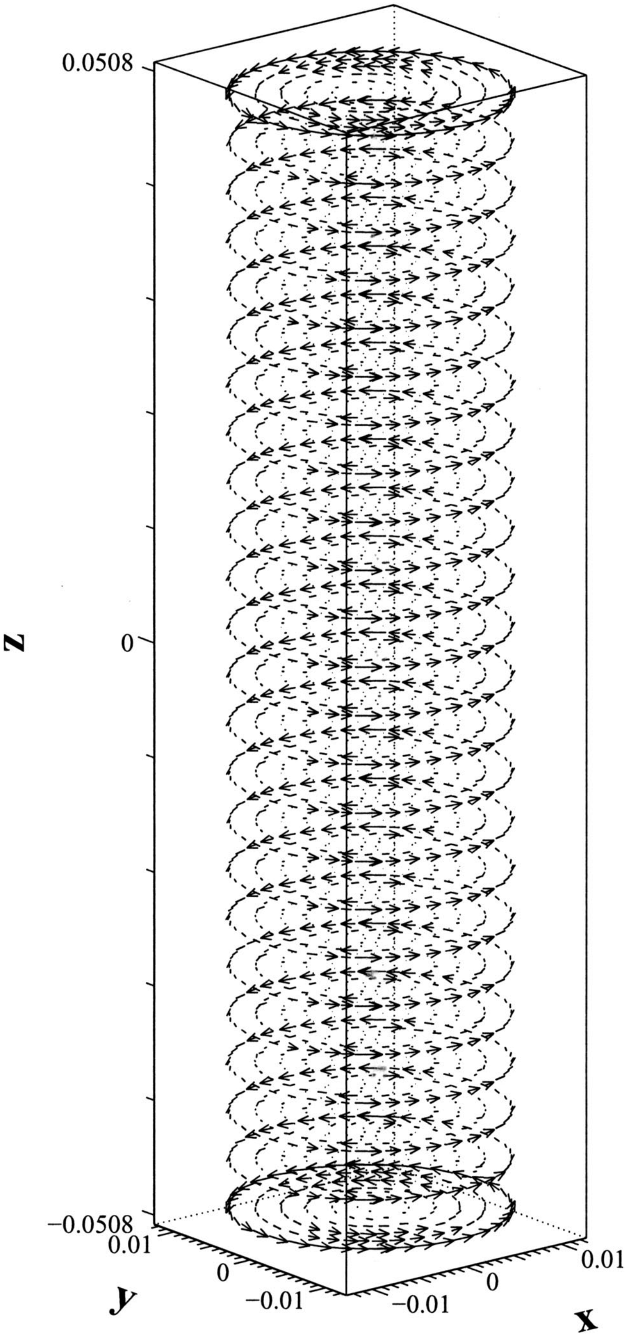


Fig. 6 Velocity field

The main observation is that already moderate magnetic fields and angular velocities result in significant counter-rotation of the melt in the two near-wall regions, centered on the axis of the applied field, where the current $\mathbf{J}_0$ is concentrated.

5 Conclusions

A new aproach to the mathematical modeling and computational simulation of fully three-dimensional, electromagnetically and thermally driven liquid-metal flow was developed and applied

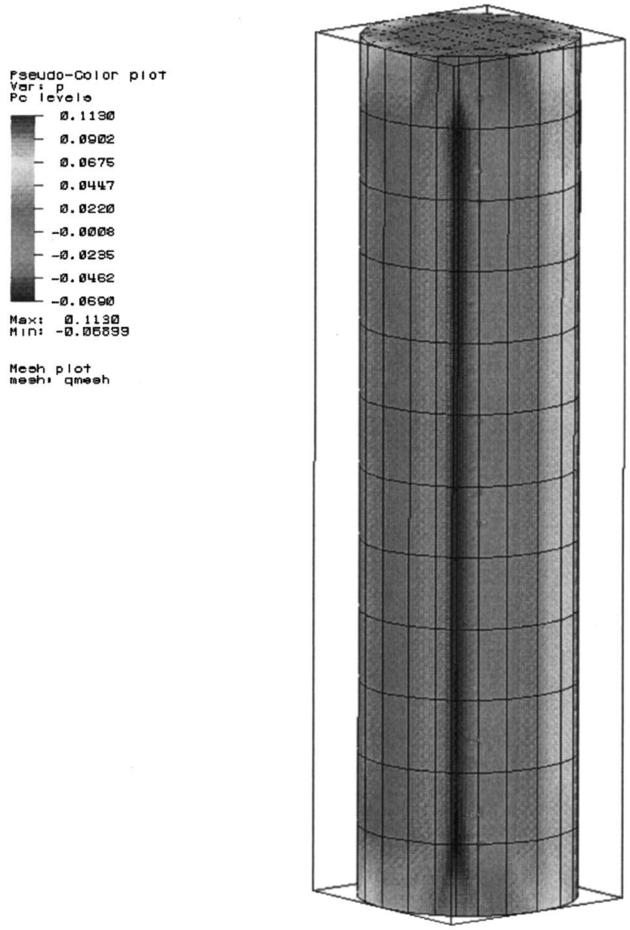


Fig. 7 Pressure distribution

to simulate the flow of a metallic melt in a cylindrical crucible, rotating steadily in a uniform magnetic field perpendicular to the cylinder axis. A finite element based discretization and iteration scheme was designed for the numerical solution of the underlying

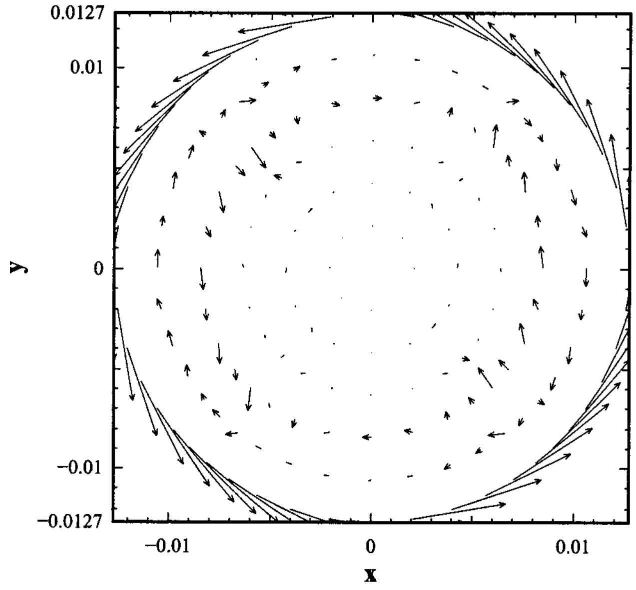


Fig. 8 Flow pattern in the plane $z=0$

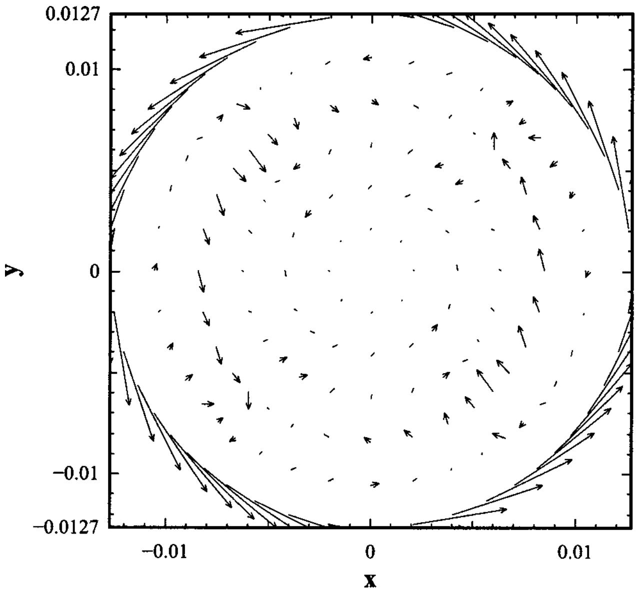


Fig. 9 Flow pattern in the plane $z=5L/12$

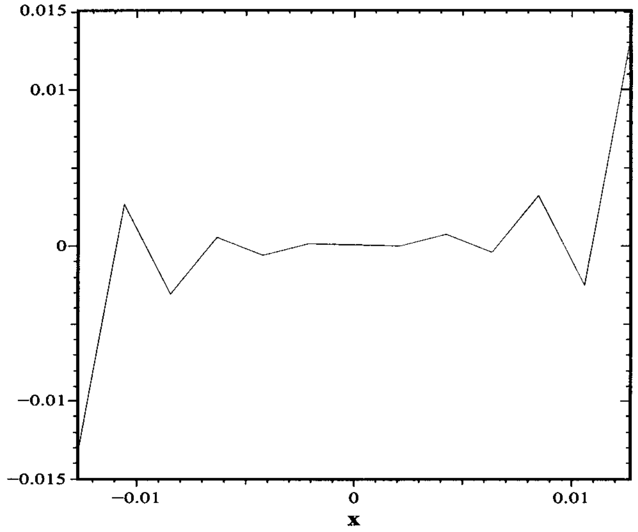


Fig. 10 Azimuthal velocity in the plane $z=0$, along the x-axis (parallel to the magnetic field)

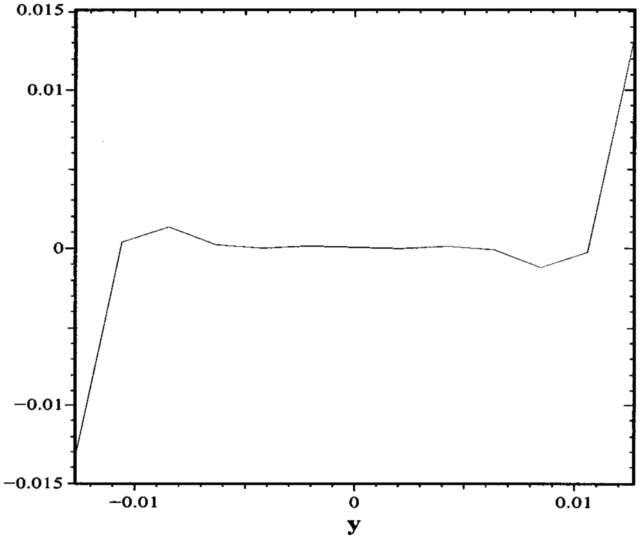


Fig. 11 Azimuthal velocity in the plane $z=0$, along the y-axis (perpendicular to the magnetic field)

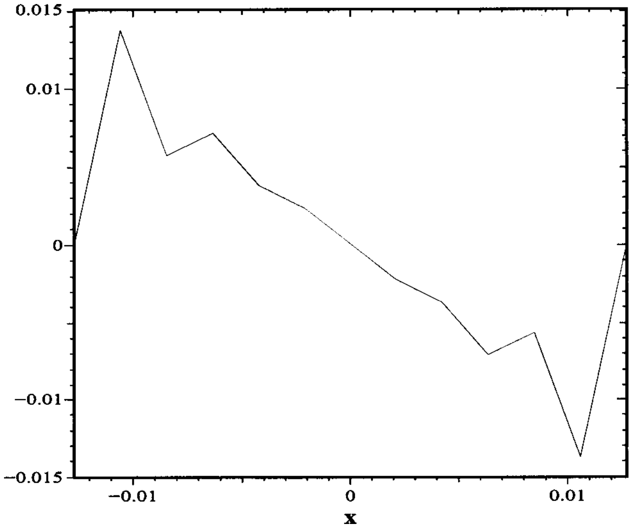


Fig. 12 Induced azimuthal velocity $V-V_0$ in the plane $z=0$, along the x-axis

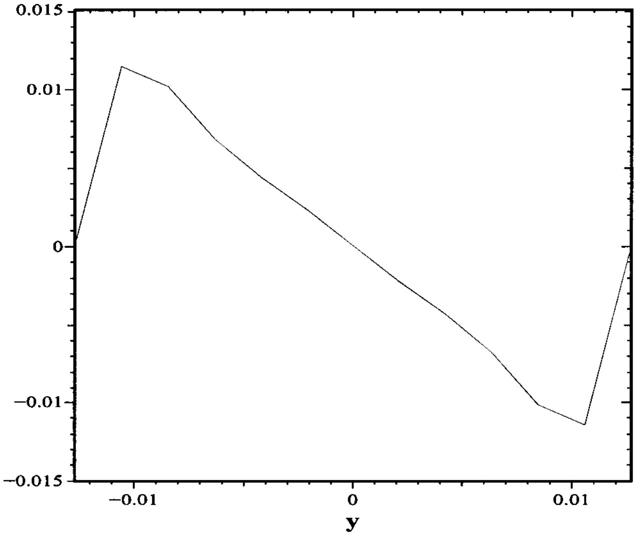


Fig. 13 Induced azimuthal velocity $V-V_0$ in the plane $z=0$, along the y-axis

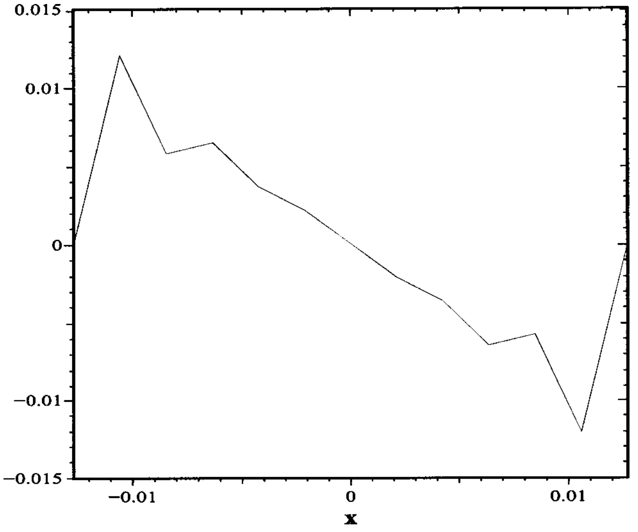


Fig. 14 Induced azimuthal velocity $V-V_0$ in the plane $z=5L/12$, along the x-axis

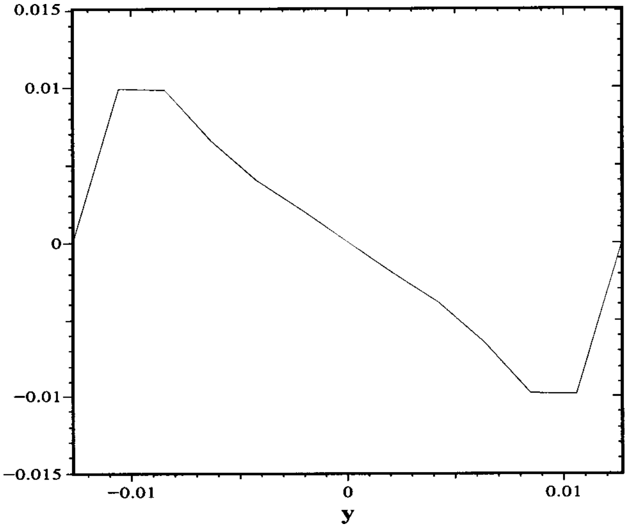


Fig. 15 Induced azimuthal velocity $V-V_0$ in the plane $z=5L/12$, along the y-axis

nonlinear PDEs. The results of preliminary computer experiments (limited to the laminar flow regime and not accounting for temperature fluctuations) were shown to agree with theoretical predictions. It was found that already modest magnetic fields and angular velocities lead to significant counter-rotation in the melt.

Acknowledgment

The work of Meir and Schmidt was supported by the National Science Foundation under Grant No. DMS-9625096.

References

[1] Prescott, P. J., and Incropera, F. P., 1993, "Magnetically Damped Convection During Solidification of a Binary Metal Alloy," ASME J. Heat Transfer, **115**, pp. 302–310.
[2] Walker, J. S., Henry, D., and Ben Hadid, H., 2002, "Magnetic Stabilization of the Buoyant Convection in the Liquid-Encapsulated Czochralski Process," J. Cryst. Growth, **243**, pp. 108–116.
[3] Hughes, M., Pericleous, K. A., and Cross, M., 1995, "The Numerical Modeling of DC Electromagnetic Pump and Brake Flow," Appl. Math. Model., **19**, pp. 713–723.
[4] Verardi, S. L. L., Cardoso, J. R., and Costa, M. C., 2001, "Three-Dimensional Finite Element Analysis of MHD Duct Flow by the Penalty Function Formulation," IEEE Trans. Magn., **37**, pp. 3384–3387.
[5] Vivès, Ch., and Ricou, R., 1985, "Fluid Flow Phenomena in a Single Phase Coreless Induction Furnace," Metall. Trans. B, **16**, pp. 227–235.
[6] Spitzer, K.-H., Dubke, M., and Schwerdtfeger, K., 1986, "Rotational Electromagnetic Stirring in Continuous Casting of Round Strands," Metall. Trans. B, **17**, pp. 119–131.
[7] Barz, R. U., Gerbeth, G., Wunderwald, U., Buhrig, E., and Gelfgat, Yu. M., 1997, "Modelling of the Isothermal Melt Flow Due to Rotating Magnetic Fields in Crystal Growth," J. Cryst. Growth, **180**, pp. 410–421.
[8] Ben Hadid, H., Henry, D., and Kaddeche, S., 1997, "Numerical Study of Convection in the Horizontal Bridgman Configuration Under the Action of a Constant Magnetic Field. Part 1. Two-Dimensional Flow," J. Fluid Mech., **333**, pp. 23–56.
[9] Ben Hadid, H., and Henry, D., 1997, "Numerical Study of Convection in the Horizontal Bridgman Configuration Under the Action of a Constant Magnetic Field. Part 2. Three-Dimensional Flow," J. Fluid Mech., **333**, pp. 57–83.
[10] Witkowski, L. M., Walker, J. S., and Marty, P., 1999, "Nonaxisymmetric Flow in a Finite-Length Cylinder With a Rotating Magnetic Field," Phys. Fluids, **11**, pp. 1821–1826.
[11] Witkowski, L. M., and Walker, J. S., 2002, "Numerical Solutions for the Liquid-Metal Flow in a Rotating Cylinder With a Weak Transverse Magnetic Field," Fluid Dyn. Res., **30**, pp. 127–137.
[12] Moffatt, H. K., 1965, "On Fluid Flow Induced by a Rotating Magnetic Field," J. Fluid Mech., **22**, pp. 521–528.
[13] Davidson, P. A., and Hunt, J. C. R., 1987, "Swirling Recirculating Flow in a Liquid-Metal Column Generated by a Rotating Magnetic Field," J. Fluid Mech., **185**, pp. 67–106.
[14] Fujisaki, K., 2001, "In-Mold Electromagnetic Stirring in Continuous Casting," IEEE Trans. Ind. Appl., **37**, pp. 1098–1104.
[15] Natarajan, T. T., and El-Kaddah, N., 2004, "Finite Element Analysis of Elec-

tromagnetic and Fluid Flow Phenomena in Rotary Electromagnetic Stirring of Steel,'' Appl. Math. Model., **28**, pp. 47–61.

[16] Natarajan, T. T., and El-Kaddah, N., 1999, ''A Methodology for Two-Dimensional Finite Element Analysis of Electromagnetically Driven Flow in Induction Stirring Systems,'' IEEE Trans. Magn., **35**, pp. 1773–1776.

[17] Natarajan, T. T., and El-Kaddah, N., 2002, ''A New Method for Three-Dimensional Numerical Simulation of Electromagnetic and Fluid-Flow Phenomena in Electromagnetic Separation of Inclusions From Liquid Metal,'' Metall. Mater. Trans. B, **33**, pp. 775–785.

[18] Meir, A. J., and Schmidt, P. G., 1999, ''Analysis and Numerical Approximation of a Stationary MHD Flow Problem With Nonideal Boundary,'' SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal., **36**, pp. 1304–1332.

[19] Meir, A. J., and Schmidt, P. G., 2001, ''On Electrically and Thermally Driven Liquid-Metal Flows,'' Nonlinear Anal. Theory, Methods Appl., **47**, pp. 3281–3294.

[20] Meir, A. J., and Schmidt, P. G., 1998, ''Analysis and Finite Element Simulation of MHD Flows, With an Application to Seawater Drag Reduction,'' *Proc. International Symposium on Seawater Drag Reduction*, J. C. S. Meng, ed., Naval Undersea Warfare Center, Newport, RI, pp. 401–406.

[21] Meir, A. J., and Schmidt, P. G., 1999, ''Analysis and Finite Element Simulation of MHD Flows, With an Application to Liquid Metal Processing,'' *Proc. Fluid Flow Phenomena in Metals Processing*, 1999 TMS Annual Meeting, N. El-Kaddah et al., eds., Minerals, Metals, and Materials Society, Warrendale, PA, pp. 561–569.

[22] Schmidt, P. G., 1999, ''A Galerkin Method for Time-Dependent MHD Flow With Nonideal Boundaries,'' Commun. Appl. Anal., **3**, pp. 383–398.

[23] Bakhtiyarov, S. I., Overfelt, R. A., Meir, A. J., and Schmidt, P. G., 2003, ''Experimental Measurements of Velocity, Potential, and Temperature Distributions in Molten Metals During Electromagnetic Stirring,'' ASME J. Appl. Mech., **70**, pp. 351–358.

[24] Hughes, W. F., and Young, F. J., 1966, *The Electromagnetodynamics of Fluids*, Wiley, New York.

[25] Meyer, J.-L., Durand, F., Ricou, R., and Vivès, C., 1984, ''Steady Flow of Liquid Aluminum in a Rectangular-Vertical Ingot Mold, Thermally or Electromagnetically Activated,'' Metall. Trans. B, **15**, pp. 471–478.

[26] Brezzi, F., and Fortin, M., 1991, *Mixed and Hybrid Finite Element Methods*, Springer, New York.

[27] Girault, V., and Raviart, P.-A., 1986, *Finite Element Approximation of the Navier-Stokes Equations, Theory and Algorithms*, Springer, New York.

Ganesh Thiagarajan[1]
e-mail: ganesht@umkc.edu
Department of Civil Engineering,
University of Missouri,
Kansas City, MO 64110

Yonggang Y. Huang
Department of Mechanical and Industrial
Engineering,
University of Illinois—Urbana-Champaign,
Urbana, IL 61801

K. Jimmy Hsia
Department of Theoretical and Applied
Mechanics,
University of Illinois—Urbana-Champaign,
Urbana, IL 61801

Fracture Simulation Using an Elasto-Viscoplastic Virtual Internal Bond Model With Finite Elements

A virtual internal bond (VIB) model for isotropic materials has been recently proposed by Gao (Gao, H., 1997, "Elastic Waves in a Hyperelastic Solid Near its Plane Strain Equibiaxial Cohesive Limit," Philos. Mag. Lett. ***76****, pp. 307–314) and Gao and Klein (Gao, H., and Klein, P., 1998, "Numerical Simulation of Crack Growth in an Isotropic Solid With Randomized Internal Cohesive Bonds," J. Mech. Phys. Solids* ***46****(2), pp. 187–218), in order to describe material deformation and fracture under both static and dynamic loading situations. This is made possible by incorporating a cohesive type law of interaction among particles at the atomistic level into a hyperelastic framework at the continuum level. The finite element implementation of the hyperelastic VIB model in an explicit integration framework has also been successfully described in an earlier work by the authors. This paper extends the isotropic hyperelastic VIB model to ductile materials by incorporating rate effects and hardening behavior of the material into a finite deformation framework. The hyperelastic VIB model is formulated in the intermediate configuration of the multiplicative decomposition of the deformation gradient framework. The results pertaining to the deformation, stress-strain behavior, loading rate effects, and the material hardening behavior are studied for a plate with a hole problem. Comparisons are also made with the corresponding hyperelastic VIB model behavior.*
[DOI: 10.1115/1.1796451]

1 Introduction

It is well established that plastic deformation takes place in regions of high stress concentrations and that fracture at the crack tip is preceded by some degree of plastic deformation. To be more realistic the numerical simulations of crack initiation, propagation, and branching must, therefore, include the elastoplastic behavior of the material. Among the finite-element-based numerical models, which are becoming widely accepted, is the cohesive surface modeling of fracture. Many researchers, including Barenblatt [1], Dugdale [2], Willis [3], Xia and Shih [4], and Xu and Needleman [5], to mention a few, have worked on propagating and advancing this methodology. The fundamental basis of these models lies in defining discrete cohesive surfaces in which the traction and separation at the boundaries are described by nonlinear cohesive laws. These models do not require any separate fracture criterion. However, these surfaces, which lie in between element boundaries, must be defined a priori, and separate cohesive elements must be introduced in between boundaries of the regular finite elements.

In contrast to the approach described above, Gao [6,7] and Gao and Klein [8] proposed an approach called the Virtual Internal Bond (VIB) model, wherein the *constitutive model* directly *incorporates a cohesive-type law*. In the VIB approach, the continuum is treated as a random network of material points, interconnected by bonds, which obeys a cohesive law. The bonds are physically described by a bond energy $U(l)$, where l is the bond length, and its derivative with respect to the bond length $U'(l)$ is the cohesive bond force. By describing this bond energy in a hyperelastic framework of finite deformation, the appropriate stress and strain measures, such as the Green-Lagrange strain tensor and the second Piola-Kirchoff stress tensor, can be derived. The macroscopic description of the continuum is determined by the Cauchy-Born rule [9] of crystal elasticity, by equating the macroscopic strain energy function at the continuum level to the potential energy stored in the cohesive bonds at the microscale. The description of the bond length l in terms of the Green-Lagrange strain tensor provides the link between the two scales.

Klein and Gao [10] have described the application of the VIB model (based on the hyperelastic framework) to fracture initiation and propagation and have further studied the crack dynamics using this model. An implicit second-order integration scheme was used to simulate quasi-static and dynamic loading problems. Zhang et al. [11] implemented this model, using an implicit integration scheme, in a UMAT subroutine in ABAQUS [12]. The softening region of the cohesive models presents a major issue with the numerical implementation, using implicit integration schemes. Thiagarajan et al. [13] found that the explicit integration scheme is better suited for the finite element (FE) implementation of the VIB model. The model was implemented using the user subroutine VUMAT in ABAQUS. The influence of mesh shape and size, loading rate, and other related issues were studied for both quasi-static and dynamic impact loading cases [13]. The experimental verification and validation of the VIB model was studied for the case of dynamic fragmentation of brittle materials under impact loading by Thiagarajan et al. [14].

The work presented here stems from the previous work by the authors and is motivated by the observation that it is necessary to incorporate plastic deformation in regions of high stress concentrations. There are two basic approaches of incorporating plasticity into the VIB model. One is at the atomistic level by the consideration of individual dislocations (millions of them) and other microstructural features. This method is numerically daunting

---

[1] Corresponding author.

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, January 21, 2003; final revision, November 13, 2003. Editor: H. Gao. 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.

even with the current level of supercomputers. The second approach, which is adopted in this research, is to incorporate plasticity and/or viscoplasticity at the continuum level by using well-established plasticity and viscoplasticity models.

The incorporation of viscoplastic effects into the VIB model is done within the framework of the multiplicative decomposition of the deformation gradient $\mathbf{F}=\mathbf{F}^e\mathbf{F}^p$ proposed by Lee [15] and advanced by numerous authors. This description is outlined in a later section. The fracture simulation of a ductile material is studied by treating it as an elasto-viscoplastic solid. The elastic behavior is modeled as a hyperelastic VIB material and the plastic and/or viscoplastic response is defined in the so-called intermediate configuration. The model is studied using 6061 Al for which the viscoplastic material properties are available.

In this paper, Section 2 gives a brief description of the well-known finite deformation kinematic formulations in order to establish the scientific basis for the choice of the intermediate configuration. The elasto-viscoplastic constitutive formulation for the intermediate configuration is then described in Section 3. Section 4 describes the details of the explicit integration scheme details, and Section 5 describes the numerical finite element implementation, along with the results from various case studies. The conclusions are presented in Section 6.

2 General Kinematic Formulations

The choice of an appropriate strain and strain-rate measure is critical to the elasto-viscoplastic formulation of fracture for ductile materials using the VIB model. At the outset, various kinematic frameworks, both small strain and finite strain, are described below along with the possible situations to which they may be applied. The choice of the kinematic formulation that is best suited for the elastoplastic VIB model is then described in detail.

2.1 Small Strain Formulation. For problems where total strains are small, the additive decomposition of total strains (of the order of 0.001) are given as

$$\epsilon_{ij}=\epsilon_{ij}^{e}+\epsilon_{ij}^{p} \tag{1}$$

where the elastic strains are of the order of 0.001 (0.1%).

The uniaxial strains for ductile materials can reach the order of unity due to plastic flow, especially in metal-forming problems. In these cases the elastic strains (0.001) are very small compared to the total strains and can be neglected. The hydrostatic stress in such cases does not affect yielding, and only the deviatoric part of the stress causes yielding. The plastic flow can be expressed as a functional as shown below:

$$\dot{\epsilon}_{ij}^{p}=f_{ij}(\sigma,\dot{\sigma}) \tag{2}$$

As the elastic strains are neglected, the total strain rate can be expressed in terms of the derivatives of the velocity field as follows:

$$\dot{\epsilon}_{ij}=\dot{\epsilon}_{ij}^{p} \tag{3}$$

$$=\frac{1}{2}\left(\frac{\partial v_i}{\partial x_j}+\frac{\partial v_j}{\partial x_i}\right) \tag{4}$$

Final deformations are obtained by integrating Eq. (4) over time. This formulation is well known as the rigid plastic model and, although the total strains are finite, there is no need to use finite strain kinematics.

2.2 Finite Deformation Theory. The finite deformation framework is applicable to cases where the total strains are finite and the magnitude of elastic and plastic strains are comparable. Examples of these include cases where the inertia forces are high due to rapid loading as in explosive and impact loading. Elastic strains of the magnitude of 25% or more have been reported for explosive loading [16], which would necessitate a kinematic framework where both elastic and plastic strains can be of finite magnitude. The kinematics involved in the formulation of finite elastic and inelastic deformations are based on the multiplicative decomposition of the deformation gradient and described in detail below.

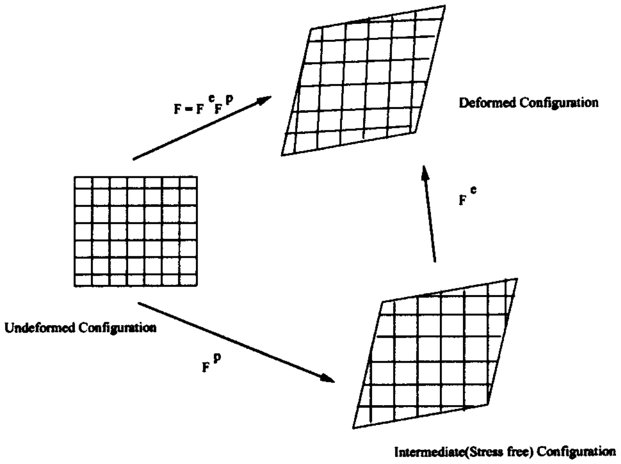


Fig. 1 Multiplicative decomposition of deformation gradient F with its effect on shape and microstructure

2.3 Multiplicative Decomposition of Deformation Gradient. The *effective decoupling* of the elastic and plastic laws can be achieved by using the unstressed configuration for deformation changes. It is assumed that yielding is not affected by the hydrostatic Cauchy stresses and that the elastic constants are not affected by plastic deformations. In this paper, capital-letter subscripts are used for the initial configuration while lowercase subscripts are used for the deformed configuration.

A multiplicative framework for large deformation inelastic behavior is adopted [15–21] as follows:

$$\mathbf{F}=\mathbf{F}^e\cdot\mathbf{F}^p \qquad \det\mathbf{F}^e,\det\mathbf{F}^p>0$$
$$F_{iJ}=F_{iI}^{e}F_{IJ}^{p} \tag{5}$$

where $\mathbf{F}$ is the deformation gradient, $\mathbf{F}^e$ is the elastic deformation gradient and $\mathbf{F}^p$ is the plastic deformation gradient. As shown in Fig. 1, such a characterization allows for the definition of an imaginary intermediate relaxed (stress-free) configuration $\bar{\Omega}$. Ω_0 and Ω are the configurations at time $t=t_0$ and $t=t_{n+1}$, respectively. In this paper any quantity with an overbar $\overline{(\ )}$ refers to its value in the intermediate configuration. The intermediate configuration is commonly considered as a stress-free configuration obtained by elastically unloading the body from the current configuration. It can be physically considered to represent the total effect of dislocations without any lattice distortions (mapped by $\mathbf{F}^p$), while the lattice distortions and rotations transform the intermediate configuration to the current configuration (mapped by $\mathbf{F}^e$). The basic kinematic variables associated with the three configurations are now described.

2.3.1 Undeformed Configuration Ω0. The primary kinematic tensors associated with the undeformed configuration Ω_0 are the left Cauchy-Green deformation tensor $\mathbf{C}$ and the Green-Lagrange strain tensor $\mathbf{E}$, which are expressed as follows:

$$\mathbf{C}=\mathbf{F}^T\cdot\mathbf{F} \quad \text{or} \quad C_{IJ}=F_{iI}F_{iJ} \tag{6}$$

$$\mathbf{E}=\tfrac{1}{2}(\mathbf{C}-\mathbf{G}) \quad \text{or} \quad E_{IJ}=\tfrac{1}{2}(C_{IJ}-G_{IJ}) \tag{7}$$

where $\mathbf{G}$ is the metric tensor in the undeformed configuration. In the Cartesian coordinate system $G_{IJ}=\delta_{IJ}$ where δ_{IJ} is the Kronecker delta. Since the plastic deformation gradient maps the initial configuration to the intermediate configuration, the plastic parts of the left Cauchy-Green deformation tensor $\mathbf{C}^p$ and the Green-Lagrange strain tensor $\mathbf{E}^p$ can be expressed as follows:

$$\mathbf{C}^p=\mathbf{F}^{pT}\cdot\mathbf{F}^p \quad \text{or} \quad C^p_{IJ}=F^p_{iI}F^p_{iJ} \tag{8}$$

$$\mathbf{E}^p=\tfrac{1}{2}(\mathbf{C}^p-\mathbf{G}) \quad \text{or} \quad E^p_{IJ}=\tfrac{1}{2}(C^p_{IJ}-G_{IJ}) \tag{9}$$

From the above definitions, the elastic component of the Green-Lagrange strain tensor is defined using the additive decomposition [22] as

$$\mathbf{E}^e=\mathbf{E}-\mathbf{E}^p \tag{10}$$

2.3.2 Intermediate Configuration $\overline{\Omega}$. __The Green-Lagrange strain in the intermediate configuration $\mathbf{E}$ can be defined as a push-forward transformation of $\mathbf{E}$ using $\mathbf{F}^p$ as follows:

$$\overline{\mathbf{E}}=\mathbf{F}^{p^{-T}}\mathbf{E}\mathbf{F}^{p^{-1}} \tag{11}$$

$$=\overline{\mathbf{E}}^e+\overline{\mathbf{E}}^p \tag{12}$$

where

$$\overline{\mathbf{E}}^e=\tfrac{1}{2}(\mathbf{F}^{e^T}\cdot\mathbf{F}^e-\mathbf{I}) \tag{13}$$

$$\overline{\mathbf{E}}^p=\tfrac{1}{2}(\mathbf{I}-\mathbf{F}^{p^{-T}}\cdot\mathbf{F}^{p^{-1}}) \tag{14}$$

The deformation gradients $\mathbf{F}^e$, $\mathbf{F}^p$ may not strictly be continuously differentiable mappings. They have sometimes been described as point matrix functions.

2.3.3 Current or Deformed Configuration Ω. The deformation gradients mapping the initial and the intermediate configurations into the current configuration (Fig. 1) are the total $\mathbf{F}$ and elastic $\mathbf{F}^e$ parts, respectively. Consequently, the primary strain tensors associated with the current configuration may be expressed as

$$\mathbf{b}=\mathbf{F}\cdot\mathbf{F}^T \quad \text{or} \quad b_{ij}=F_{iI}F_{jI} \tag{15}$$

$$\mathbf{e}=\tfrac{1}{2}(\mathbf{g}-\mathbf{b}^{-1}) \quad \text{or} \quad e_{ij}=\tfrac{1}{2}(g_{ij}-b^{-1}_{ij}) \tag{16}$$

$$\mathbf{b}^e=\mathbf{F}^e\cdot\mathbf{F}^{eT} \quad \text{or} \quad b^e_{ij}=F^e_{iI}F^e_{jI} \tag{17}$$

$$\mathbf{e}^e=\tfrac{1}{2}(\mathbf{g}-\mathbf{b}^{e-1}) \quad \text{or} \quad e^e_{ij}=\tfrac{1}{2}(g_{ij}-b^{e-1}_{ij}) \tag{18}$$

where $\mathbf{b}$ and $\mathbf{b}^e$ are the total and elastic right Cauchy-Green deformation tensors and $\mathbf{e}$ and $\mathbf{e}^e$ are the total and elastic Eulerian strain tensors, respectively. The additive decomposition of the Eulerian strain tensor leads to

$$\mathbf{e}^p=\mathbf{e}-\mathbf{e}^e \tag{19}$$

Following Marsden and Hughes [23], the tensors $\mathbf{e}$ and $\mathbf{e}^p$ may be referred to as the push forward, using $\mathbf{F}^p$, of the tensors $\mathbf{E}$ and $\mathbf{E}^p$, respectively.

2.4 Rate of Deformation and Spin Tensors. The velocity gradient, rate of deformation, and spin tensors can be expressed in the current and intermediate configurations as described below.

2.4.1 Current Configuration. The velocity gradient of a particle defined in the current configuration is expressed as

$$\mathbf{L}=\frac{\partial\mathbf{v}}{\partial\mathbf{x}}=\dot{\mathbf{F}}\mathbf{F}^{-1}=\mathbf{D}+\mathbf{W} \tag{20}$$

$$\mathbf{D}=symm(\mathbf{L}) \tag{21}$$

$$\mathbf{W}=skew(\mathbf{L}) \tag{22}$$

where $\mathbf{D}$ is the rate of deformation or velocity tensor representing the stretching part and $\mathbf{W}$ is the spin rate tensor. By substituting the multiplicative decomposition of the deformation gradient in the above equation, one can derive the following expression:

$$\mathbf{L}=(\dot{\mathbf{F}}^e\cdot\mathbf{F}^{e^{-1}})+(\mathbf{F}^e\cdot\dot{\mathbf{F}}^p\cdot\mathbf{F}^{p^{-1}}\cdot\mathbf{F}^{e^{-1}})=(\mathbf{L}^e)+(\mathbf{L}^p) \tag{23}$$

$$=(\mathbf{D}^e+\mathbf{W}^e)+(\mathbf{D}^p+\mathbf{W}^p) \tag{24}$$

where $\mathbf{L}^e$ and $\mathbf{L}^p$ are the definitions of the elastic and plastic velocity gradient tensors and $\mathbf{D}^e$, $\mathbf{D}^p$, $\mathbf{W}^e$, $\mathbf{W}^p$ are the elastic and plastic parts of the stretching-rate and spin-rate tensors, respectively. In theory it can be assumed that not only the symmetric part $\mathbf{D}^P$ but also the skew symmetric part $\mathbf{W}^p$ is governed by constitutive equations. For rate-dependent materials [24], these constitutive equations can be expressed as

$$\mathbf{D}^P=\mathbf{F}^e\mathbf{D}^p(F^{eT}F^e,\alpha)\mathbf{F}^{e^{-1}} \tag{25}$$

$$\mathbf{W}^P=\mathbf{F}^e\mathbf{W}^p(F^{eT}F^e,\alpha)\mathbf{F}^{e^{-1}} \tag{26}$$

where α is a set of parameters that describe structural changes caused by lattice defects. Wang's [25] representation theorem yields $\mathbf{W}^p=0$, thereby ignoring the effect of the plastic spin-rate tensor.

2.4.2 Intermediate Configuration. Since the velocity gradient has a covariant-contravariant character, using an appropriate pull back from the current configuration, the expression for its representation in the intermediate configuration results in

$$\overline{\mathbf{L}}=\mathbf{F}^{e^{-1}}\mathbf{L}\mathbf{F}^e \tag{27}$$

$$=\mathbf{F}^{e^{-1}}\dot{\mathbf{F}}^e+\dot{\mathbf{F}}^p\mathbf{F}^{p^{-1}}=\overline{\mathbf{L}}^e+\mathbf{L}^p \tag{28}$$

where $\overline{\mathbf{L}}^e$ and $\overline{\mathbf{L}}^p$ are the elastic and plastic velocity gradient tensors. The rate of deformation tensors in the intermediate configuration can be expressed as

$$\overline{\mathbf{D}}=symm(\overline{L})=\tfrac{1}{2}(\overline{C}^e\overline{L}+\overline{L}^T\overline{C}^e)=\mathbf{F}^{eT}\mathbf{D}\mathbf{F}^e \tag{29}$$

$$\overline{\mathbf{D}}^e=symm(\overline{L}^e)=\tfrac{1}{2}(\overline{C}^e\overline{L}^e+\overline{L}^{eT}\overline{C}^e)=\mathbf{F}^{eT}\mathbf{D}^e\mathbf{F}^e \tag{30}$$

$$\overline{\mathbf{D}}^p=symm(\overline{L}^p)=\tfrac{1}{2}(\overline{C}^e\overline{L}^p+\overline{L}^{pT}\overline{C}^e)=\mathbf{F}^{eT}\mathbf{D}^p\mathbf{F}^e \tag{31}$$

3 Elastic-Viscoplastic Constitutive Assumptions

In the approach proposed in this paper, the hyperelastic-viscoplastic response of the VIB model is based on the following assumptions:

1. The intermediate configuration is assumed to be the basis for the definition of the plastic variables. As the intermediate configuration is considered to be a fixed configuration associated only with lattice deformations, the elastic bonds between the particles are not affected. This configuration is obtained by elastically unloading the material from the current configuration. The original hyperelastic VIB model is hence set up in the intermediate configuration.
2. The *flow rule* is specified in terms of the plastic part of the velocity gradient $\overline{\mathbf{L}}^p$. This is assumed to be a function of the Piola-Kirchoff stress and other internal variables.
3. The solid is considered to be isotropic in the intermediate configuration leading to the assumption that the plastic part of the spin tensor is zero $(\overline{\mathbf{W}}^p=0)$.
4. A unified viscoplastic model is assumed, where inelastic strains include plastic and creep strains, by specifying a single set of flow rules and evolutionary equations.
5. The stress rate is assumed to have no effect on the evolutionary equations, i.e., instantaneous plasticity is neglected.
6. Plastic deformations are assumed to be incompressible or $\det(\mathbf{F}^p)=1$.

3.1 Intermediate Configuration Formulation. Based on the hypothesis of formulating the hyperelastic part of the elasto-viscoplastic VIB formulation in the intermediate configuration, the final kinematic quantities used in this formulation are summarized as follows:

$$\mathbf{F}=\mathbf{F}^e\cdot\mathbf{F}^p$$

$$\mathbf{L}=\dot{\mathbf{F}}\mathbf{F}^{-1}=\mathbf{D}+\mathbf{W}$$

$$\mathbf{L}=(\dot{\mathbf{F}}^e\cdot\mathbf{F}^{e^{-1}})+(\mathbf{F}^e\cdot\dot{\mathbf{F}}^p\cdot\mathbf{F}^{p^{-1}}\cdot\mathbf{F}^{e^{-1}})=(\mathbf{L}^e)+(\mathbf{F}^e\overline{\mathbf{L}}^p\mathbf{F}^{e^{-1}})$$
$$=\mathbf{L}^e+\mathbf{L}^p=(\mathbf{D}^e+\mathbf{W}^e)+(\mathbf{D}^p+\mathbf{W}^p)$$

$$\overline{\mathbf{L}}=\mathbf{F}^{e^{-1}}\mathbf{L}\mathbf{F}^e=\mathbf{F}^{e^{-1}}\dot{\mathbf{F}}^e+\dot{\mathbf{F}}^p\mathbf{F}^{p^{-1}}=\overline{\mathbf{L}}^e+\overline{\mathbf{L}}^p$$

$$\overline{\mathbf{L}}^e=\mathbf{F}^{e^{-1}}\cdot\dot{\mathbf{F}}^e$$

$$\overline{\mathbf{L}}^p=\dot{\mathbf{F}}^p\cdot\mathbf{F}^{p^{-1}}$$

3.2 Hyperelastic Constitutive Equations. The original Virtual Internal Bond (VIB) elastic model has been formulated in the reference configuration [6,7,10]. The Cauchy-Born rule of crystal elasticity is used to derive the overall constitutive equations by equating the internal strain energy of the bonds to the potential energy stored in the continuum due to external forces. The initial and the deformed configurations are defined using the Lagrangian coordinates $\mathbf{X}=X_I$ and the Eulerian coordinates $\mathbf{x}=\mathbf{x}(\mathbf{X},t)=x_i(X_I,t)$, respectively. The deformation gradient $\mathbf{F}$ and the Green-Lagrange strain tensor $\mathbf{E}$ are used in the basic elastic description.

Consider an arbitrary microstructural bond at an angle θ and ϕ, where θ is the angle between the bond and the vertical positive z-axis, and ϕ is the angle in the horizontal plane with respect to the positive x-axis (in polar coordinates), respectively. The unit vector along this direction is given as $\xi=(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta)$ with respect to the undeformed configuration. The stretch of this bond can be given as

$$l=l_o\sqrt{1+2\xi_I E_{IJ}\xi_J} \tag{32}$$

The macroscopic strain energy density function is derived using the Cauchy-Born rule [9,26] as

$$\psi(E_{IJ})=\langle U(l)\rangle \tag{33}$$

where $\langle\ldots\rangle$ represents the weighted average with respect to the bond density function D_d. $U(l)$ is the potential energy function between each bond. Assuming that all bonds have the same initial length l_o, for the general case this is given as,

$$\langle\cdots\rangle=\int_0^{2\pi}\int_0^{\pi}\cdots D_d(\theta,\phi)\sin\theta d\theta d\phi \tag{34}$$

The term $D_d(\theta,\phi)\sin(\theta)d\theta d\phi$ represents the number of bonds per unit volume between the bond angles $(\theta,\theta+d\theta)$ and $(\phi,\phi+d\phi)$. For isotropic solids the bond density function is taken as a constant D_o. Hence the macroscopic strain energy density function can be now given as

$$\psi(E_{IJ})=D_o\int_0^{2\pi}\int_0^{\pi}U(l)\sin\theta d\theta d\phi \tag{35}$$

For a two-dimensional isotropic solid subjected to plane stress, the bond density function can be expressed as $D_d\delta[\theta-(\pi/2)]$ and the strain energy density function becomes

$$\psi=D_o\int_0^{2\pi}U(l)d\phi \tag{36}$$

From the strain energy density function ψ, the symmetric second Piola-Kirchoff stress S_{IJ} and the elastic modulus A_{IJKL} can be derived as follows:

$$\mathbf{S}=\frac{\partial\psi}{\partial\mathbf{E}}\quad\text{or}\quad S_{IJ}=\frac{\partial\psi}{\partial E_{IJ}} \tag{37}$$

$$A_{IJKL}=\frac{\partial^2\psi}{\partial E_{IJ}\partial E_{KL}} \tag{38}$$

The modulus derived from this potential satisfies the major and minor symmetries, $A_{IJKL}=A_{JIKL}=A_{IJLK}=A_{KLIJ}$, as well as the Cauchy symmetry, $A_{IJKL}=A_{IKJL}$. This results in only one isotropic elastic constant being needed. This is due to the fact that the Cauchy symmetry is satisfied by the fourth-order isotropic elasticity tensor only for the case of $\lambda=\mu$, where λ and μ are the two Lamé constants.

3.2.1 Adaptation to the Proposed Model. In the proposed model, since the intermediate configuration is considered to be a fixed configuration associated only with lattice deformations, the elastic bonds between the particles are not affected in this configuration. Hence, the elastic VIB response in the intermediate configuration can now be expressed as

$$\overline{\mathbf{S}}=\frac{\partial\overline{\psi}(\overline{\mathbf{E}}^e)}{\partial\overline{\mathbf{E}}^e}\quad\text{or}\quad \overline{S}_{IJ}=\frac{\partial\overline{\psi}}{\partial\overline{E}^e_{IJ}} \tag{39}$$

where $\overline{\mathbf{S}}$ is the equivalent of second Piola-Kirchoff stress in the intermediate configuration.

3.3 Viscoplastic Response Formulation. The intermediate configuration can be uniquely described by the plastic part of the deformation gradient $\mathbf{F}^p$, and the evolution equations can be set up to describe it using an appropriate flow rule. The two internal variables, namely, the backstress $\overline{\alpha}$ and the effective accumulated plastic strain $\overline{\epsilon^p}=\int_0^t\dot{\overline{\epsilon}}^p dt$, and the evolution equations in the rate-dependent form are described here. The constitutive equations formulated here is based on the elasto-viscoplastic response of a material following the J_2 flow theory.

The expressions for the deviatoric form of the second Piola-Kirchoff stress and the backstress (in the intermediate configuration) can be expressed as follows:

$$\overline{\mathbf{S}}^d=\overline{\mathbf{S}}-\tfrac{1}{3}(\overline{\mathbf{S}}:\overline{\mathbf{C}}^e)\overline{\mathbf{C}}^{e^{-1}} \tag{40}$$

$$\overline{\alpha}^d=\overline{\alpha}-\tfrac{1}{3}(\overline{\alpha}:\overline{\mathbf{C}}^e)\overline{\mathbf{C}}^{e^{-1}} \tag{41}$$

and the hydrostatic pressure p is given by the equation $3Jp=\overline{\mathbf{S}}:\overline{\mathbf{C}}^e$, where J is the determinant of the deformation gradient.

Yield function. The yield function in the intermediate configuration is described as

$$\overline{\Phi}(\overline{\mathbf{S}},\overline{\alpha},\overline{\epsilon^p})=0 \tag{42}$$

For the J_2 flow theory the yield function takes the form

$$\overline{\Phi}=\overline{\sigma}^2_{eq}-\kappa^2=0 \tag{43}$$

where the equivalent stress in the intermediate configuration is given as

$$\sigma^2_{eq}=\tfrac{3}{2}(\overline{\mathbf{S}}^d-\overline{\alpha}^d):(\overline{\mathbf{S}}^d-\overline{\alpha}^d)\overline{\mathbf{C}}^e\overline{\mathbf{C}}^e \tag{44}$$

Flow rule. The flow rule is expressed as the evolution of the plastic part of the velocity gradient as follows:

$$\overline{\mathbf{L}}^p=\dot{\mathbf{F}}^p\mathbf{F}^{p^{-1}}=\overline{\mathbf{D}}^p+\overline{\mathbf{W}}^p \tag{45}$$

$$=\dot{\lambda}\overline{\mathbf{n}}(\overline{\mathbf{S}},\overline{\alpha},\overline{\epsilon^p}) \tag{46}$$

where $\overline{\mathbf{n}}$ is a second-order tensor representing the direction of plastic flow. Adopting the associated flow rule, this direction is normal to the yield surface. Using Eq. (44), the plastic flow direction can be expressed as

$$\overline{\mathbf{n}}=\frac{3}{2\overline{\sigma}_{eq}}(\overline{\mathbf{S}}^d-\overline{\alpha}^d)\overline{\mathbf{C}}^e\overline{\mathbf{C}}^e \tag{47}$$

and the evolution of the effective plastic strain rate is then given as follows:

$$\dot{\overline{\epsilon}}^p=\dot{\lambda}=\frac{\overline{\Phi}(\overline{\mathbf{S}},\overline{\alpha},\overline{\epsilon^p})}{\eta} \tag{48}$$

where η is the viscosity parameter of the material (also referred to as the fluidity coefficient). The plastic part of the spin tensor ($\overline{\mathbf{W}^p}=0$) is not considered. These considerations result in a reduced form of the flow rule shown below

$$\overline{\mathbf{D}}^p=\dot{\lambda} symm(\overline{n}) \tag{49}$$

Hardening rule. The evolution of backstress is given by a kinematic hardening rule as follows:

$$\dot{\overline{\alpha}}=\dot{\lambda}\overline{\mathbf{h}}(\overline{\mathbf{S}},\overline{\alpha},\overline{\epsilon^p}) \tag{50}$$

where $\overline{\mathbf{h}}$ is the hardening function.

4 Integration Procedure

The details and issues involved with the numerical implementation of the model proposed above are described in this section. The steps outlined here are presented in a format that is suitable for implementation using the ABAQUS subroutine VUMAT. In the material model, which involves the integration of the constitutive equations, the following data are given as input from ABAQUS: $\mathbf{F}_t$, $\mathbf{F}_{t+1}$, Δt, which represent the deformation gradients at the previous and current time step and the increment of time, respectively. The elastic and plastic deformation gradients at time t denoted by $\mathbf{F}_t^e$ and $\mathbf{F}_t^p$, respectively, and the effective plastic strain $\overline{\epsilon^p}$ are designated and computed as state variables within the material model subroutine. The Cauchy stress σ_t from the previous time step is also returned back by ABAQUS for use if needed. The Cauchy stress tensor at time (t+1), σ_{t+1}, along with the updated values of $\mathbf{F}_{t+1}^p$, $\mathbf{F}_{t+1}^e$, $\overline{\epsilon^p}$ are calculated and stored at the end of the current time step.

4.1 Numerical Integration Procedure

Step 1. The flow rule is integrated using an *implicit exponential approximation*, keeping the flow direction and the plastic modulus fixed during the current time step, as follows:

$$*\mathbf{F}_{n+1}^p=\exp[\Delta\overline{\epsilon^p}_{n+1}\overline{\mathbf{n}}_n]\overline{\mathbf{F}}_n^p \tag{51}$$

The tensor exponential function in Eq. (51) can be represented by a series representation as shown below

$$\exp[\mathbf{A}]=\sum_{n=0}^{\infty}\frac{1}{n!}[\mathbf{A}]^{\mathbf{n}} \tag{52}$$

Considering the first two terms of the expansion of Eq. (51) and using the series representation, the flow rule and the resulting trial plastic deformation gradient can be expressed as given below

$$*\mathbf{F}_{t+1}^p=[\mathbf{I}+\Delta\overline{\epsilon^p}_{t+1}\overline{\mathbf{n}}_t]\overline{\mathbf{F}}_t^p \tag{53}$$

where $\mathbf{I}$ is the second order identity tensor.

Step 2. The *trial* elastic deformation gradient for the current step is computed as follows:

$$\begin{gathered}*\mathbf{F}_{t+1}^e=(\mathbf{F})_{t+1}\mathbf{F}_{t+1}^{p^{-1}}\\ *F_{iJ}^e=(F_{iK})_{t+1}(F_{kJ})_{t+1}^{p^{-1}}\end{gathered} \tag{54}$$

Step 3. The Green strain can then be computed as shown below

$$*\overline{\mathbf{E}}_{t+1}^e=\tfrac{1}{2}[*\mathbf{F}_{t+1}^{e^T}*\mathbf{F}_{t+1}^e-\mathbf{I}] \tag{55}$$

Step 4. The *trial elastic* second Piola-Kirchoff stress (based on the computation of new bond lengths in the intermediate configuration and assuming that the hyperelastic potential is based on the elastic stretch only) is calculated using the following equation:

$$\begin{gathered}*\overline{\mathbf{S}}_{t+1}=\frac{\partial\overline{\psi}(*\overline{\mathbf{E}}_{t+1}^e)}{\partial*\overline{\mathbf{E}}_{t+1}^e}\\ *(\overline{S_{IJ}})_{t+1}=\frac{\partial\overline{\psi}}{\partial*\overline{E_{IJ}^e}}\end{gathered} \tag{56}$$

Step 5. The trial backstress is then updated by the following equation:

$$*\overline{\alpha}_{t+1}=\overline{\alpha}_t+\Delta\overline{\epsilon^p}_{t+1}\mathbf{h}_n \tag{57}$$

Step 6. The equivalent plastic strain, actual equivalent stress, and internal scalar or tensor variables can then be solved using a Newton iteration. The equivalent plastic strain increment can be expanded as

$$\Delta\overline{\epsilon^p}_{t+1}=\Delta t\dot{\overline{\epsilon}}^p(*\overline{\mathbf{S}}_{t+1},*\mathbf{F}_{t+1}^p,\Delta\overline{\epsilon^p}_{t+1},\overline{\alpha}_{t+1}) \tag{58}$$

The equivalent plastic strain increment in Eq. (58) is solved from the steps described using the Newton-Raphson technique with an initial guess of zero. This value is incremented until Eq. (58) is satisfied to some acceptable tolerance. The algorithm described above is a combination of both explicit and implicit steps.

5 Numerical Implementation

The numerical implementation of the finite deformation viscoplastic VIB model, described in the above sections, has been tested on a plate with a center-hole (PWH) problem. A number of issues, such as crack initiation and propagation, stress-strain behavior of cracked elements, development of plastic strain in elements, time of solution, etc., have been studied. The results are presented in this paper. Simulations have also been run using the elastic VIB model described in [13,14] and compared with the viscoplastic model.

In the simulations, 6061 Al has been used with a Young's Modulus of 70 GPa and a Poisson ratio of 0.3. The material has an initial yield stress $\overline{\sigma}_0=135$ MPa. The size of the quarter plate is taken as 100×100 mm. The plastic flow for this material is described by a power law given by the following equation [27]:

$$\overline{\sigma}=a+b(c+\kappa)^d \tag{59}$$

where the material constants with $a=25$ MPa, $b=466$ MPa, $c=0.003$, and the exponent $d=0.293$. κ is the equivalent plastic strain. The fluidity coefficient η used in the effective plastic strain-rate equation (48) is 5000 MPa/s.

5.1 Plate With Hole Problem: Unidirectional Displacement Loading. The loading cases studied for this problem are shown in Fig. 2. The displacement at the top edge of the plate is increased linearly in all cases with the rate being controlled by the

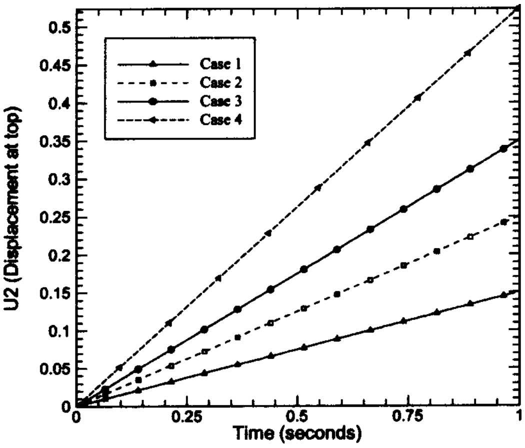


Fig. 2 Loading cases studied for plate with hole

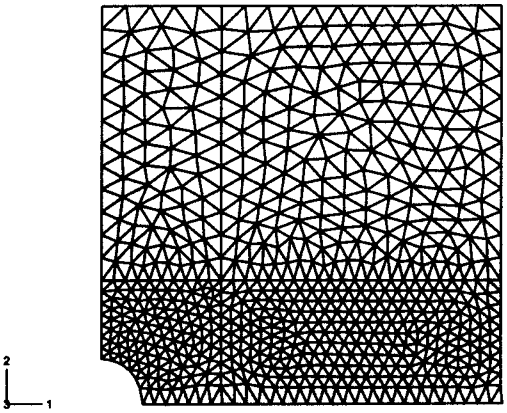


Fig. 3 Final deformation for elastic VIB model (case 2)

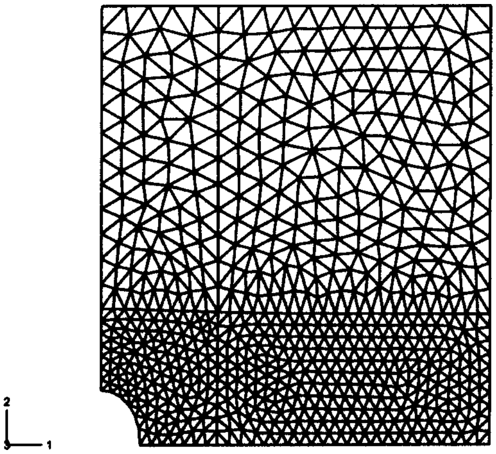


Fig. 5 Final deformation for viscoplastic VIB model (case 3)

final displacement value. All displacements are given during a real time of 1 s. The four cases correspond to a maximum displacement (U2) of 1.5, 2.5, 3.5, and 5.3 mm, respectively, of the top of the plate. Both the elastic VIB model and the elasto-viscoplastic VIB (EPVIB) model have been run on the same data set.

For case 1 (U2=1.5 mm), no cracking was observed for both the models, and hence, the results are not shown graphically. Figures 3 and 4 show the deformation of the VIB and EPVIB models, respectively, for case 2 (U2=2.5 mm). A fully cracked band at the bottom edge can be seen for the elastic case (Fig. 3) while no cracking is observed in the EPVIB case (Fig. 4). For subsequent load cases the elastic deformation pattern is not shown as the bottom edge is fully cracked in all cases and is similar to that shown in Fig. 3. Figure 5 shows the final deformation for loading case 3 (U2=3.5 mm) with the EPVIB model. A small crack is seen to develop at the bottom left edge with two or three elements showing cracks. Figure 6 shows the deformation pattern for case 4 (U2=5.3 mm) at time 0.7 s from which it can be seen that the crack tip located at about one-third of the base length from the edge of the hole.

Figure 7 shows the stress distribution along the path defined by the bottom edge of the plate (for loading case 4), showing the stress values with increasing time, after cracking. The peak stress is indicative of the current location of the crack tip, shifts to the right. After the crack has fully propagated throughout the bottom edge, the stresses in these elements drop significantly, as shown in Fig. 7 at time steps $t=0.9$ s. Figure 8 shows the plot of crack tip location with time up to the time of 0.875 s. It was observed earlier that at 0.9 s the crack suddenly propagates throughout the bottom of the plate.

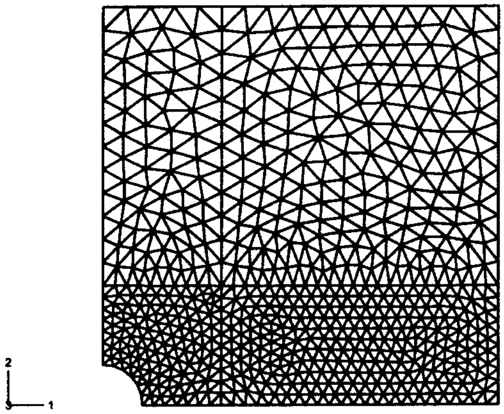


Fig. 4 Final deformation for viscoplastic VIB model (case 2)

The stress-strain behavior comparison of a critical element is studied by plotting the stress along the loading direction σ_{22} with the logarithmic strain LE_{22}, for the left-bottom corner element (the first element at the edge of the hole). The results have been presented comparing the elastic and the viscoplastic response for various loading cases. For loading case 3 the response is shown in

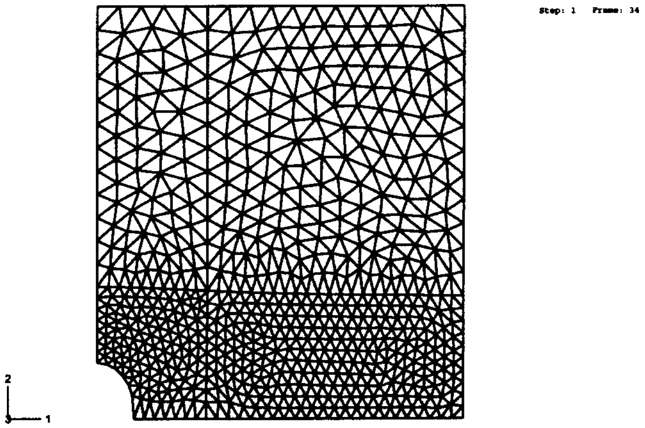


Fig. 6 Final deformation for viscoplastic VIB model (t=0.7 s) (case 4)

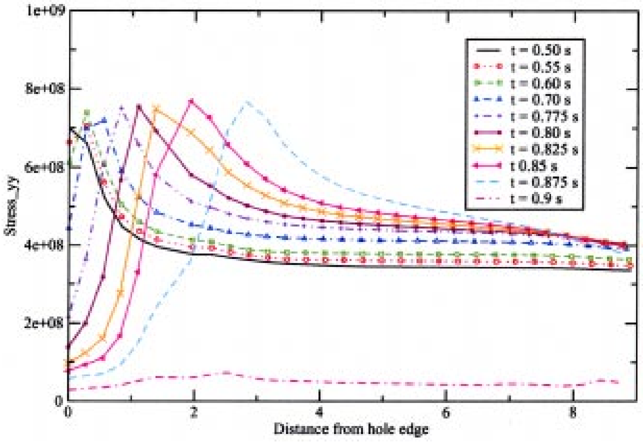


Fig. 7 Stress (σ_{yy}) distribution along bottom edge of plate (after cracking)

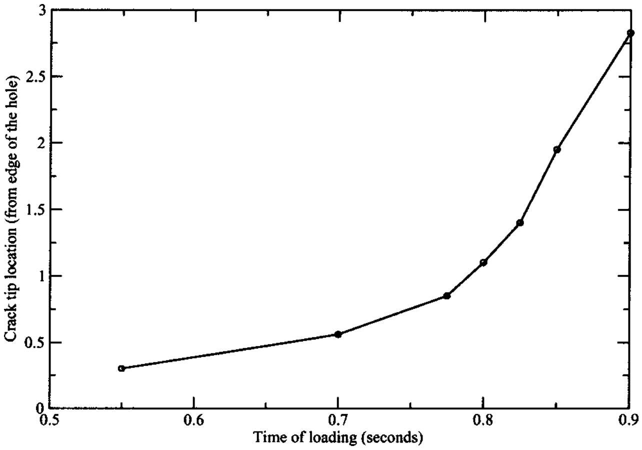


Fig. 8 Plot of crack tip location with time

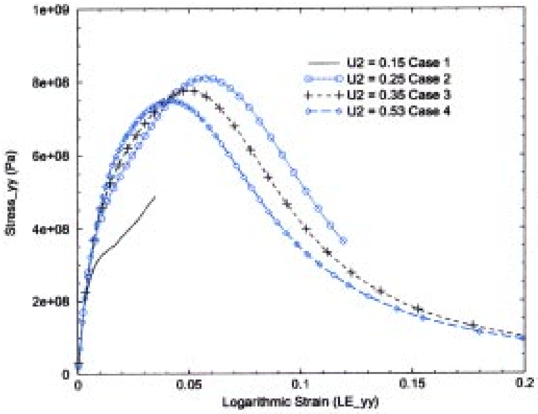


Fig. 11 Comparison of viscoplastic model stress-strain curves

Fig. 9. The peak of the viscoplastic response is higher than the elastic response and is due to the incorporation of the hardening behavior exhibited by the material.

Figure 10 shows the plot of the equivalent Mises stress with the equivalent plastic strain, for the left-bottom corner element. The intent of this figure is to study the effect of rate of loading on the development of plastic strains. After yielding, for loading cases 1 and 2 equivalent plastic strains are allowed to develop as the loading rates are fairly low. In cases 3 and 4, which are simulated at a higher loading rate, the equivalent Mises stresses developed are higher while the plastic strains developed are lower compared to loading cases 1 and 2. The sudden drop in the stress in loading cases 3 and 4 are a consequence of the cracking of the element.

Figure 11 shows the comparative response for various loading cases for the viscoplastic model for the left corner element. The elastic response, which is not shown here, for these loading rates was observed to be identical for all the loading cases up to the point of maximum stress. The deviations in the softening region are due to numerical issues. For the viscoplastic model (Fig. 11), as the loading rate increases the peak stress and peak strain sustained appears to decrease. This might indicate that, for higher loading rates, the tendency to fracture is higher as plastic strains do not have time to develop.

The influence of material parameters (such as yield strength, fluidity coefficient, and hardening coefficient) on fracture properties, particularly, the cohesive strength and strain at the cohesive strength, have been studied. Figures 12–14 show the stress-strain behavior of the left-bottom corner element of the plate for different values of yield strength, fluidity coefficient, and hardness coefficient, respectively. The loading rate was kept the same in all these simulations.

From Fig. 12, which shows the variation for different yield strength values with the hardness and fluidity coefficient remaining the same, it can be seen that cohesive strength is not affected very much by the change in yield strength. The cohesive strain for

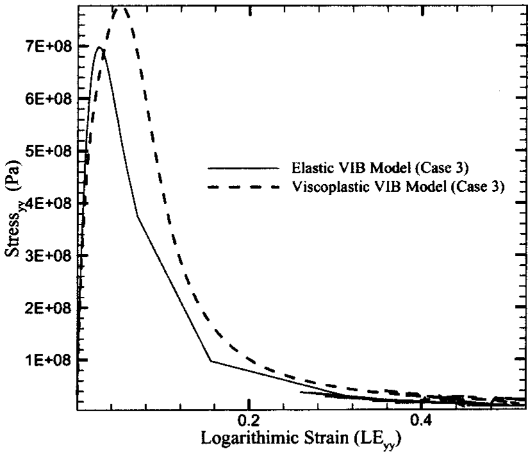


Fig. 9 Stress-strain curve for left-bottom corner element (case 3)

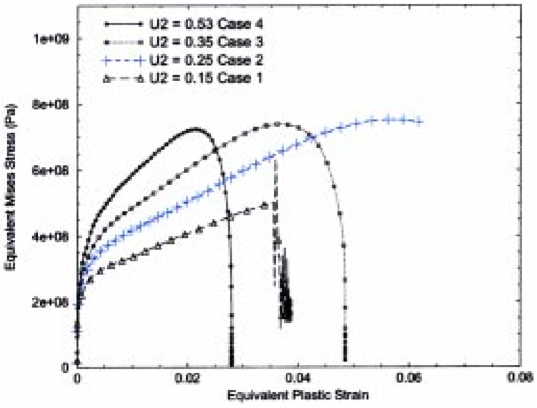


Fig. 10 Mises stress-equivalent plastic strain plots for left-corner element

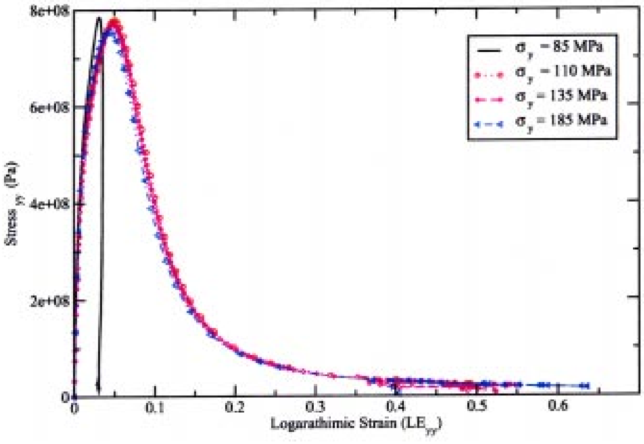


Fig. 12 Comparison of stress strain curves for different yield strengths

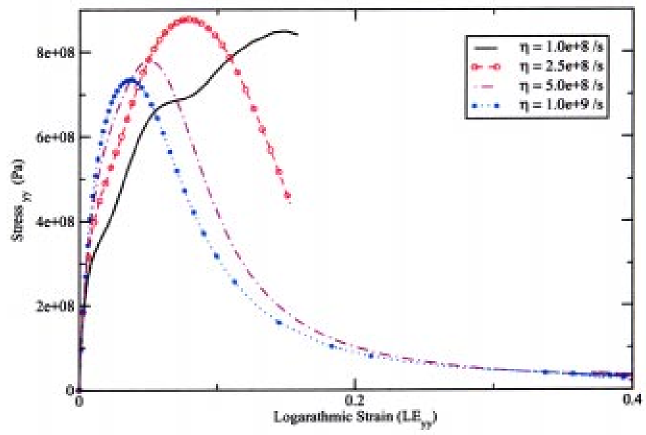


Fig. 13 Comparison of stress strain curves for different fluidity coefficients

yield strength of an 85 MPa value is lower compared to the yield strength values of 110, 135, and 185 MPa, which have very similar cohesive strain values.

Figure 13 shows the variation of the stress-strain behavior for the left-bottom corner element with varying fluidity coefficient values. It can be clearly seen that as the fluidity coefficient increases, both the cohesive stress and the cohesive strain values decrease, indicating that the there is a faster tendency to fracture. Figure 14, which shows the variation of the stress strain curves for different values of hardness coefficient κ, indicates that the hardness coefficient has negligible influence on fracture.

5.2 Plate With Hole Problem: Biaxial Loading. The PWH problem was repeated with a biaxial load. The magnitude of loading was similar to that of loading case 3 of the uniaxial case with the loading now applied in two directions. A uniform velocity in both the x and y directions was applied over one second. Figures 15 and 16 show the final cracked configuration for the elastic and viscoplastic cases, respectively. A crack emanating from approximately the center of the quarter circular arc and propagating at a 45-deg angle can be clearly seen. The elastic model cracks earlier than the viscoplastic model case and also propagates much further. From Fig. 17, which shows a comparison of the stress and the logarithmic strain for the biaxial and uniaxial cases, it can be seen that the cohesive stress for biaxial stretching is lower and occurs at a lower strain value when compared to the uniaxial case. This observation is consistent with the theoretical derivation shown in (Fig. 6 of Gao and Klein [8]). Figure 18 shows the comparison of the Mises stress and the equivalent plastic strain plots. Since the cohesive stress for the biaxial case is lower, the overall development of the equivalent plastic strain is also lower.

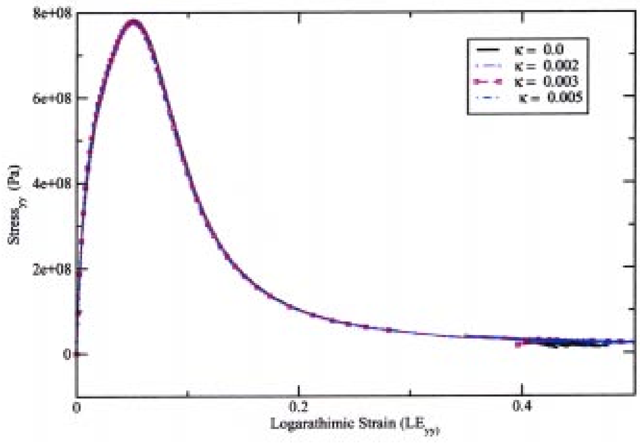


Fig. 14 Comparison of stress strain curves for different hardness coefficients

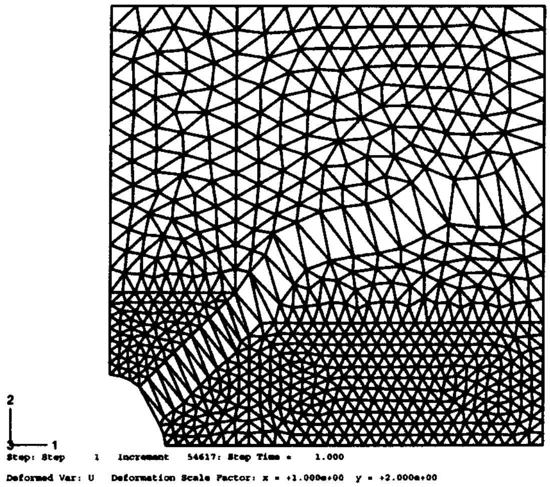


Fig. 15 Cracked pattern for biaxial load case with elastic VIB model

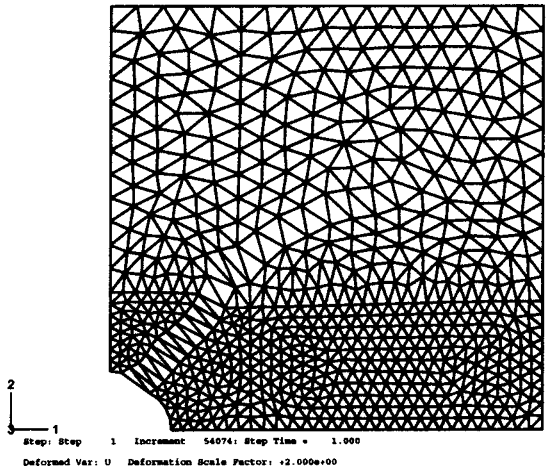


Fig. 16 Cracked pattern for biaxial load case with viscoplastic VIB model

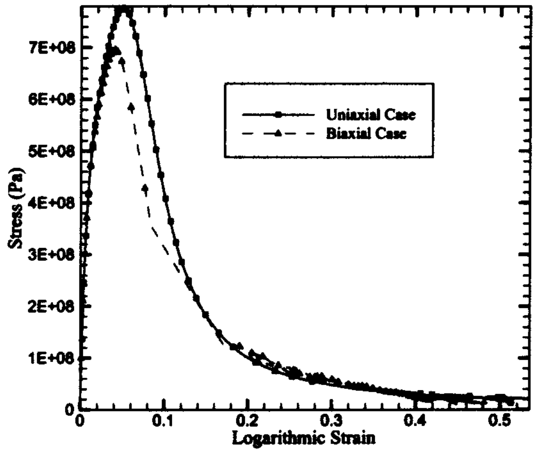


Fig. 17 Comparison of stress and logarithmic strains for the elastic and viscoplastic models

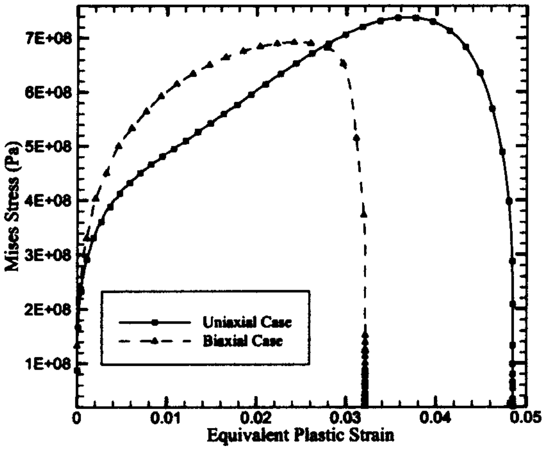


Fig. 18 Comparison of equivalent plastic strains for uniaxial and biaxial loading cases

6 Conclusions

In this paper an effective and robust finite element implementation of a hyperelasto-viscoplastic virtual internal bond model has been presented to simulate crack initiation, propagation, and branching in ductile materials. The following conclusions can be drawn from the paper.

1. Since the model is set up in an explicit integration framework, the softening region does not pose any difficulties due to possible negative eigenvalues or singularities in the stiffness matrix.
2. The effect of the plastic zone ahead of the crack tip is clearly demonstrated in the deformation patterns exhibited by the elastic and the viscoplastic VIB model.
3. The effect of strain hardening is also seen in the form of increased cohesive stress limits in the viscoplastic case.
4. The effect of loading rate on the development of plastic strains and crack formation has also been shown. It is seen that as loading rate increases, while the Mises stresses are higher, the plastic strains developed are lower. The material tends to crack earlier as the loading rate increases.
5. In parametric studies in which the effect of yield strength, fluidity coefficient, and hardness coefficient on fracture was studied, it was found that the yield strength and the hardness coefficient did not have a significant influence on the cohesive stress. However, an increase in the fluidity coefficient resulted in a decrease in the cohesive stress as well as the strain value.
6. The cohesive stress for biaxial stretching is lower and occurs at a lower strain value when compared to the uniaxial case. This observation is consistent with the theoretical derivation shown in (Fig. 6 of Gao and Klein [8]).

Acknowledgments

G.T. acknowledges the partial support by National Computational Science Alliance (Grant No. OND), and utilized the SGI Origin 2000 and the University of Missouri Research Board Grant (UMRB). Y.H. acknowledges the support from NSF (Grant No. 9983779), ONR (Grant No. N00014-01-1-0205, program officer Dr. Y. D. S. Rajapakse), and ASCI Center for Simulation of Advanced Rockets at the University of Illinois, supported by U.S. Department of Energy through the University of California under Subcontract No. B523819. K.J.H. also acknowledges partial financial support from NSF.

References

[1] Barenblatt, G. I., 1959, ''The Formation of Equilibrium Cracks During Brittle Fracture: General Ideas and Hypotheses, Axially Symmetric Cracks,'' J. Appl. Math. Mech., **23**, pp. 622–636.
[2] Dugdale, D. S., 1960, ''Yielding of Steel Sheets Containing Silts,'' J. Mech. Phys. Solids, **8**, pp. 100–104.
[3] Willis, J. R., 1967, ''A Comparison of the Fracture Criteria of Griffith and Barenblatt,'' J. Mech. Phys. Solids, **15**, pp. 151–162.
[4] Xia, L., and Shih, F. C., 1995, ''Ductile Crack Growth-I. A Numerical Study Using Computational Cells With Microstructurally Based Length Scales,'' J. Mech. Phys. Solids, **43**(2), pp. 233–259.
[5] Xu, X. P., and Needleman, A., 1994, ''Numerical Simulation of Fast Crack Growth in Brittle Solids,'' J. Mech. Phys. Solids, **42**(9), pp. 1397–1434.
[6] Gao, H., 1997, ''Elastic Waves in a Hyperelastic Solid Near its Plane Strain Equibiaxial Cohesive Limit,'' Philos. Mag. Lett., **76**, pp. 307–314.
[7] Gao, H. A., 1996, ''Theory of Local Limiting Speed in Dynamic Fracture,'' J. Mech. Phys. Solids, **44**, pp. 1453–1474.
[8] Gao, H., and Klein, P., 1998, ''Numerical Simulation of Crack Growth in an Isotropic Solid With Randomized Internal Cohesive Bonds,'' J. Mech. Phys. Solids, **46**(2), pp. 187–218.
[9] Tadmor, E. B., Ortiz, M., and Phillips, R., 1996, ''Quasicontinuum Analysis of Defects in Solids,'' Philos. Mag. A, **73**, pp. 1529–1563.
[10] Klein, P., and Gao, H., 1998, ''Crack Nucleation and Growth as Strain Localization in a Virtual Internal Bond Continuum,'' Eng. Fract. Mech., **61**, pp. 21–48.
[11] Zhang, P., Klein, P., Huang, Y., Gao, H., and Wu, P. D., 2002, ''Numerical Simulation of Cohesive Fracture by the Virtual Internal Bond Method,'' Comput. Model. Eng. Sci., **3**, pp. 263–278.
[12] *ABAQUS/Explicit Users Manual*, HKS Inc.
[13] Thiagarajan, G., Hsia, K. J., and Huang, Y. Y., 2004, ''Finite Element Implementation of Virtual Internal Bond Model for Simulating Crack Behavior,'' Eng. Fract. Mech., **71**, pp. 401–423.
[14] Thiagarajan, G., Hsia, K. J., and Huang, Y. Y., 2002, ''Experimental Verification and Validation of Virtual Internal Bond Parameters for Simulating Dynamic Crack Behavior,'' *Proc. 39th Annual Technical Meeting of Society of Engineering Science-SES 2002*.
[15] Lee, E. H., 1969, ''Elastic Plastic Deformations at Finite Strain,'' ASME J. Appl. Mech., **36**, pp. 1–6.
[16] Rice, M. H., McQueen, R. G., and Walsh, J. M., 1958, ''Compression of Solids by Strong Shock Waves,'' Solid State Phys., **6**, pp. 1–63.
[17] Lee, E. H., and Liu, D. T., 1967, ''Finite-Strain Elastic-Plastic Theory With Application to Plane Wave Analysis,'' J. Appl. Phys., **38**, pp. 19–27.
[18] Lee, E. H., and Germain, P., 1972, ''Elastic-Plastic Theory at Finite Strain,'' *Problems of Plasticity*, A. Sawczuk, ed., Noordhoff International Publishing, pp. 117–133.
[19] Simo, J. C., and Ortiz, M., 1985, ''A Unified Approach to Finite Deformation Elastoplastic Analysis Based on the Use of Hyperelastic Constitutive Equations,'' Comput. Methods Appl. Mech. Eng., **49**, pp. 221–245.
[20] Weber, G., and Anand, L., 1990, ''Finite Deformation Constitutive Equations and a Time Integration Procedure for Isotropic, Hyperelastic-Viscoplastic Solids,'' Comput. Methods Appl. Mech. Eng., **79**, pp. 173–202.
[21] Needleman, A., 1985, ''On Finite Element Formulation for Large Elastic-Plastic Deformations,'' Comput. Struct., **20**, pp. 247–257.
[22] Green, A. E., and Naghdi, P. M., 1965, ''A General Theory of an Elastic-Plastic Continuum,'' Arch. Ration. Mech. Anal., **18**, pp. 251–281.
[23] Marsden, J. E., and Hughes, T. J. R., 1983, *Mathematical Foundations of Elasticity*, Dover, New York.
[24] Kratochvil, J., 1972, ''Comment on Elastic Plastic Rotations,'' *Problems of Plasticity*, A. Sawczuk, ed., Noordhoff International Publishing, pp. 413–416.
[25] Wang, C. C., 1970, ''A New Representation Theorem for Isotropic Functions,'' Arch. Ration. Mech. Anal., **36**, p. 198.
[26] Milstein, F., 1980, ''Review: Theoretical Elastic Behavior at Large Strains,'' J. Mater. Sci., **15**, pp. 1071–1084.
[27] Wang, W. M., Pozivilova, A., and Sluys, A. J., 2000, ''Implicit Algorithms for Finite Deformation Viscoplasticity,'' *European Congress on Computational Methods in Applied Sciences and Engineering, ECCOMAS 2000*, pp. 1–15.

A Superposition Framework for Discrete Dislocation Plasticity

M. P. O'Day

W. A. Curtin

Division of Engineering,
Brown University,
Providence, RI 02912

A superposition technique is introduced that allows for the application of discrete dislocation (DD) plasticity to a wide range of thermomechanical problems with reduced computational effort. Problems involving regions of differing elastic and/or plastic behavior are solved by superposing the solutions to i) DD models only for those regions of the structure where dislocation phenomena are permitted subject to either zero traction or displacement at every point on the boundary and ii) an elastic (EL) (or elastic/cohesive-zone) model of the entire structure subject to all desired loading and boundary conditions. The DD subproblem is solved with standard DD machinery for an elastically homogeneous material. The EL subproblem requires only a standard elastic or elastic/cohesive-zone finite element (FE) calculation. The subproblems are coupled: the negative of the tractions developed at the boundaries of the DD subproblem are applied as body forces in the EL subproblem, while the stress field of the EL subproblem contributes a driving force to the dislocations in the DD subproblem structure. This decomposition and the generic boundary conditions of the DD subproblem permit the DD machinery to be easily applied as a "black-box" constitutive material description in an otherwise elastic FE formulation and to be used in a broader scope of applications due to the overall enhanced computational efficiency. The method is validated against prior results for crack growth along a plastic/rigid bimaterial interface. Preliminary results for crack growth along a metal/ceramic bimaterial interface are presented. [DOI: 10.1115/1.1794167]

1 Introduction

The proliferation of increasingly smaller structures has highlighted the necessity of developing accurate modeling techniques for material deformation at these scales. Currently, micron scale analysis is important for micromachines and microelectronic components, as well as for the fundamental modeling of fracture processes. A wealth of experimental evidence has shown that in metal specimens having characteristic dimensions less than 100 μm, plastic flow exhibits a size effect: smaller is stronger [1,2]. Classical continuum plasticity does not include a length scale, precluding any size effects. Thus the application of classical plastic constitutive laws to micron scale specimens is questionable. The micron-size scale effectively lies in an intermediate regime that is too large for fully atomistic modeling, but small enough that individual dislocation effects are important and cannot be averaged into a classical continuum plasticity constitutive law. Over the last decade several new methods have been developed that can more accurately predict deformation at smaller scales.

Numerous nonlocal plasticity theories have been developed in an attempt to reproduce size effects [3–5]. These theories introduce a length scale that aims to account for the effect of geometrically necessary dislocations. As continuum theories, most nonlocal models can be incorporated into existing finite element (FE) architecture with only slight modifications. However, there remain several drawbacks to nonlocal formulations. Like classical plasticity, these are phenomenological theories that do not account for the fundamental basis of plasticity, i.e., the motion of dislocations. The specification of the various length scales, usually chosen by a fit to experimental data, is also an outstanding issue [6].

An alternative technique is the discrete dislocation (DD) method of Van der Giessen and Needleman [7]. The DD approach solves boundary value problems (bvp) for isotropic elastic bodies containing mobile dislocations, which carry the information about plastic deformation. No assumptions about the material plastic constitutive response are necessary although a set of constitutive rules governing dislocation motion, nucleation, and annihilation are required. The DD method has the benefit of being a true mechanism-based theory of plasticity; plastic flow arises directly from dislocation motion. The DD method has recently been applied to study flow in composites of varying microstructure [8], crack growth in plastic materials [9], fatigue crack growth [10], and size effects in model Al/Si alloys [11]. The DD method has also proven useful in acting as a numerical experiment for comparison with various nonlocal theories [6,12]. One limitation of the method is its significant computational cost, particularly for bodies containing elastic inhomogeneities [7] and/or large numbers of dislocations. For this reason much of the work to date has focused on elastically homogeneous systems, often with some degree of symmetry to further simplify the computation. Consequently, many interesting physical problems that demand noncontinuum treatments have yet to be examined within the powerful DD framework.

Here a new superposition technique is presented that allows for a computationally efficient solution of elastically inhomogeneous DD problems. The problem of interest is solved by superposing the solutions to i) a DD model of only that portion of the structure where dislocation phenomena are permitted subject to special boundary conditions and ii) an elastic (EL) model of the entire structure subject to all desired loading and boundary conditions. The DD subproblem is homogeneous and solved with standard DD machinery, including a contribution to the Peach-Koehler forces on the dislocations coming from the EL subproblem. The EL subproblem is solved with a standard elastic FE calculation including special body forces that emerge from the DD problem. This technique is basically a special application of the Eshelby method [13]. It is similar to the coupled atomistic/continuum method of Shilkrot et al. [14], although here the formulation is purely continuum in nature. The new technique also simplifies the application of the DD method to other problems because it separates out a generic DD subproblem that can be considered as a material constitutive law. Finally, the superposition method is

---

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, February 24, 2003; final revision, October 30, 2003. Associate Editor: E. 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 of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

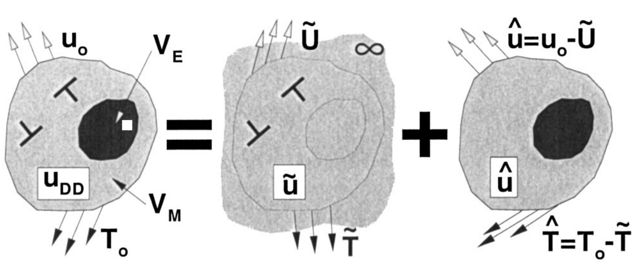


Fig. 1 General discrete dislocation boundary value problem (fields $\mathbf{u}_{DD}$, $\boldsymbol{\sigma}_{DD}$) is written as the superposition of: (i) dislocation fields in infinite space of homogeneous matrix material ($\tilde{\mathbf{u}}$,$\tilde{\boldsymbol{\sigma}}$) and (ii) corrective fields to account for the inclusion and proper boundary conditions ($\hat{\mathbf{u}}$,$\hat{\boldsymbol{\sigma}}$)

ideal for using the DD machinery in a parallel-computing environment. The technique is validated against prior results on two-dimensional (2D) plane-strain crack growth along a rigid/elastic-plastic bimaterial interface, and preliminary results on the fracture of more realistic elastic-plastic/elastic bimaterial interfaces is presented. The superposition formulation and examples discussed will be limited to 2D plane strain. Fundamentally, the technique is equally valid in 3D, although the DD mechanics becomes considerably more complex [15].

The formulation is presented in dyadic notation. Vectors and tensors are given by bold faced symbols, denotes the inner product, and: the trace product. With respect to a Cartesian basis $\mathbf{e}_i$, $\mathbf{a}\cdot\mathbf{b}=a_ib_i$, $\mathbf{A}:\mathbf{B}=A_{ij}B_{ji}$, and $(\boldsymbol{\mathcal{L}}:\mathbf{B})_{ij}=\boldsymbol{\mathcal{L}}_{ijkl}B_{lk}$, with implied summation over repeated indices. Latin indices run from 1 to 3, Greek indices from 1 to 2 only. The gradient operator is denoted as ∇. The fourth-order identity tensor is $\mathcal{I}$.

The remainder of this paper is organized as follows. Section 2 contains a concise overview of the standard DD formulation for inhomogeneous bodies. Section 3 presents the new superposition technique. The material parameters, validation, and preliminary bimaterial results are presented in Section 4. Section 5 discusses other applications of this formulation, and summarizes our results. Computational efficiency for problems involving a nonlinear cohesive zone requires special techniques in the EL subproblem. The Appendix describes an efficient technique for solving the incremental FE equations for an *elastic* bimaterial model with a nonlinear cohesive zone, which is used to solve the EL subproblem in the bimaterial interface crack growth problems.

2 Discrete Dislocation Methodology

The standard DD formulation for the inhomogeneous problem of an elastic-plastic body containing an elastic inclusion has been derived by Van der Giessen and Needleman [7]. The derivation is briefly reviewed here, with a focus on when the implementation becomes computationally expensive. This motivates the development of the new superposition technique, presented in Section 3.

The discrete dislocation formulation models edge dislocations as line defects in an isotropic elastic material, constrained to glide on a fixed slip plane. Long-range dislocation interactions occur through their continuum elastic fields. Short-range interactions are governed by constitutive rules for dislocation motion, nucleation, and annihilation. In addition, dislocations can become pinned at obstacles and are released when the resolved shear stress on the dislocation exceeds the obstacle strength. Nucleation occurs by the expansion of Frank-Read dislocation loops, which in 2D is represented by the creation of a dislocation dipole.

The general discrete dislocation boundary value problem is shown in Fig. 1 for a body of volume V, subject to boundary conditions $\mathbf{u}=\mathbf{u_o}$ on S_u, and $\mathbf{T}=\mathbf{T_o}$ on S_t. The body is composed of an elastic-plastic "matrix" region V_M and an elastic "inclusion" region V_E with tensors of elastic moduli $\mathcal{L}$ and $\mathcal{L}^E$, respectively. The problem of interest is solved as the superposition of a problem containing dislocations in an infinite body of homogeneous *matrix* material, yielding the displacement, strain, and stress fields $\tilde{\mathbf{u}}$, $\tilde{\boldsymbol{\epsilon}}$, and $\tilde{\boldsymbol{\sigma}}$, and a complementary problem that corrects for the actual boundary conditions and the presence of the inclusion, yielding the fields $\hat{\mathbf{u}}$, $\hat{\boldsymbol{\epsilon}}$, and $\hat{\boldsymbol{\sigma}}$. The fields in the problem of interest are then obtained by superposition as

$$\mathbf{u}_{DD}=\tilde{\mathbf{u}}+\hat{\mathbf{u}} \quad \boldsymbol{\epsilon}_{DD}=\tilde{\boldsymbol{\epsilon}}+\hat{\boldsymbol{\epsilon}} \quad \boldsymbol{\sigma}_{DD}=\tilde{\boldsymbol{\sigma}}+\hat{\boldsymbol{\sigma}} \quad \text{in V} \tag{1}$$

The elastic fields of an isolated dislocation in an infinite body are known analytically [16] and given by $\mathbf{u}^i$, $\boldsymbol{\epsilon}^i$, $\boldsymbol{\sigma}^i$. Again by superposition, the ($\tilde{\ }$) fields from n_d dislocations are given by

$$\tilde{\mathbf{u}}=\sum_i \mathbf{u}^i \quad \tilde{\boldsymbol{\epsilon}}=\sum_i \boldsymbol{\epsilon}^i \quad \tilde{\boldsymbol{\sigma}}=\sum_i \boldsymbol{\sigma}^i \quad (i=1,\ldots,n_d) \tag{2}$$

These fields produce tractions and displacements at the real boundary of interest given by

$$\begin{gathered}\boldsymbol{\nu}\cdot\tilde{\boldsymbol{\sigma}}=\tilde{\mathbf{T}} \quad \text{on } S_t\\ \mathbf{u}=\tilde{\mathbf{U}} \quad \text{on } S_u\end{gathered} \tag{3}$$

where $\boldsymbol{\nu}$ is the outward normal to S.

The corrective field is designed such that when superposed with the infinite space dislocation fields, the desired boundary value problem is obtained. The governing equations for the corrective fields are thus

$$\nabla\cdot\hat{\boldsymbol{\sigma}}=\mathbf{0} \quad \hat{\boldsymbol{\epsilon}}=\nabla\hat{\mathbf{u}} \quad \text{in V} \tag{4}$$

$$\hat{\boldsymbol{\sigma}}=\mathcal{L}:\hat{\boldsymbol{\epsilon}} \quad \text{in } V_M \tag{5}$$

$$\hat{\boldsymbol{\sigma}}=\mathcal{L}^{\mathbf{E}}:\hat{\boldsymbol{\epsilon}}+(\mathcal{L}^{\mathbf{E}}-\mathcal{L}):\tilde{\boldsymbol{\epsilon}} \quad \text{in } V_E \tag{6}$$

subject to the "corrective" boundary conditions

$$\begin{gathered}\boldsymbol{\nu}\cdot\hat{\boldsymbol{\sigma}}=\hat{\mathbf{T}}=\mathbf{T_o}-\tilde{\mathbf{T}} \quad \text{on } S_t\\ \hat{\mathbf{u}}=\mathbf{u_o}-\tilde{\mathbf{U}} \quad \text{on } S_u\end{gathered} \tag{7}$$

Since the ($\tilde{\ }$) fields are singular only at the dislocation cores and since dislocations occur in dipole pairs or terminate on traction-free surfaces, the ($\tilde{\ }$) fields on the boundary S and the boundary conditions (7) for the corrective field problem are smooth, and thus the corrective field problem can be solved with the conventional FE method.

With the dislocation structure and all fields known at some instant, the evolved structure and fields are desired after an increment in applied loading. Based on the known dislocation structure, the boundary fields $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{T}}$ are calculated. The corrective FE problem is then solved for an increment of applied loading. With the total fields determined, the evolution of the dislocation structure is accomplished by i) evaluation of the Peach-Koehler force on each dislocation and ii) application of the rules for dislocation motion, nucleation, and annihilation. The updated dislocation structure and new fields are now known, and this procedure is repeated for all subsequent increments. The Peach-Koehler force $f^{(I)}$ on the Ith dislocation is computed as

$$f^{(I)}=n_i^{(I)}\left(\hat{\sigma}_{ij}+\sum_{J\neq I}\sigma_{ij}^{(J)}\right)b_j^{(I)} \tag{8}$$

where $n_i^{(I)}$ is the slip plane normal and $b_i^{(I)}$ is the Burgers vector of the Ith dislocation. A key point is that the method does not solve for equilibrium dislocation distributions. The dislocation velocity is linearly related to the Peach-Koehler force; no dissipative mechanism to slow the dislocations is included. At any instant, the dislocation structure is a snapshot of the constantly evolving dislocation structure. Since an equilibrium solution is not being sought, no self-consistent iteration between the two subproblems is necessary. The FE framework used here is that of Cleveringa et al. [9], which is quasi-static and uses a virtual work expansion to step forward in time without iteration.

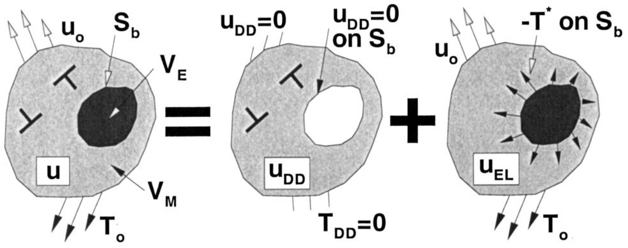


Fig. 2 New superposition framework showing decomposition into two subsidiary problems: discrete dislocation (DD) subproblem solved with the standard formulation subject to generic boundary conditions, and elastic (EL) subproblem, which contains all specific boundary conditions and loading, solved with standard elastic FE

The additional term $\hat{\mathbf{p}}=(\mathcal{L}^{\mathrm{E}}-\mathcal{L}):\tilde{\boldsymbol{\epsilon}}=(\mathcal{L}^{\mathrm{E}}:\mathcal{L}^{-1}-\mathcal{I}):\tilde{\boldsymbol{\sigma}}$ in Eq. (6) that corrects for the presence of the inclusion is known as the polarization stress. In an incremental FE scheme the polarization stress must be computed at each inclusion integration point, which requires the stress field of each dislocation to be evaluated at these points. This is a major computational limitation when applying the DD method to elastically inhomogeneous structures. As the number of dislocations becomes large and/or the number of inclusion elements increases, evaluation of $\hat{\mathbf{p}}$ can dominate the FE calculation. For this reason most of the DD literature has focused on homogeneous materials. This motivates our development of a new DD technique.

3 Superposition Method for DD Plasticity

The new superposition technique is shown schematically in Fig. 2. The general problem is exactly the same as in Fig. 1, however, a different decomposition is used. The desired boundary value problem is solved as the superposition of a DD subproblem subject to generic boundary conditions and a fully elastic (EL) subproblem subject to all actual boundary conditions.

The DD subproblem models only that part of the structure where dislocations are permitted to exist; the regions of elastic inhomogeneity are not modeled in the DD subproblem. The generic boundary conditions of the DD subproblem are chosen as $\mathbf{u}_{DD}=\mathbf{0}$ and $\mathbf{T}_{DD}=\mathbf{0}$ on S_u and S_t, respectively. Additionally $\mathbf{u}_{DD}=\mathbf{0}$ is prescribed on the boundary S_b between the matrix and inclusion. Thus, the only information about the full problem that is used in the solution of the DD subproblem is the geometry of the plastic region and the knowledge of whether displacement or traction boundary conditions are applied on boundaries shared by the DD subproblem and the full problem. The incremental solution of the DD subproblem is then obtained exactly as described in the previous section, i.e., as the superposition of an infinite space dislocation problem and a corrective problem. An outcome of the solution of the DD subproblem at any instant is a traction $\mathbf{T}^*$ along the boundary S_b, which is used in the EL subproblem as described below.

The EL subproblem models the entire structure and is subject to all the true boundary conditions on S. The region of the structure containing dislocations is modeled as an isotropic elastic material. Information about the plastic deformation in the plastic region of the material is transmitted to the remainder of the structure through the addition of a body force $-\mathbf{T}^*$ along S_b in the EL subproblem, which is the negative of the traction $\mathbf{T}^*$ obtained from the DD subproblem. The EL subproblem can be solved by standard FE methods. In the absence of nonlinear regions, such as a cohesive zone surface, the EL subproblem is fully linear and the FE equations can thus be solved very quickly because inversion or decomposition of the entire elastic stiffness matrix must be accomplished only once at the start of the calculation.

The solution to the full problem of interest is then obtained by superposition of the DD and EL subproblems as

$$\mathbf{u}=\mathbf{u}_{DD}+\mathbf{u}_{EL} \quad \boldsymbol{\epsilon}=\boldsymbol{\epsilon}_{DD}+\boldsymbol{\epsilon}_{EL} \quad \boldsymbol{\sigma}=\boldsymbol{\sigma}_{DD}+\boldsymbol{\sigma}_{EL} \tag{9}$$

Superposition is permitted since, in the region where superposition is being used, both problems are linearly elastic at any instant. All plasticity is completely contained in the motion and position of the dislocations within the underlying elastic material and, therefore, does not preclude the application of superposition. Although termed an ‘‘elastic’’ subproblem, the EL subproblem need only be linear in the region where the discrete dislocation superposition is being applied. That superposition yields the correct boundary conditions for the desired problem during any increment is clear from the schematic in Fig. 2: the boundary conditions are satisfied exactly, the linear field equations in each part of the problem have been solved via FEM, the tractions $\mathbf{T}^*$ and $-\mathbf{T}^*$ cancel upon superposition, and the boundary condition $\mathbf{u}_{DD}=\mathbf{0}$ has no effect on the displacement of the boundary S_b as calculated in the elastic subproblem.

The EL subproblem influences the dislocation structure and its evolution because the Peach-Koehler force is calculated on each dislocation using the full field $\boldsymbol{\sigma}$ minus the dislocation self-interaction, which thus includes the contribution $\boldsymbol{\sigma}_{EL}$. In all applications, it must be remembered that the dislocations are driven by the full, true field and not simply the fields $\boldsymbol{\sigma}_{DD}$ calculated in the DD subproblem. The expression for the Peach-Koehler force in this superposition framework is

$$f^{(I)}=n_i^{(I)}\left(\hat{\sigma}_{ij}+\sum_{J\neq I}\sigma_{ij}^{(J)}+\sigma_{ij}^{EL}\right)b_j^{(I)} \tag{10}$$

which differs from the corresponding expression for the standard formulation (8) by the inclusion of the $\boldsymbol{\sigma}_{EL}$ contribution.

Before proceeding to validate and use the new superposition method, some comments are warranted. Operationally, at each increment the new superposition method requires two FE calculations: one for the corrective fields in the DD subproblem and one for the entire EL subproblem. The overall DD subproblem is solved with standard DD machinery (described in Section 2), and as the region is elastically homogeneous, there are no polarization stresses. The new superposition method is thus advantageous in elastically inhomogeneous problems with large numbers of dislocations and/or many inclusion elements. The calculation of $\hat{\mathbf{p}}$ is eliminated at the cost of an additional FE calculation for the EL subproblem.

The DD subproblem is largely independent of the particular problem under study. Aside from adding the field $\boldsymbol{\sigma}_{EL}$ to drive the dislocations, the DD subproblem may know nothing, or only little, about the actual problem (geometry and loading) under study. If the plastic zone of the problem is constrained to occur within a finite region of space, then the DD problem can be further confined within that box, with no knowledge whatsoever about the full geometry. In this sense, the DD subproblem serves as a ‘‘black-box’’ constitutive material law for the plastic flow of the plastically deforming material. The DD subproblem is, however, limited to small strains.

Because the DD subproblem is largely disconnected from the actual problem of interest, the decomposition of the problem also provides opportunities for parallel coding for a wide variety of problems, as will be discussed further in Section 5.

4 Application to Bimaterial Interface Fracture

Crack growth in plastically deforming materials is an attractive application of the DD methodology. It is well known that in continuum plasticity the maximum opening stress ahead of the crack tip is, at most, about five times the yield strength. Such low near-tip stresses are unable to cause crack growth in many cases involving nonductile fracture modes. Furthermore, for brittlelike fracture occurring by cleavage of atomic planes, the fracture process zone is small (nanoscale), and the peak stresses required for

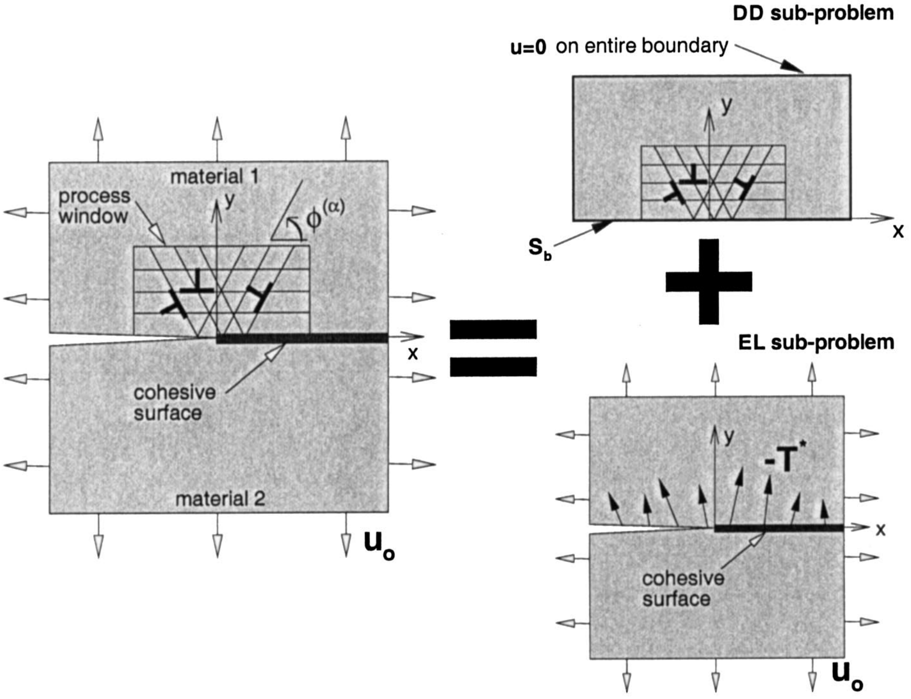


Fig. 3 Decomposition of the bimaterial fracture problem into DD and EL subproblems

material separation are large. Continuum plasticity is not expected to handle the high local stresses and stress gradients prevailing around the crack tip in such situations. Cleveringa et al. [9] thus applied the DD methodology to crack growth in a homogeneous elastic-plastic material and predicted R-curve behavior for the onset of crack growth. Their results demonstrated that dislocation organization on the scale of microns can generate the high stresses needed to grow the crack. On the other hand, the dislocation structures also generated stress fields away from the crack tip that were in generally good agreement with the continuum fields predicted by using a perfectly plastic constitutive law.

The bimaterial interface fracture problem is also suited toward treatment via the DD methodology for similar reasons. At the continuum level, the fields around a semi-infinite interface crack separating two dissimilar *elastic* half spaces were given by Rice [17]. More recently Tvergaard and Hutchinson [18–20] and Tvergaard [21–23] have explored the effects of plasticity in one or both materials. Tvergaard [23] has shown that for an elastic-plastic/elastic bimaterial, the toughness increases with increasing modulus of the purely elastic material. When the purely elastic material is at least twice as stiff as the elastic-plastic material, fracture is largely suppressed when the peak cohesive strength is roughly four times the yield stress [23]. This is expected since the cohesive strength approaches the maximum possible opening stress of the continuum model. Continuum models have been augmented by additional assumptions about the near-tip behavior. Thus Tvergaard [21,22] modeled brittle fracture with the dislocation free zone (DFZ) model of Suo et al. [24]. The DFZ model for cleavage crack growth assumes that the crack tip does not emit dislocations, stays nanoscopically sharp, and is surrounded by a thin *elastic* strip with the far-field region governed by continuum plasticity. The elastic singular field gives rise to high near tip stresses and plastic flow provides dissipation. Fracture in the elastic strip region is governed by linear elastic fracture mechanics (LEFM); crack growth is characterized by a critical stress intensity at the crack tip. The DD model, in contrast, makes no assumptions about a dislocation free region, but such a region may emerge naturally from the solution to the boundary value problem, providing insight into the size and evolution of the dislocation free region. Furthermore, fracture is an outcome of the DD solution and does not require any a priori modeling changes.

The elastic-plastic/elastic bimaterial interface problem and its decomposition using our superposition framework are shown schematically in Fig. 3. Two materials, an elastic-plastic upper half-space and an elastic lower half-space are separated by an interface that is described by a cohesive zone model (CZM). The overall material is loaded by displacements corresponding to the desired elastic K-field (see below). Far to the left, the cohesive zone is fully open and the crack surfaces are traction free. Far to the right, the cohesive zone is essentially closed and the displacements are continuous across the interface. Due to the existence of the cohesive zone, there is no crack-tip singularity in this problem. In decomposing this problem, the DD subproblem is used to model the elastic-plastic upper half plane and the EL subproblem models the entire body as an elastic/CZM problem. The boundary S_b is taken to be the entire lower boundary of the DD subproblem, i.e., the entire upper crack surface in the EL subproblem. This choice gives a smooth traction $\mathbf{T}^*$ along the entire boundary S_b that is exported to the EL subproblem. Choosing a more traditional crack boundary condition of zero traction along the open crack and zero displacement along the originally closed crack would yield a singularity in the traction $\mathbf{T}^*$ at the original crack tip in the DD subproblem. This would be difficult to resolve and handle properly in the superposition framework. The latter boundary conditions would also unnecessarily bias the entire problem toward the original crack tip.

As the bimaterial specimen is loaded, dislocations generated in the plastic material may pass out of the traction-free and/or partially open regions of the cohesive surface. This behavior is entirely physical, as the surface and cohesive zone in regions where its stiffness is becoming small, absorb the dislocations. The occurrence of this phenomenon in the new superposition technique requires comment. The DD subproblem contains all the dislocations, but subject to the boundary condition $\mathbf{u}_{DD}=\mathbf{0}$ everywhere. However, the Peach-Koehler force on a dislocation is always evaluated using the *total* fields, and hence, the dislocations move, correctly, as if in the full problem. The zero displacement boundary condition imposed in the DD subproblem is not actually seen by the dislocations. Thus, a dislocation can experience a driving force to move it out of the DD subproblem domain through the surface S_b. Physically, this leaves a step having magnitude of the Burger's vector along S_b. In terms of the DD superposition of Fig. 1, the $\tilde{\mathbf{u}}$ displacement field has a step where the dislocation exited the surface (due to the existence of the other half of the dislocation dipole that remains inside the material). Recalling Eq. (1), to satisfy the imposed $\mathbf{u}_{DD}=\mathbf{0}$ constraint on S_b, the $\hat{\mathbf{u}}$ corrective field must have an equal and opposite step at the same point on S_b. Thus, the DD subproblem of Fig. 3 is no longer smooth, and there is a singularity in the traction $\mathbf{T}^*$ at the point along S_b where the dislocation has passed through. It is not possible to accurately resolve the $\mathbf{T}^*$ singularities associated with dislocation surface steps on S_b, but an exact resolution is *not* required. If identical meshes are employed in the overlapping regions of the EL and DD subproblems, which is also desirable for other reasons, then the *approximate, nonsingular* FE traction $\mathbf{T}^*$ in the DD subproblem is exactly cancelled by the field $-\mathbf{T}^*$ applied in the EL subproblem. The field $-\mathbf{T}^*$ in the EL subproblem creates the physical surface step in the problem, resolved to the accuracy of the FE mesh. In other words, part of the traction $\mathbf{T}^*$ in the DD subproblem serves to generate an elastic field to *eliminate* the surface step, and the step is then *recreated* by the application of $-\mathbf{T}^*$ in the EL subproblem. Comparison of the new superposition technique to the standard DD method below shows that the surface steps are handled precisely via the superposition.

4.1 Model Parameters and Loading. The results presented here are based on the geometry and dislocation parameters previously used by Cleveringa et al. and Deshpande et al. [9,10] to study mode I and fatigue crack growth, respectively. In fact, we have used the actual code of Deshpande et al. [10], modified to include $\boldsymbol{\sigma}_{EL}$ from the EL subproblem to the P-K force, thus treating the entire DD code as a "black-box." We consider a specimen of 1000×1000 μm, and take the origin of an xy-coordinate system to be at the center of the sample. The initial crack tip is located at the origin and assumed to be open ($\mathbf{T}=\mathbf{0}$) for $y=0$, $x<0$ and a cohesive zone describes the interface properties for $y=0$, $x>0$. The sample is meshed with 120×220 bilinear quadrilateral elements; displacement boundary conditions correspond to an applied remote K field. The properties of the metal, i.e., the elastic-plastic upper half plane ($y>0$), are consistent with Aluminum ($E_1=70$ GPa, $\nu_1=0.33$). The elastic modulus E_2 of the ceramic ($y<0$) will be varied (with $\nu_2=0.33$). Dislocation activity in the upper half plane is restricted to a 15×15 μm "process window" of 80×80 elements, graded around the crack tip. The slip plane geometry is representative of an FCC type single crystal, with three slip systems oriented at +60°, −60°, and 0° relative to the crack plane $y=0$, spaced at $100b$ and initially dislocation-free. Only edge dislocations are considered, with Burgers vector of magnitude $b=0.25$ nm. The dislocation glide velocity is linear in the Peach-Koehler force with viscous drag coefficient $B=10^{-4}$ Pa s, and climb is not permitted. Dislocation nucleation occurs by the expansion of Frank-Read sources, randomly dispersed in the process window with density $\rho_{nuc}=66/\mu\text{m}^2$. Nucleation occurs when the Peach-Koehler force at a source exceeds a critical value of $\tau_{nuc}b$ for a time period of $t_{nuc}=10$ ns. The value of τ_{nuc} is chosen from a Gaussian distribution with mean strength $\overline{\tau}_{nuc}=50$ MPa and standard deviation $0.2\overline{\tau}_{nuc}$. Dislocations of opposite sign are annihilated when they come within a critical distance of $6b$. The process window also contains a random distribution of obstacles with density $\rho_{obs}=170/\mu\text{m}^2$ that pin dislocations until the Peach-Koehler force reaches the obstacle strength $\tau_{obs}=150$ MPa. To fully resolve the dislocation activity a time step of $\Delta t=0.5$ ns was used, which necessitates the use of a high loading rate of $\dot{K}=100\ \text{GPa m}^{1/2}/\text{s}$. Since an iterative numerical scheme is not used, the time step is not chosen from stability considerations. Instead the time step must be small enough to resolve t_{nuc}, which is the intrinsic time scale of the problem.

In the EL subproblem, we use the coupled normal-shear cohesive law relating the displacement jump $\boldsymbol{\Delta}$ across the interface to the traction $\mathbf{T}$ introduced by Xu and Needleman [25]. The tractions are obtained from a potential ϕ as $\mathbf{T}=\partial\phi/\partial\boldsymbol{\Delta}$, with

$$\phi=\phi_n+\phi_n\exp\left(-\frac{\Delta_n}{\delta_n}\right)\left\{\left[1-r+\frac{\Delta_n}{\delta_n}\right]\frac{1-q}{r-1}\right.$$
$$\left.-\left[q+\left(\frac{r-q}{r-1}\right)\frac{\Delta_n}{\delta_n}\right]\exp\left(-\frac{\Delta_t^2}{\delta_t^2}\right)\right\} \tag{11}$$

where δ_n and δ_t are the normal and tangential characteristic lengths, respectively, and the work of normal and tangential separation are

$$\phi_n=e\sigma_{\max}\delta_n, \quad \phi_t=\tau_{\max}\delta_t\sqrt{\frac{e}{2}}$$

where $\sigma_{\max}$ and $\tau_{\max}$ are the normal and shear cohesive strengths, respectively. The normal-shear coupling is included through parameters

$$q=\frac{\phi_t}{\phi_n}, \quad r=\frac{\Delta_n^*}{\delta_n}$$

where Δ_n^* is the value of Δ_n after a complete shear separation with $T_n=0$. The characteristic lengths are taken to be $\delta_n=\delta_t=0.5$ nm, and the interface strengths are $\sigma_{\max}=0.3$ GPa and $\tau_{\max}=0.699$ GPa. Then $\phi_n=\phi_t=0.408\ \text{J m}^{-2}$, giving $q=1$. The parameter r is taken to be zero.

Assuming small-scale yielding, the remote boundary conditions are characterized by the bimaterial elastic K-field. The displacement field for the upper half plane is [19]

$$u_1+\mathrm{i}u_2=\frac{|K|}{2\mu_1\cosh(\pi\epsilon)}\sqrt{\frac{R}{2\pi}}\left\{\frac{(3-4\nu_1)\mathrm{e}^{\mathrm{i}\theta/2+\epsilon(\theta-\pi)-\mathrm{i}\overline{\psi}}}{1-2\mathrm{i}\epsilon}\right.$$
$$\left.-\frac{\mathrm{e}^{-\mathrm{i}\theta/2-\epsilon(\theta-\pi)-\mathrm{i}\overline{\psi}}}{1-2\mathrm{i}\epsilon}-\mathrm{i}\sin\theta\,\mathrm{e}^{\mathrm{i}\theta/2+\epsilon(\theta+\pi)+\mathrm{i}\overline{\psi}}\right\} \tag{12}$$

where $K=K_1+\mathrm{i}K_2$, and $R=\sqrt{x^2+y^2}$. The form for the lower half plane is similar. The elastic mismatch is included through ϵ, given as

$$\epsilon=\frac{1}{2\pi}\ln\left(\frac{1-\beta}{1+\beta}\right)$$

where

$$\beta=\frac{1}{2}\,\frac{\mu_1(1-2\nu_2)-\mu_2(1-2\nu_1)}{\mu_1(1-\nu_2)+\mu_2(1-\nu_1)}$$

The mode mixity is described by a phase angle ψ such that

$$\tan\psi=\frac{\mathrm{Im}[(K_1+\mathrm{i}K_2)L^{\mathrm{i}\epsilon}]}{\mathrm{Re}[(K_1+\mathrm{i}K_2)L^{\mathrm{i}\epsilon}]} \tag{13}$$

where L is a reference length used to characterize the remote field and ψ varies with r as

$$\overline{\psi}=\psi+\epsilon\ln(r/L) \tag{14}$$

The remote loading is thus characterized by $|K|$, ψ, and L. Physically, ψ measures the ratio of shear to normal stress on the interface a distance L from the tip, as predicted by the elastic solution [19]. We take $L=10\ \mu$m, which is on the order of the process window size. Tvergaard [23] defines a reference stress intensity factor

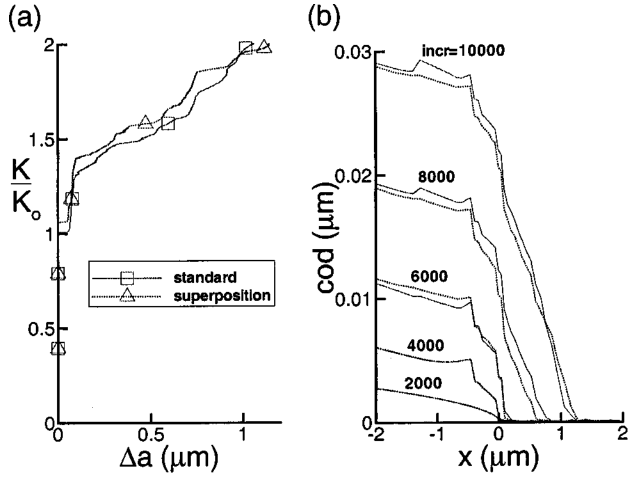


Fig. 4 Comparison of superposition and standard DD methods for rigid substrate: (*a*) crack growth resistance curves and (*b*) crack opening displacements

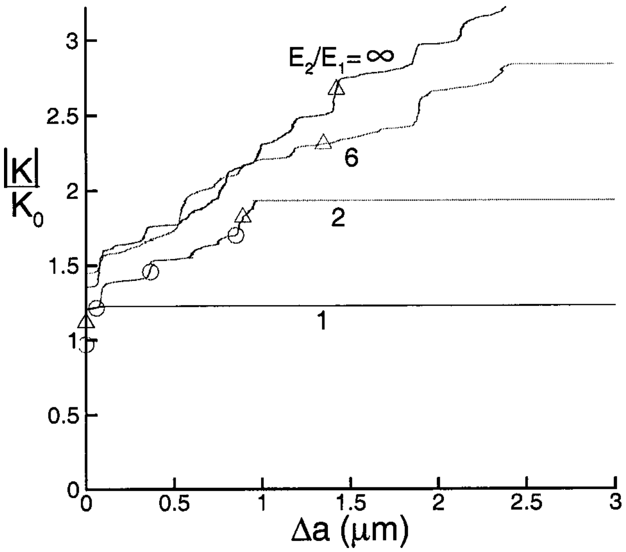


Fig. 5 Normalized applied stress intensity factor $|K|/K_0$ versus crack extension Δa for various substrate moduli

$$K_0=\left[\frac{1-\nu_1^2}{E_1}+\frac{1-\nu_2^2}{E_2}\right]^{-1/2}\left(\frac{2\Gamma_o}{1-\beta^2}\right)^{1/2} \tag{15}$$

where Γ_o is the work of separation of a mode-independent cohesive law. With $q=1$ (as is the case here), the cohesive law (11) is mode-independent, thus we set $\Gamma_o=\phi_n=\phi_t$ in (15) to calculate the reference K_0.

4.2 Validation. The superposition technique was validated by comparing to existing results obtained with the standard DD method for the problem of an elastic-plastic half-space on a rigid substrate. Use of a rigid substrate permits the DD problem to be solved only in the upper half-space, so that the standard method is appropriate and efficient. In both cases, exactly the same material was modeled, i.e., the exact same random distribution of sources and obstacles was used for both methods. Figure 4 shows the crack growth resistance curve and the evolution of the crack opening displacement with load level as obtained from the standard method and the superposition method. Excellent agreement in both quantities is obtained. Due to the chaotic nature of the discrete dislocation simulations, as discussed by Deshpande et al. [26], the fields are expected to agree within 10%. The variations in the R-curves are well within the range of chaotic effects. It is worth noting the excellent agreement in the crack surface features associated with dislocations exiting through the crack surface as the deformation proceeds. Simpler problems involving single dislocation sources are essentially exactly reproduced. These results fully validate the superposition technique and its numerical implementation. The superposition technique, including the treatment of dislocations passing through surfaces with fixed displacement boundaries of the DD subproblem, can thus be extended to the analysis of new problems with confidence.

4.3 Results. Here we present some preliminary results on the fracture toughness versus elastic mismatch for a bimaterial interface. The experimental results of Liechti and Chai [27] and the simulations of Tvergaard [23] both demonstrate a strong dependence of the R-curve on the elastic mismatch and phase angle of the applied loading, with significantly less-tough behavior found for modest ratios of elastic mismatch E_2/E_1 as compared to the rigid substrate case. Thus, the standard DD method for the rigid substrate problem is expected to greatly overpredict toughness relative to realistic metal/ceramic systems and cannot be extended to realistic systems without high computational cost.

Figure 5 shows the crack growth resistance curves for bimaterial specimens of Aluminum, as modeled above, on substrates of varying elastic modulus. The ratio of $E_2/E_1=6$ corresponds roughly to Aluminum on SiC, while the ratio $E_2/E_1=2$ corresponds to Aluminum on Si. With increasing stiffness ratio, the bimaterial interface shows a rapidly increasing toughness. For the $E_2/E_1=\infty$ specimen the simulation was stopped before a failure point was reached. Figure 6 shows the opening stress σ_{22} and dislocation positions at a load near the failure point for $E_2/E_1=1$, 2, and 6, and at a high load for the $E_2/E_1=\infty$ specimen. When $E_2/E_1=1$ relatively few dislocations are nucleated, the crack remains sharp, and fracture occurs in an almost entirely brittle manner, with negligible toughening. That failure occurs at $|K|/K_0>1$ is due to the neglect of the shear toughness of the interface in choosing the reference K_0 [see comments after Eq. (15)]. The other R-curves all exhibit regimes of toughening and spurts of brittle crack extension. The periods where a large increase in $|K|$ occurs over a small Δa correspond to the crack tip field repeatedly activating one (or several) nearby sources. Crack blunting follows as one dislocation of a dipole pair passes out of the crack surface, while the other slips away from the crack surface into the material. Eventually the dislocations in the material provide a backstress sufficient to shut down the source(s) or a near-tip dislocation configuration generates high stresses resulting in a spurt of crack growth. This semi-brittle extension occurs with a relatively small increase in applied $|K|$. This brittle extension is either stopped when the new crack tip field activates other nearby sources, or if sufficiently weak sources are not nearby and/or the crack is growing very rapidly the extension continues and the specimen fails. When comparing the R-curves of Fig. 5 it is worth noting that since the applied loading was defined by $\psi=0$ at $L=10\,\mu$m the variation in substrate modulus results in varying amounts of shear closer to the crack tip. The variation of phase along the interface is particular to the inhomogeneous crack problem, and such mixed mode effects will be examined in detail in a future publication.

The evolution of the opening stress and dislocation structure for the $E_2/E_1=2$ specimen are shown in Fig. 7. In Fig. 7(*a*) at $|K|/K_0=0.966$ the crack tip is sharp and very few dislocations have been nucleated. When $|K|/K_0=1.208$ the majority of the slip is occurring on two slip planes with $+60°$ orientation. As one dipole slips into the material, the other glides toward the crack tip, passes out, and blunts the crack. As loading continues, additional sources become activated; at $|K|/K_0=1.449$ one or two $-60°$ slip planes have a cluster of dislocations near the initial crack tip. At $|K|/K_0=1.691$ the crack has extended roughly 0.85 μm and new

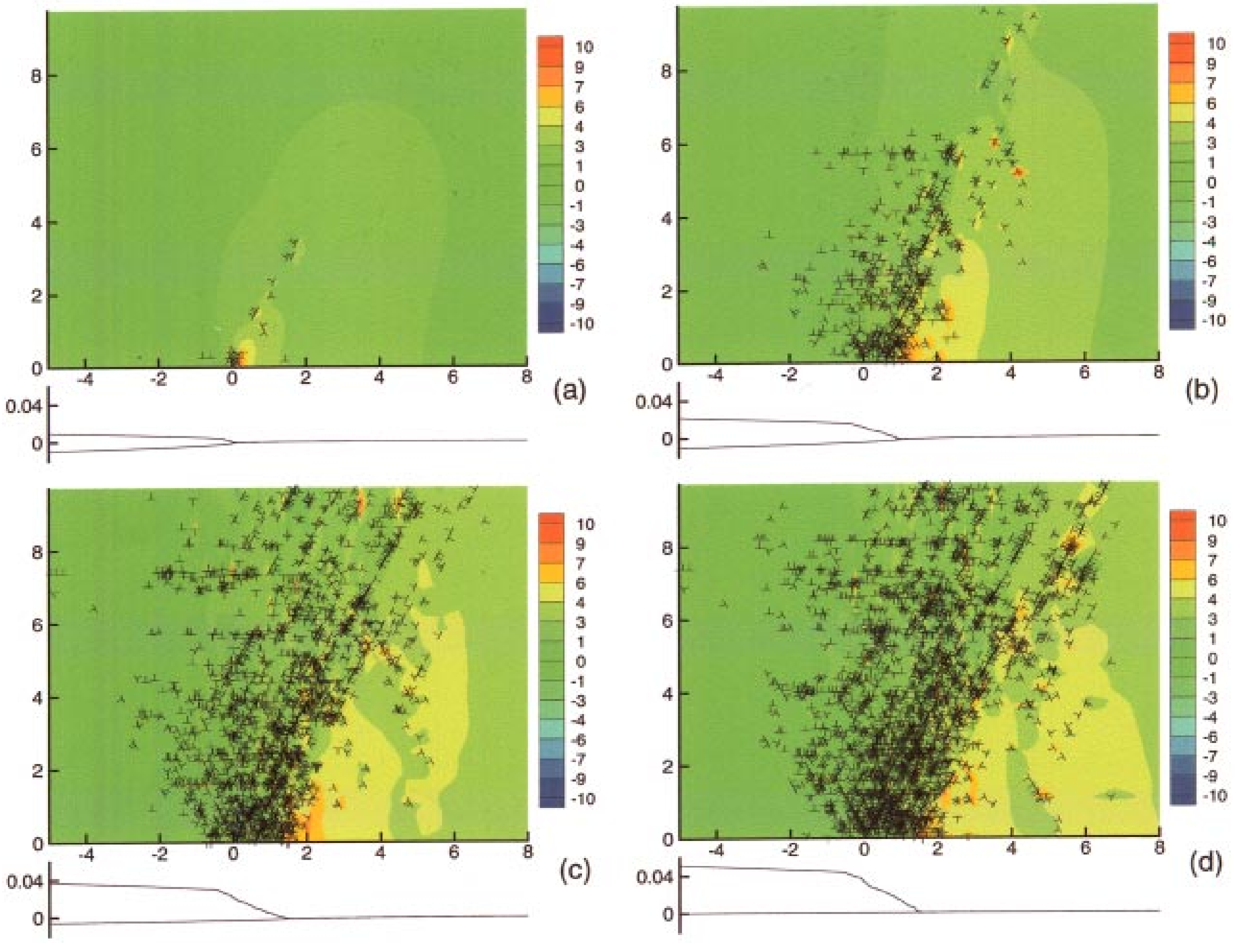


Fig. 6 (Color online) **Dislocations and normalized opening stress $\sigma_{22}/\mu \times 10^3$ in a 10×13 μm region near the crack tip for various substrate moduli just prior to failure** (**see triangles in Fig. 5**): (**a**) $E_2/E_1=1$ **with** $|K|/K_0=1.118$, (**b**) $E_2/E_1=2$ **with** $|K|/K_0=1.811$, (**c**) $E_2/E_1=6$ **with** $|K|/K_0=2.304$, **and** (**d**) $E_2/E_1=\infty$ **with** $|K|/K_0=2.286$. **The crack opening profiles for each case are plotted below the** x **axis. All distances are in microns.**

sources are activated. A snapshot just before failure of this specimen is shown in Fig. 6(*b*) where $|K|/K_0=1.811$.

The model DD parameters used for Aluminum generate a yield stress of about 60 MPa [28]. As discussed above, the normal cohesive strength is 300 MPa, thus $\sigma_{\max}/\sigma_y=5$. In the continuum models of Tvergaard [23], such a large ratio of cohesive strength to yield stress essentially precludes crack growth, except when the substrate is very soft $(E_2/E_1=1)$. In the DD model, however, brittle fracture is reached in all cases except that for the rigid substrate, as seen in Fig. 5. This highlights the dual role of dislocations in elastic-plastic fracture: the motion and interaction of dislocations provides dissipation and increased toughness, but the dislocations also generate local stresses reaching the cohesive strength and thus driving crack growth. The latter effect is missing in standard continuum simulations. Qualitatively, however, the R-curves found here do agree with those found by Tvergaard at a lower ratio of $\sigma_{\max}/\sigma_y=3$ (Fig. 4. in [23]).

The trend of increasing toughness with increasing substrate modulus can be justified by the following qualitative analysis. The work of the applied loading is apportioned between the elastic deformations of the substrate and plastic material and the dissipation due to plastic flow. The plastic flow itself is driven by the stresses, and hence stored work, in the plastic material. Thus, with increasing rigidity of the substrate, more work is put into the plastic material and a larger fraction of that work is dissipated by plastic flow. In the rigid limit, all of the deformation and stored elastic energy resides in the plastic material, and the deformation drives a high degree of plastic dissipation and, hence, a high overall toughness. As the substrate becomes very soft, all of the energy is stored in the deformation of the substrate, preempting plastic flow and dissipation in the plastic material and causing completely brittle fracture.

The preliminary DD results can provide insight into the relevance or necessity of the dislocation free zone concept to bimaterials. Excluding the $E_2/E_1=1$ specimen, which has such few dislocations that it exhibits almost completely brittle fracture, there is no consistent observation of a DFZ throughout the load history. When $E_2/E_1=2$ a strip approximately 0.1 μm from the *free* crack surface does not contain dislocations (see Fig. 7). This is expected as nearby dislocations (on a ±60° plane) will be attracted to the free surface and pass out. However, for the $E_2/E_1=2$, 6, ∞ ratios, particularly at elevated loading, a DFZ is not observed down to the resolution of this analysis (see Fig. 7); the minimum near-tip element size is 0.05 μm and the intrinsic cohesive length scale $\delta_c(=\sigma_{\max}\delta_n/E)$ for the problem is 0.11 μm and so "crack-tip" phenomena are not resolved below this scale. In addition, the crack profiles in Figs. 6 and 7 all exhibit significant blunting before failure, which is inconsistent with the DFZ assumption of a nanoscopically sharp crack. However, the preliminary DD results suggest that crack growth is due primarily to high near-tip stresses, resulting from the applied load and near-tip dislocation structures, and not the blunting of the crack as the cohe-

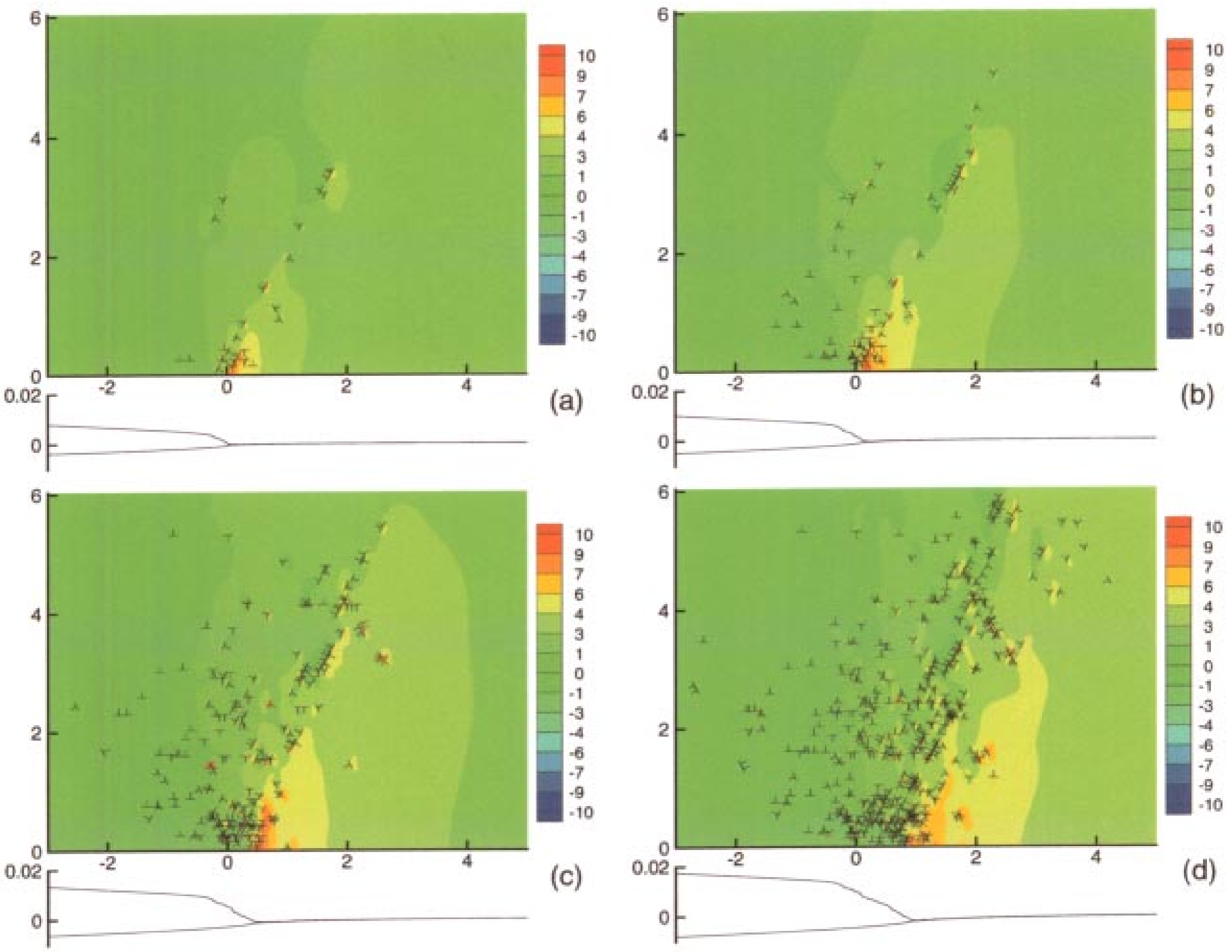


Fig. 7 (Color online) Dislocations and normalized opening stress $\sigma_{22}/\mu \times 10^3$ in a 8×6 μm region near the crack tip for $E_2/E_1 = 2$ at four stages of loading (see circles in Fig. 5): (a) $|K|/K_0 = 0.966$, (b) 1.208, (c) 1.449, and (d) 1.691. The crack opening profiles at each load are plotted below the x axis. All distances are in microns.

sive zone absorbs dislocations. The results here also support the observation of Wei and Hutchinson [29] that even when a DFZ is appropriate, it may be so small that a nonlocal plasticity theory is required to match to the large gradients present in the elastic DFZ strip.

This represents a preliminary overview of the elastic-plastic/elastic bimaterial problem. A much more detailed analysis will be performed in a future publication, including examination of the influences of i) cohesive law details (e.g., shear versus normal strengths and shear/normal coupling); ii) the ratio of the cohesive strength to yield stress; and iii) mode mixity of the remote loading, an analysis of the underlying dislocation structures and their relationships to observed trends, and a study of statistical effects.

5 Discussion and Conclusions

We have presented a general new technique for the efficient extension of the DD plasticity method to problems with elastic inhomogeneities. The usefulness of the method lies in its isolation of the DD part of the model from many of the particular features of the problem under consideration. The DD calculation only requires information about the geometry of the region where dislocations may be present, and is supplied, via the superposition, with driving forces on the dislocations coming from the elastic subproblem. This construction effectively allows for the use DD plasticity as a ‘‘black-box’’ constitutive input in any desired region.

The existing DD method involves continuum representations of many atomistic phenomena, including nucleation, glide, pinning, and annihilation. Work is currently in progress on incorporating additional physical features of dislocation behavior (i.e., junctions, source generation, and stage II hardening) into the 2D formulation [30]. The superposition technique can assimilate any such changes without difficulty because it is unaffected by any of the inner workings of the DD method.

The superposition method can also be used to handle problems containing multiple plastic domains. In such problems, each plastic domain requires its own DD calculation and generates tractions $\mathbf{T}^*$ on the relevant surfaces S_b that are incorporated into the single EL subproblem. Upon solution of the full EL subproblem, the stress fields are exported back to the DD calculations in each domain. Some physical applications of this approach are to metal/ceramic multilayers, polycrystalline structures (each grain treated separately, see Fig. 8), and layered metal/metal structures where variations in elastic and plastic properties exist from layer to layer. For the polycrystalline structure of Fig. 8 a separate DD subproblem is constructed for each grain. The tractions T^* in the EL subproblem are calculated at the appropriate boundaries in the DD subproblem; boundaries common to two grains will have T^* contributions from each grain.

The superposition method also serves as a basis for parallel computations. The DD problem(s) can be divided into any number of subregions, either for physical or mathematical convenience. Each subvolume DD calculation requires only the calculation of

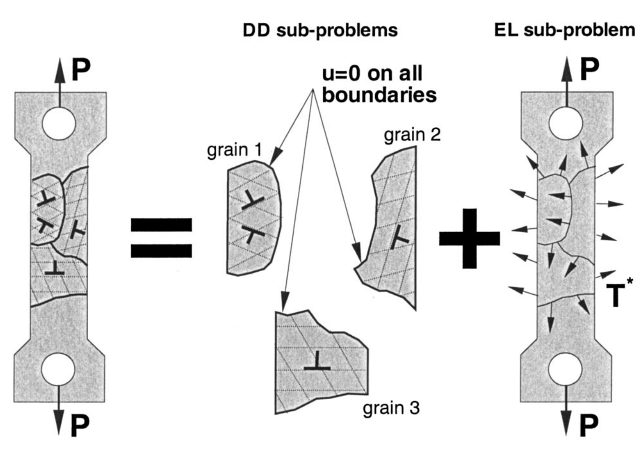


Fig. 8 Schematic of the superposition technique applied to a 2D polycrystalline structure. Each DD problem is independent and thus the computation is easily parallelized.

dislocation/dislocation interactions within that volume. All other dislocation interactions are fully accounted for through the external (EL sub-problem) fields and exported tractions. Since each subvolume DD calculation is self-contained, these calculations can be performed independently and simultaneously on separate processors. For instance, for the polycrystalline model in Fig. 8 each grain can be handled simultaneously on a separate processor. Dislocations passing from one subvolume to another are handled seamlessly just as we have treated dislocations passing through a cohesive surface in the bimaterial problem.

In conclusion, the DD model allows, in principle, for the study of numerous problems where the characteristic dimensions are on the order of tens of microns, at which scales continuum plasticity is unable to reproduce observed size effects. The new DD superposition technique developed here extends the practical range of application of the DD method to problems with elastic inhomogeneities, multiple plastic domains, and larger physical sizes. Preliminary application to bimaterial interface cracking shows the power and generality of the approach. Future work will address the detailed analysis of crack growth along bimaterial interfaces as well as other problems to which the general DD approach is well suited.

Acknowledgments

The authors gratefully acknowledge the support of this work by U.S. Air Force Office of Scientific Research through the MURI program "Virtual design and testing on materials: A multiscale approach," Grant No. F49620-99-1-0272, and by General Motors through the GM/Brown Collaborative Research Laboratory on Computational Materials Research. The authors thank Professor Alan Needleman and Professor Vikram Deshpande for many useful discussions pertaining to this work and for providing the DD code used here.

Appendix: Static Condensation of FE Equations

This appendix describes the implementation of static condensation, which provides for an optimally efficient solution of the EL finite element equations. The incremental FE equations are derived from an expansion of the principle of virtual work, then the elastic degrees-of-freedom (dof) are condensed out, leaving a system of equations for only the cohesive dof. This reduces the size of the system of equations to be solved by two orders of magnitude for the present bimaterial model.

In the superposition method, two finite element solutions are required at each increment for i) the corrective fields of the DD sub-problem and ii) the fields of the EL subproblem. The system of equations for the DD corrective fields is completely linear as the cohesive zone is included only in the EL subproblem. The entire stiffness matrix is constant, can be decomposed once, and stored. The incremental solution for the corrective fields is thus very efficient.

Standard equation solution techniques are, however, largely inefficient for the EL subproblem, which has two linear elastic regions joined by a nonlinear cohesive zone. When only a small subset of the global stiffness matrix is nonlinear (those associated with the cohesive dof), partial decomposition can provide a fast incremental solution. This allows the stiffness decomposition *up to* the first nonlinear dof to be computed once and stored (as it is invariant). In all future increments only the remainder of the stiffness matrix must be decomposed. Clearly for a fixed number of dof, the closer to the end of the matrix the first nonlinear dof appears, the greater the benefit of partial decomposition. In the EL subproblem, optimum node numbering (left to right, top to bottom) places the nonlinear dof in the middle of the unknown nodal dof vector. Even using partial decomposition, over half of the global stiffness matrix must be decomposed during each increment. Due to the large model size (each elastic region has over 20,000 dof) and number of increments required, this is not practical. Instead we use static condensation [31] to obtain a system of equations for only the nonlinear cohesive dof. Once this system is solved, the remaining elastic dof are obtained by matrix multiplication with an appropriate numerical Green's function. This reduces the size of the system of equations to be solved by two orders of magnitude, greatly increasing the efficiency of the EL subproblem incremental solution.

The FE equations are developed from the minimization of potential energy. At time t an elastic body with a cohesive surface S_c, subject to only displacement boundary conditions, has potential energy

$$\Pi^{(t)}=\frac{1}{2}\int_V \sigma_{ij}^{(t)}\epsilon_{ij}\mathrm{d}V-\int_{S_c}T_i^{(t)}u_i\mathrm{d}S_c \tag{A1}$$

The discretizations of the displacement, velocity, and strain fields are

$$u_i=N_{iI}U_I,\quad \dot{u}_i=v_i=N_{iI}\dot{U}_I=N_{iI}V_I,\quad \epsilon_{ij}=u_{i,j}=N_{i,jI}U_I \tag{A2}$$

where U_I and V_I are nodal displacements and velocities, respectively, N_{iI} are the shape functions, and capital indices refer to nodal quantities. Using this discretization in Eq. ($A1$) gives

$$\Pi^{(t)}=\tfrac{1}{2}K_{IJ}U_IU_J-F_I^cU_I^c \tag{A3}$$

where the global stiffness matrix is $K_{IJ}=\int_V N_{i,jI}C_{ijkl}N_{k,lJ}\mathrm{d}V$, and the global cohesive nodal force vector is $F_I^c=\int_{S_c}T_iN_{iI}\mathrm{d}S_c$.

The incremental equations are obtained by expanding the potential energy at time $t+\Delta t$ about an equilibrium state at time t

$$\begin{aligned}\Pi^{(t+\Delta t)}=\Pi^{(t)}+\Delta t\dot{\Pi}=&\tfrac{1}{2}\int_V\sigma_{ij}\epsilon_{ij}\mathrm{d}V-\int_{S_c}T_iu_i\mathrm{d}S_c\\&+\Delta t\left[\tfrac{1}{2}\int_V\dot{\sigma}_{ij}\epsilon_{ij}\mathrm{d}V+\tfrac{1}{2}\int_V\sigma_{ij}\dot{\epsilon}_{ij}\mathrm{d}V-\int_{S_c}\dot{T}_iu_i\mathrm{d}S_c\right.\\&\left.-\int_{S_c}T_i\dot{u}_i\mathrm{d}S_c\right]\end{aligned} \tag{A4}$$

Rewriting the rate of change of cohesive traction as $\dot{T}_i=\mathrm{d}T_i/\mathrm{d}t=(\mathrm{d}T_i/\mathrm{d}u_j)(\mathrm{d}u_j/\mathrm{d}t)=-k_{ij}^c\dot{u}_j$ allows the cohesive stiffness contribution to be written as $K_{IJ}^{cc}=\int_{S_c}N_{iI}k_{ij}^cN_{iJ}\mathrm{d}S_c$. Using the discretization ($A2$), Eq. ($A4$) becomes

$$\Pi^{(t+\Delta t)}=\tfrac{1}{2}K_{IJ}U_IU_J-F_I^cU_I^c+\Delta t[K_{IJ}V_JU_I+K_{IJ}^{cc}V_J^cU_I^c-F_I^cV_I^c] \tag{A5}$$

The standard FE equations could be obtained at this point by minimizing potential energy with respect to nodal displacements, i.e., $\mathrm{d}\Pi^{(t+\Delta t)}/\mathrm{d}U_K=0$. Here, we use static condensation, which is based on a careful labeling of nodal vectors and matrices according to dof being elastic (i.e., linear, not on the cohesive zone) or cohesive (nonlinear). Then the elastic dof will be condensed out, so that we are left with a system of equations for only the nonlinear cohesive dof.

The total number of dof n is composed of elastic and cohesive dof n_e and n_c, respectively. Any dof can also be classified as belonging to material 1 (upper) or material 2 (lower). Using this notation, the global displacement vector $\mathbf{U}$ can be written as a composition of subvectors

$$\mathbf{U}=\begin{bmatrix}\mathbf{U}^{e(1)}\\ \mathbf{U}^{c(1)}\\ \mathbf{U}^{c(2)}\\ \mathbf{U}^{e(2)}\end{bmatrix},\quad \mathbf{U}^{e}=\begin{bmatrix}\mathbf{U}^{e(1)}\\ \mathbf{U}^{e(2)}\end{bmatrix},\quad \mathbf{U}^{c}=\begin{bmatrix}\mathbf{U}^{c(1)}\\ \mathbf{U}^{c(2)}\end{bmatrix} \tag{A6}$$

where $\mathbf{U}$, or any vector, is split into "local" vectors $\mathbf{U}^e$ and $\mathbf{U}^c$ formed by the union of subvectors as shown and having respective dimensions of n_e and n_c.

The global stiffness matrix can be similarly decomposed into submatrices as

$$\mathbf{K}=\begin{bmatrix}\mathbf{K}^{ee(1)} & \mathbf{K}^{ec(1)} & \mathbf{0}\\ \mathbf{K}^{ce(1)} & \mathbf{K}^{cc} & \mathbf{K}^{ce(2)}\\ \mathbf{0} & \mathbf{K}^{ce(2)} & \mathbf{K}^{ee(2)}\end{bmatrix} \tag{A7}$$

The relevant local matrices are defined as

$$\mathbf{K}^{ee}=\begin{bmatrix}\mathbf{K}^{ee(1)} & \mathbf{0}\\ \mathbf{0} & \mathbf{K}^{ee(2)}\end{bmatrix},\quad \mathbf{K}^{ec}=\begin{bmatrix}\mathbf{K}^{ec(1)}\\ \mathbf{K}^{ec(2)}\end{bmatrix},$$

$$\mathbf{K}^{ce}=[\mathbf{K}^{ce(1)}\ \ \mathbf{K}^{ce(2)}] \tag{A8}$$

Note that when indicial notation is used to represent a local array, the superscripts denote the range of the indices, e.g., K_{IJ}^{ec} has dimensions $n_e\times n_c$. The cohesive stiffness $\check{K}_{IJ}^{cc}$ necessarily has contributions from the cohesive zone *and* the elastic material. These contributions can be separated as

$$\check{K}_{IJ}^{cc}=K_{IJ}^{cc}+\overline{K}_{IJ}^{cc} \tag{A9}$$

where $\overline{K}_{IJ}^{cc}=\int_V N_{i,jI}C_{ijkl}N_{k,lJ}\mathrm{d}V$ is the elastic stiffness contribution to cohesive dof. Local matrices $\mathbf{K}^{ec}$ and $\mathbf{K}^{ce}$ couple elastic and cohesive dof, by construction $\mathbf{K}^{ce}=(\mathbf{K}^{ec})^T$.

The only stiffness submatrix that changes during the incremental procedure is $\check{K}_{IJ}^{cc}$, and all of its variations are contained in K_{IJ}^{cc}. All other stiffness submatrices are strictly a function of elastic constants and undeformed geometry (mesh) and are thus invariant. Rewriting Eq. (A5) in local form gives

$$\begin{aligned}\Pi^{(t+\Delta t)}={}&\tfrac{1}{2}K_{IJ}^{ee}U_I^eU_J^e+K_{IJ}^{ec}U_J^cU_I^e+\tfrac{1}{2}\overline{K}_{IJ}^{cc}U_I^cU_J^c-F_I^cU_I^c\\&+\Delta t[K_{IJ}^{ee}U_I^eV_J^e+K_{IJ}^{ce}V_J^eU_I^c+K_{JI}^{ce}V_J^cU_I^e+\check{K}_{IJ}^{cc}V_J^cU_I^c\\&-F_I^cV_I^c]\end{aligned} \tag{A10}$$

The FE equations are derived by applying the *two* equilibrium conditions, $\mathrm{d}\Pi/\mathrm{d}U_K^e=0$ and $\mathrm{d}\Pi/\mathrm{d}U_K^c=0$, giving

$$K_{IJ}^{ee}V_J^e=-K_{IJ}^{ec}V_J^c-\frac{1}{\Delta t}[K_{IJ}^{ee}U_J^e+K_{IJ}^{ec}U_J^c] \tag{A11}$$

$$\check{K}_{IJ}^{cc}V_J^c=-K_{IJ}^{ce}V_J^e-\frac{1}{\Delta t}[\overline{K}_{IJ}^{cc}U_J^c+K_{IJ}^{ce}U_J^e-F_I^c] \tag{A12}$$

At this point, it is necessary to treat constrained and unconstrained dof separately. Thus all elastic (cohesive) dof $n_e\,(n_c)$ are classified as either constrained $n_{ec}\,(n_{cc})$ or unconstrained $n_{ef}\,(n_{cf})$. For the remainder of this section all implicit summations span only *unconstrained* dof, while summations over constrained dof are written explicitly. The implementation of this convention is best illustrated with a brief example. Consider the sums $K_{IJ}^{ec}V_J^c$ and $\Sigma_{n_{cc}}K_{IJ}^{ec}V_J^c$, which according to our convention span unconstrained and constrained cohesive dof, respectively. The key point is that the matrix K_{IJ}^{ec} (vector V_J^c) has size $n_e\times n_c\,(n_c)$ and contains constrained and unconstrained dof, which are interspersed irregularly. When performing the sum (over unconstrained dof) $K_{IJ}^{ec}V_J^c$, we must inquire at every J as to the existence of a constraint on that particular dof. Only if J is an unconstrained dof is the multiplication performed and the product added to the total for a given I; if J is a constrained dof, then it is skipped over and does not contribute to the sum. It is essential to recall this convention in the following formulation.

The equilibrium Eqs. (A11) and (A12) are rewritten as

$$K_{IJ}^{ee}V_J^e=-K_{IJ}^{ec}V_J^c-\frac{1}{\Delta t}T_I^e-R_I^e \tag{A13}$$

$$\check{K}_{IJ}^{cc}V_J^c=-K_{JI}^{ec}V_J^e-\frac{1}{\Delta t}T_I^c-R_I^c \tag{A14}$$

where R_I is a nodal force vector and T_I is an equilibrium correction, specifically,

$$R_I^e=\sum_{n_{ec}}K_{IJ}^{ee}V_J^e+\sum_{n_{cc}}K_{IJ}^{ec}V_J^c \tag{A15}$$

$$T_I^e=K_{IJ}^{ee}U_J^e+K_{IJ}^{ec}U_J^c+\sum_{n_{ec}}K_{IJ}^{ee}U_J^e+\sum_{n_{cc}}K_{IJ}^{ec}U_J^c \tag{A16}$$

$$R_I^c=\sum_{n_{ec}}K_{JI}^{ec}V_J^e+\sum_{n_{cc}}\check{K}_{IJ}^{cc}V_J^c, \tag{A17}$$

$$T_I^c=\overline{K}_{IJ}^{cc}U_J^c+K_{JI}^{ec}U_J^e-F_I^c+\sum_{n_{ec}}K_{JI}^{ec}U_J^e+\sum_{n_{cc}}\overline{K}_{IJ}^{cc}U_J^c \tag{A18}$$

Now Eq. (A-13) can be solved for V_J^e in terms of V_J^c as

$$V_K^e=-(K_{IK}^{ee})^{-1}K_{IJ}^{ec}V_J^c-\frac{1}{\Delta t}(K_{IK}^{ee})^{-1}T_I^e-(K_{IK}^{ee})^{-1}R_I^e \tag{A19}$$

which is used in Eq. (A14) to provide the system of equations for the unconstrained cohesive dof V_J^c

$$\begin{aligned}[\check{K}_{IJ}^{cc}-K_{KI}^{ec}(K_{LK}^{ee})^{-1}K_{LJ}^{ec}]V_J^c={}&\frac{1}{\Delta t}[K_{KI}^{ec}(K_{LK}^{ee})^{-1}T_L^e-T_I^c]\\&+K_{KI}^{ec}(K_{LK}^{ee})^{-1}R_L^e-R_I^c\end{aligned} \tag{A20}$$

where the term in brackets on the left-hand side is an effective stiffness matrix $(n_c\times n_c)$ and the right-hand side is an effective nodal force vector (n_c). Both are known at the beginning of each increment, so the system of Eqs. (A20) can be solved for V_J^c. The unknown elastic velocity vector V_J^e is obtained by matrix multiplication from Eq. (A13)

$$V_K^e=-(K_{IK}^{ee})^{-1}K_{IJ}^{ec}V_J^c-\frac{1}{\Delta t}(K_{IK}^{ee})^{-1}T_I^e-(K_{IK}^{ee})^{-1}R_I^e \tag{A21}$$

The elastic dof have been condensed out of the system of Eqs. (A20) to be solved. Equation (A21) involves the numerical Green's function, which, given the cohesive velocity vector V_J^c, outputs the elastic velocity vector V_K^e. An obvious drawback is the fact that the inverse of the elastic stiffness matrix $(K_{LK}^{ee})^{-1}$ must be computed before any incremental solution can begin. This is a huge (symmetric) matrix, but not banded. Before the incremental scheme begins, the elastic stiffness matrix K_{IJ}^{ee} must be inverted, which is significant overhead. Once this has been done, much of the matrix multiplication, i.e., $K^{ec}_{KI}(K^{ee}_{LK})^{-1}K^{ec}_{LJ}$, $K^{ec}_{KI}(K^{ee}_{LK})^{-1}$, and $(K^{ee}_{IK})^{-1}K^{ec}_{IJ}$ in Eqs. $(A20)$ and $(A21)$ can be done once and stored. This allows the incremental solution to proceed with optimal efficiency. The static condensation was verified independently of the DD machinery and was in exact agreement with the standard FE method.

References

[1] Stolken, J. S., and Evans, A. G., 1998, ''A Microbend Test Method for Measuring the Plasticity Length Scale,'' Acta Mater., **46**, pp. 5109–5115.

[2] Fleck, N. A., Muller, G. M., Ashby, M. F., and Hutchinson, J. W., 1994, ''Strain Gradient Plasticity: Theory and Experiment,'' Acta Metall. Mater., **42**, pp. 475–487.

[3] Fleck, N. A., and Hutchinson, J. W., 1997, ''Strain Gradient Plasticity,'' J. W. Hutchinson and T. T. Wu, eds., Adv. Appl. Mech., **33**, pp. 295–361.

[4] Gurtin, M. E., 2002, ''A Gradient Theory of Single-Crystal Viscoplasticity That Accounts for Geometrically Necessary Dislocations,'' J. Mech. Phys. Solids, **50**, pp. 5–32.

[5] Acharya, A., and Bassani, J. L., 2000, ''Lattice Incompatibility and a Gradient Theory of Crystal Plasticity,'' J. Mech. Phys. Solids, **48**, pp. 1565–1595.

[6] Shu, J. Y., Fleck, N. A., Van der Giessen, E., and Needleman, A., 2001, ''Boundary Layers in Con-Strained Plastic Flow: Comparison of Nonlocal and Discrete Dislocation Plasticity,'' J. Mech. Phys. Solids, **49**, pp. 1361–1395.

[7] Van der Giessen, E., and Needleman, A., 1995, ''Discrete Dislocation Plasticity: A Simple Planar Model,'' Modell. Simul. Mater. Sci. Eng., **3**, pp. 689–735.

[8] Cleveringa, H. H. M., Van der Giessen, E., and Needleman, A., 1997, ''Comparison of Discrete Dislocation and Continuum Plasticity Predictions for a Composite Material,'' Acta Mater., **45**, pp. 3163–3179.

[9] Cleveringa, H. H. M., Van der Giessen, E., and Needleman, A., 2000, ''A Discrete Dislocation Analysis of Mode I Crack Growth,'' J. Mech. Phys. Solids, **48**, pp. 1133–1157.

[10] Deshpande, V. S., Needleman, A., and Van der Giessen, E., 2002, ''Discrete Dislocation Modeling of Fatigue Crack Propagation,'' Acta Mater., **50**, pp. 831–846.

[11] Benzerga, A. A., Hong, S. S., Kim, K. S., Needleman, A., and Van der Giessen, E., 2001, ''Smaller is Softer: An Inverse Size Effect in a Cast Aluminum Alloy,'' Acta Mater., **49**, pp. 3071–3083.

[12] Bittencourt, E., Needleman, A., Gurtin, M. E., and Van der Giessen, E., 2003, ''A Comparison of Nonlocal Continuum and Discrete Dislocation Plasticity Predictions,'' J. Mech. Phys. Solids, **51**, pp. 281–310.

[13] Eshelby, J. D., 1961, ''Elastic Inclusions and Inhomogeneities,'' *Progress in Solid Mechanics, Vol. II*, I. N. Sneddon and R. Hill, eds., North-Holland, Amsterdam, 89–140.

[14] Shilkrot, L. E., Curtin, W. A., and Miller, R. E., 2002, ''A Coupled Atomistic/Continuum Model of Defects in Solids,'' J. Mech. Phys. Solids, **50**, pp. 2085–2106.

[15] Weygand, D., Friedman, L. H., Van der Giessen, E., and Needleman, A., 2002, ''Aspects of Boundary-Value Problem Solutions With Three-Dimensional Dislocation Dynamics,'' Modell. Simul. Mater. Sci. Eng., **10**, pp. 437–468.

[16] Hirth, J. P., and Lothe, J., 1968, *Theory of Dislocations*, McGraw-Hill, New York.

[17] Rice, J. R., 1988, ''Elastic Fracture Mechanics Concepts for Interfacial Cracks,'' ASME J. Appl. Mech., **55**, pp. 98–103.

[18] Tvergaard, V., and Hutchinson, J. W., 1992, ''The Relation Between Crack Growth Resistance and Fracture Process Parameters in Elastic-Plastic Solids,'' J. Mech. Phys. Solids, **40**, pp. 1377–1397.

[19] Tvergaard, V., and Hutchinson, J. W., 1993, ''The Influence of Plasticity on Mixed Mode Interface Toughness,'' J. Mech. Phys. Solids, **41**, pp. 1119–1135.

[20] Tvergaard, V., and Hutchinson, J. W., 1994, ''Effect of T-Stress on Mode I Crack Growth Resistance in a Ductile Solid,'' Int. J. Solids Struct., **31**, 823–833.

[21] Tvergaard, V., 1997, ''Cleavage Crack Growth Resistance Due to Plastic Flow Around a Near-Tip Dislocation-Free Region,'' J. Mech. Phys. Solids, **45**, pp. 1007–1023.

[22] Tvergaard, V., 1999, ''Effect of Plasticity on Cleavage Crack Growth Resistance at an Interface,'' J. Mech. Phys. Solids, **47**, pp. 1095–1112.

[23] Tvergaard, V., 2001, ''Resistance Curves for Mixed Mode Interface Crack Growth Between Dissimilar Elastic-Plastic Solids,'' J. Mech. Phys. Solids, **49**, pp. 2689–2703.

[24] Suo, Z., Shih, C. F., and Varias, A. G., 1993, ''A Theory for Cleavage Cracking in the Presence of Plastic Flow,'' Acta Metall. Mater., **41**, pp. 1551–1557.

[25] Xu, X.-P., and Needleman, A., 1993, ''Void Nucleation by Inclusion Debonding in a Crystal Matrix,'' Modell. Simul. Mater. Sci. Eng., **1**, pp. 111–132.

[26] Deshpande, V. S., Needleman, A., and Van der Giessen, E., 2001, ''Dislocation Dynamics is Chaotic,'' Scr. Mater., **45**, pp. 1047–1053.

[27] Liechti, K. M., and Chai, Y. S., 1992, ''Asymmetric Shielding in Interfacial Fracture Under In-Plane Shear,'' ASME J. Appl. Mech., **59**, pp. 295–304.

[28] Deshpande, V. S., Needleman, A., and Van der Giessen, E., 2003, ''Discrete Dislocation Plasticity Modeling of Short Cracks in Single Crystals,'' Acta Mater., **51**, 1–15.

[29] Wei, Y., and Hutchinson, J. W., 1997, ''Steady-State Crack Growth and Work of Fracture for Solids Characterized by Strain Gradient Plasticity,'' J. Mech. Phys. Solids, **45**, pp. 1253–1273.

[30] Benzerga, A. A., and Needleman, A., work in progress.

[31] Bathe, K.-J., 1982, *Finite Element Procedures in Engineering Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

End Effects in Prestrained Plates Under Compression

B. Karp
Department of Mechanical Engineering,
Ben-Gurion University of the Negev,
P.O.B. 653,
Beer-Sheva 84105, Israel

The decay of end perturbations imposed on a rectangular plate subjected to compression is investigated in the context of plane-strain incremental finite elasticity. A separation of variables in the eigenfunction formulation is used for the perturbed field within the plate. Numerical results for the leading decay exponent are given for four rubbers: three compressible and one incompressible. It was found that the lowest decay rate is governed by a symmetric field that exhibits different patterns of dependence on the prestrain for compressible and for nearly incompressible solids. Compressible solids are characterized by low sensitivity of the decay rate to prestrain level up to moderate compression, beyond which an abrupt decrease of decay rate brings it to zero. Nearly incompressible solids, on the other hand, expose a different pattern involving interchange of modes with no decrease of decay rate to zero. Both patterns show that the decay rate obtained from linear elastic analysis can be considered as a good approximation for a prebuckled, slightly compressed plate, which is long enough in comparison to its width. Along with decaying modes, the eigenfunction expansion generates a nondecaying antisymmetric mode corresponding to buckling of the plate. Asymptotic expansion of that nondecaying mode near the stress free state predicts buckling according to the classical Euler formula. A consistent interpretation of end effects in the presence of a nondecaying mode is given.
[DOI: 10.1115/1.1794703]

1 Introduction

Studies of axial decay rates of incremental end disturbances expose considerable sensitivity of the decay rate to the initial prestrain level applied to a plate [1–4]. These works are limited to semi-infinite plates subjected to tensile loads. Sensitivity analysis of finitely stretched plates in compression is reported recently by Durban and Stronge [5] for a compressible solid, suggesting high sensitivity near bifurcation points of the plate. That work, though, does not reflect explicitly on the question of validity of Saint-Venant's principle in a prebuckled plate.

The main objective of the present study is to extend previous analyses of incremental version of Saint-Venant's principle in tension to plates under compression. Here we consider the case of a finite compressive load applied to a rectangular finite plate subjected to plane strain constraint. Some objection to such analysis can rise due to possible buckling of the plate. Nevertheless, its importance stems from the existence of many structures sustaining compressive loads to which Saint-Venant's principle might be applied.

The approach and formulation in the present paper follow those given by Durban and Karp [4]. The governing equations are summarized in Section 2 followed by verification of their ellipticity for the four hyperelastic materials examined. Mathematical formulation leads to an eigensystem, for the two velocity components, that admits a separation-of-variables solution for the eigenfields. These are concisely recapitulated in Section 3. Axial decay is exponential with the decay rates obtained as eigenvalues of the transcendental equations.

Numerical solutions for the lowest exponential decay rates are detailed in Section 4 for four hyperelastic rubbers examined by Blatz and Ko [6], Storåkers [7], and Ogden [8]. Calculations have been performed over a range of prestrain within the limits of ellipticity of the governing equations. Due to the existence of a purely real eigenvalue in the solution, the discussion in Section 5 is devoted to consistent interpretation of that eigenvalue followed by implementation of the solution to the question of validity of Saint-Venant's principle.

It was found that for a plate, in the prebuckled state, Saint-Venant's principle is a reasonable assumption when the width to length ratio is lower than 1/10. The pattern of the dependence of the decay rate on the compression is found to be highly sensitive to constitutive parameters. The smallest decay rate belongs to the symmetric field and remains practically the same up to a moderate compression level.

Buckling conditions are naturally unveiled by the analysis as a sole real eigenvalue. At low levels of prestrain, asymptotical expansion for the real eigenvalue generates the Euler formula for buckling. The constitutive sensitivity of the buckling load at high compressive loads is demonstrated.

2 Perturbed Field Equations and Ellipticity

Consider a finite plate of initial width and length $2H_0$ and $2L_0$, respectively, uniformly stretched (in tension or in compression) under plane-strain conditions in the axial direction z, by uniform stress σ applied at the ends $(z=0,2L)$ (see Fig. 1). The current width and length of the plate are $2H$ and $2L$, respectively. The axial stretch λ $(L=\lambda L_0)$ is considered as a controlled variable. Assume now an incremental self-equilibrating load is superposed on the uniform stress at one of the ends (or on both) thus inducing a quasi-static perturbed velocity $\mathbf{V}$ within the plate. We wish to examine the (plane-strain) instantaneous response of the prestrained plate to that incremental disturbance. The faces $x=\pm H$ remain free of tractions in both prestrained and disturbed states.

The perturbed velocity vector is written as (see Fig. 1 for the unit triad $\mathbf{i},\mathbf{j},\mathbf{k}$)

$$\mathbf{V}=u\mathbf{i}+w\mathbf{k} \tag{2.1}$$

where both velocity components u, w depend only on x and z coordinates. The nonzero Eulerian strain rate components follow in the form

$$\varepsilon_x=u_{,x} \quad \gamma_{xz}=\tfrac{1}{2}(u_{,z}+w_{,x}) \quad \varepsilon_z=w_{,z} \tag{2.2}$$

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, April 21, 2003; final revision; June 13, 2003. Associate Editor: M.-J. Pindera. 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.

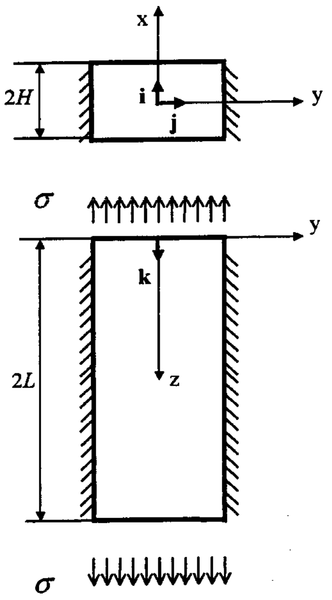


Fig. 1 Plate, of current length and width $2L$ and $2H$, respectively, subjected to uniform tension $(\sigma>0,\lambda>1)$ or compression $(\sigma<0,\lambda<1)$ stress in plane strain condition $(\lambda_y=1)$ with boundaries $x=\pm H$ free of traction

The incremental material response is governed by three in plane relations suggested by Hill [9]

$$\overset{\nabla}{\sigma}_x=a\varepsilon_x+c\varepsilon_z \quad \overset{\nabla}{\tau}_{xz}=2\mu\gamma_{xy} \quad \overset{\nabla}{\sigma}_z=(c-\sigma)\varepsilon_x+b\varepsilon_z \tag{2.3}$$

where $(\overset{\nabla}{\sigma}_x,\overset{\nabla}{\tau}_{xz},\overset{\nabla}{\sigma}_z)$ are the objective Jaumann stress rates and (a,b,c,μ) denote instantaneous material moduli, derivable from the strain energy function for hyperelastic solids and dependable on the uniform prestrain condition (detailed in [4]).

Equilibrium of stress rates requires, in view of (2.1) and (2.3), that the velocity components u, w satisfy the set of two equations [9]

$$au_{,xx}+\beta u_{,zz}+(c+\alpha)w_{,xz}=0 \tag{2.4a}$$

$$(c+\alpha)u_{,xz}+\alpha w_{,xx}+bw_{,zz}=0 \tag{2.4b}$$

with

$$\alpha=\mu-\tfrac{1}{2}\sigma \quad \beta=\mu+\tfrac{1}{2}\sigma \tag{2.5}$$

The boundary conditions for a plate with long faces free of traction are given by [10]

$$\mathbf{t}_x=\overset{\nabla}{\sigma}_x\mathbf{i}+(\overset{\nabla}{\tau}_{xz}-\sigma\gamma_{xy})\mathbf{k}=\mathbf{0} \quad \text{at} \quad x=\pm H \tag{2.6}$$

where $\mathbf{t}_x$ is the traction rate at the surface normal to the x coordinate. Condition (2.6), by substitution of (2.2) and (2.3), can be expressed by velocity components as

$$\begin{matrix} au_{,x}+cw_{,z}=0 \\ u_{,z}+w_{,x}=0 \end{matrix} \quad \text{at} \quad x=\pm H \tag{2.7}$$

Solution of the equilibrium equations (2.4) along with the boundary conditions (2.7) generates eigenfunctions with associated eigenvalues. The eigenfunctions can be regarded as a "spectral footprint" of the plate from which the actual response is composed. The actual response is determined by the particular boundary disturbance at the ends $z=0$, $2L$, not prescribed here (see recent paper by Ling et al., [11] and references cited therein for reconstruction of actual response out of end data).

Four constitutive models for hyperelastic solids are examined. The response of all four solids is represented by the strain energy function [12]

$$W=\sum_j \frac{C_j}{m_j}\left[\lambda_1^{m_j}+\lambda_2^{m_j}+\lambda_3^{m_j}-3+\frac{1}{n}(J^{-nm_j}-1)\right] \tag{2.8}$$

where λ_1, λ_2, λ_3 are the principal stretches and $J=\lambda_1\lambda_2\lambda_3$ is the volume ratio. The summation is carried over pairs C_j, m_j which, like n, are known material constants. Details of the derivation of stress-stain relations and the instantaneous moduli are given in [4].

The specific models considered in this study are two highly compressible solids; the Blatz-Ko (BK) foam rubber due to Blatz and Ko [6] with the single term representation

$$\text{BK} \quad m_1=-2 \quad C_1=-22 \quad n=0.5 \quad (\nu_0=0.25) \tag{2.9}$$

and the compressible vulcanized natural foam rubber due to Storåkers [7] (St1) with two-term representation

$$\text{St1} \quad m_1=-m_2=4.5 \quad C_1=1.85 \quad C_2=-9.2 \quad n=0.92$$

$$(\nu_0=0.324) \tag{2.10}$$

The nearly incompressible synthetic foam rubber is due to Storåkers [7] (St2) with two-term representation

$$\text{St2} \quad m_1=-m_2=3.6 \quad C_1=2.04 \quad C_2=-0.51 \quad n=25$$

$$(\nu_0=0.49) \tag{2.11}$$

and the noncompressible rubber is by Ogden [8] (OG) with three-term representation

$$\text{OG} \quad m_1=1.3 \quad m_2=4.0 \quad m_3=-2.0$$

$$C_1=69 \quad C_2=1.0 \quad C_3=-1.22 \quad n\to\infty \quad (\nu_0=0.5) \tag{2.12}$$

Here C_j have dimensions of $10^{-2}\ \mathrm{N\,mm^{-2}}$.

Ellipticity of the governing equations (2.4) will fail when [9]

$$d^2=ab\alpha\beta \tag{2.13}$$

where d is defined by

$$d=ab+\alpha\beta-(c+\alpha)^2 \tag{2.14}$$

For Blatz-Ko solid, that condition can be solved analytically to yield the limit points

$$\lambda=(7+4\sqrt{3})^{-3/8}\approx 0.3724, \quad \lambda=(7-4\sqrt{3})^{-3/8}\approx 2.685 \tag{2.15}$$

Additional inquiry into the ellipticity regimes detailed by Hill [9] reveal that for the BK rubber plate, strong ellipticity is maintained in the range

$$0.3724<\lambda<2.685 \tag{2.16}$$

in accordance with the more general result given by Knowles and Sternberg [13]. For the three rubbers given by (2.10)–(2.12) condition (2.13) is evaluated numerically leading to the conclusion that strong ellipticity is maintained for St1 rubber provided $\lambda>0.334$, while no limits on a strong ellipticity have been found for St2 and OG solids in the region examined.

3 Eigenfunctions and Eigenvalues

Separation of variables solution of (2.4) is sought via the representation

$$u=U(x)\exp\left(\frac{i\pi kz}{2h}\right) \quad w=W(x)\exp\left(\frac{i\pi kz}{2h}\right) \tag{3.1}$$

where $U(x)$ and $W(x)$ are the transverse profiles representing the eigenfunction and k is the associated eigenvalue (complex in general). While solution (3.1) is written in an exponential form in z, the possibility to obtain a harmonic, nondecaying response is included (via real eigenvalue). The notation in (3.1) is slightly different from the one employed in common studies of end effects due to the factor $i\pi/2$ in the exponential argument. The present formulation, however, is in agreement with the notation in the literature of wave guides and with some of the stability studies. Thus, according to (3.1), the imaginary part of the eigenvalue $\mathrm{Im}\{k\}$ represents the decay rate of the corresponding eigenfunction.

Inserting (3.1) into the equilibrium equations (2.4) results in two coupled ordinary differential equations

$$aU'' - \beta\left(\frac{\pi k}{2h}\right)^2 U + i(c+\alpha)\left(\frac{\pi k}{2h}\right)W' = 0 \tag{3.2a}$$

$$i(c+\alpha)\left(\frac{\pi k}{2h}\right)U' + \alpha W'' - b\left(\frac{\pi k}{2h}\right)^2 W = 0 \tag{3.2b}$$

where the prime denotes differentiation with respect to *x*. Due to the symmetry of boundary conditions around the *x-y* plane, the solution of (3.2) is commonly separated into symmetric and antisymmetric fields (indicated respectively by subscripts *s* and *a*) given by

$$U_s = A_1 \sinh\left(\Gamma_1 \frac{\pi k x}{2h}\right) + A_2 \sinh\left(\Gamma_2 \frac{\pi k x}{2h}\right) \tag{3.3a}$$

$$W_s = A_1 \eta_1 \cosh\left(\Gamma_1 \frac{\pi k x}{2h}\right) + A_2 \eta_2 \cosh\left(\Gamma_2 \frac{\pi k x}{2h}\right) \tag{3.3b}$$

and

$$U_a = A_3 \cosh\left(\Gamma_1 \frac{\pi k x}{2h}\right) + A_4 \cosh\left(\Gamma_2 \frac{\pi k x}{2h}\right) \tag{3.4a}$$

$$W_a = A_3 \eta_1 \sinh\left(\Gamma_1 \frac{\pi k x}{2h}\right) + A_4 \eta_2 \sinh\left(\Gamma_2 \frac{\pi k x}{2h}\right) \tag{3.4b}$$

Here A_1, A_2, A_3, A_4 are integration constants, Γ_1 and Γ_2 are the roots of the characteristic equation of the set (3.2)

$$a\alpha\Gamma^4 - d\Gamma^2 + b\beta = 0 \tag{3.5}$$

with solutions

$$\Gamma_1 = \sqrt{\frac{d - \sqrt{d^2 - 4ab\alpha\beta}}{2a\alpha}} \quad \Gamma_2 = \sqrt{\frac{d + \sqrt{d^2 - 4ab\alpha\beta}}{2a\alpha}} \tag{3.6}$$

where

$$\eta_p = \frac{i(a\Gamma_p^2 - \beta)}{\Gamma_p(c+\alpha)} \quad p = 1,2 \tag{3.7}$$

Compliance with boundary conditions (2.7) generates two transcendental equations for the eigenvalue *k*

$$\tanh\left(\Gamma_1 \frac{\pi k}{2}\right) - \left(\frac{Q_1}{Q_2}\right)^{\pm 1} \tanh\left(\Gamma_2 \frac{\pi k}{2}\right) = 0 \tag{3.8}$$

where the plus and minus signs correspond to symmetric and antisymmetric fields, respectively, and

$$Q_p = \frac{a\Gamma_p^2 + ic\eta_p}{\eta_p \Gamma_p + i} \quad p = 1,2 \tag{3.9}$$

Trancendental equations equivalent to (3.8) have been derived previously by Durban and Stronge [2,3]. Similar equations have been derived by Ogden and Roxburgh [14,15] in the context of analysis of stability of plates, allowing only purely real eigenvalues to be considered as a solution.

Each Eq. (3.8) generates an infinite number of eigenvalues with accompanying eigenfunctions. While in the linearly elastic case the completeness of the eigenfunction set is proved by Buchwald [16] and Gregory [17], for the incremental case it is not yet established. Nevertheless, it is conceivable that an arbitrary instantaneous surface disturbance at the ends will contain a combination of these eigenfunctions. In that sense it has been argued by Karp and Durban [18] that an eigenvalue with smallest $\mathrm{Im}\{k\}$ provides a lower bound on the axial decay rate of end disturbance at any given prestress. Regardless of the completeness question, it has been shown by Durban and Stronge [3] that the eigenfunctions of a plate with faces $x = \pm H$ free of traction are self-equilibrated, a property of central importance in the analysis of validity of Saint-Venant's principle.

4 Numerical and Asymptotical Results

With the known dependence of the instantaneous moduli (a,b,c,μ) on the prestretch λ, the eigenvalues *k* are calculated numerically by solving the transcendental equations (3.8) for stretch λ as a controlled variable. The standard Muller method is employed to yield a finite set of the smallest complex eigenvalues in a prestrain range of $0.5 < \lambda < 1.5$. The eigenvalues are ordered in increasing order of their (positive) imaginary part, namely,

$$0 \leqslant \mathrm{Im}\{k_0\} < \mathrm{Im}\{k_1\} < \mathrm{Im}\{k_2\} < \cdots \tag{4.1}$$

Due to intersection of modes with the increase of prestrain (in compression and in tension), this labeling and ordering, adopted from Linear Elasticity, is consistent only near the stress-free state. In what follows, the labeling of the eigenvalues in the stress-free state according to (4.1) is retained in all regions, even though that ordering in some regions is no longer valid. The discussion here is limited to two first eigenvalues k_0 and k_1, for the antisymmetric and symmetric fields, respectively. Additional eigenvalues are presented as an exception, in cases when the intersection of modes makes other eigenvalues to expose a lower imaginary part.

4.1 BK Rubber. Figure 2 shows a variation of the first two eigenvalues with prestrain λ for the BK solid in the extended range of prestrain $0.5 < \lambda < 2.3$, including the tension range. For ease of reference, the plot in Fig. 2 is subdivided into the following three regions:

a. The vicinity of the stress-free state $0.8 < \lambda < 1.5$
b. High compression $\lambda < 0.8$
c. High tension $\lambda > 1.5$

Region A. To begin with, we notice the continuity of the k_1 eigenvalue on the transition from tension to compression through the stress-free state $\lambda = 1$. That continuity is observed for all complex eigenvalues, not shown in Fig. 2. At the stress-free state, all eigenvalues coincide with the linear elasticity values obtained from the Fadle-Papkovich equation [19]. First symmetric eigenvalue is

$$\pi/2k_1 = 1.1254 + i2.1061 \tag{4.2}$$

First antisymmetric eigenvalue k_0, is purely real in the range of $\lambda_p < \lambda < 1$ where λ_p is a stretch at which k_0 reaches peak value k_p and coincides with the trivial solution of the Fadle-Papkovich equation at $\lambda = 1$. That eigenvalue admits asymptotical expansion in the vicinity of the origin—near the stress-free state for a small eigenvalue $|k| \ll 1$. Asymptotic expansion, according to the lines given by Durban and Karp [4] for tension, can be generalized to include compression to yield

$$\frac{\pi}{2}k_0 \approx |i\sqrt{3\Delta}| \tag{4.3}$$

where, $\Delta \equiv \lambda - 1$ with $|\Delta| \ll 1$. Extending that relation to compression supports the numerical result according to which the eigenvalue k_0 is purely real in compression $(\Delta < 0)$ and purely imaginary in tension $(\Delta > 0)$. In that sense, this eigenvalue has a continuous nature in transition from tension to compression expressed by (4.3), similarly to all other eigenvalues. The numerical results in Fig. 2 near $\lambda = 1$ are found to be in a good agreement with the asymptotical relation given by (4.3).

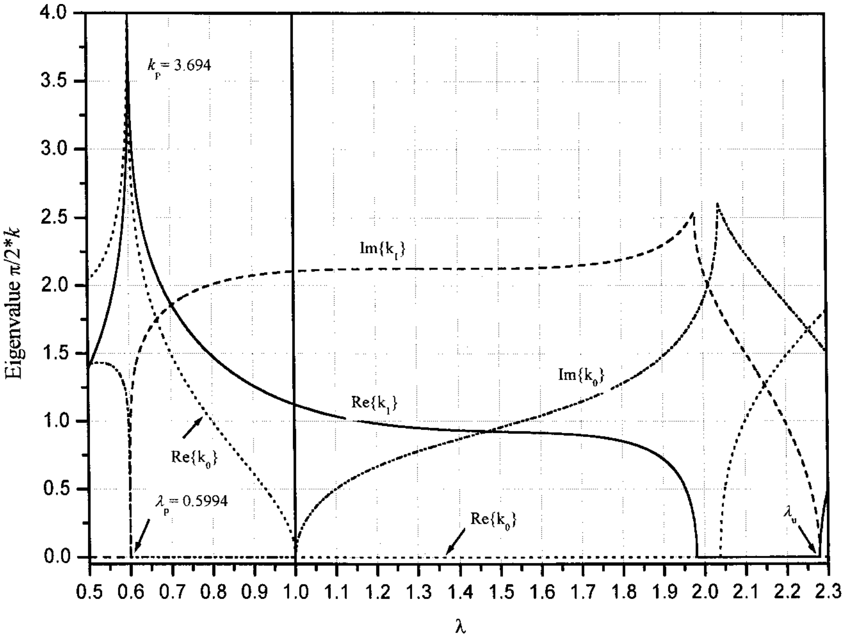


Fig. 2 Two lowest eigenvalues (multiplied by factor $\pi/2$) in tension ($\lambda>1$) and in compression ($\lambda<1$) for the BK rubber (2.9). k_0 and k_1 are the antisymmetric and the symmetric eigenvalues, respectively.

Region B. At high compression, the imaginary part of k_1 reaches zero at a prestrain λ_p, while the real part reaches a peak k_p. The values of these parameters, λ_p and k_p, are derivable along the following lines.

We begin with the observation that the imaginary part of k_1 is perpendicular to the λ axis at the peak point λ_p. That perpendicularity ($dk/d\lambda=\infty$) is exposed by all higher eigenvalues, as well as by the imaginary part of the antisymmetric k_0 and can be expressed alternatively by

$$\frac{d\lambda}{dk}=0 \tag{4.4}$$

The relation between the eigenvalue k and the prestretch λ is given implicitly by the transcendental equation (3.8). For the present purpose, it is convenient to rewrite the transcendental equation (3.8) in an equivalent form

$$F(\lambda,k)\equiv(Q_1-Q_2)\sinh\left[(\Gamma_1+\Gamma_2)\frac{\pi k}{2}\right] \pm(Q_1+Q_2)\sinh\left[(\Gamma_1-\Gamma_2)\frac{\pi k}{2}\right]=0 \tag{4.5}$$

Using this implicit function F, condition (4.4) is equivalent to condition

$$\frac{\partial F(\lambda,k)}{\partial k}=0 \tag{4.6}$$

That condition applied to (4.5) reads

$$(\Gamma_1+\Gamma_2)(Q_1-Q_2)\sinh\left[(\Gamma_1+\Gamma_2)\frac{\pi k}{2}\right] \pm(\Gamma_1-\Gamma_2)(Q_1+Q_2)\sinh\left[(\Gamma_1-\Gamma_2)\frac{\pi k}{2}\right]=0 \tag{4.7}$$

which should hold for any k. For Blatz-Ko rubber, that condition can be simplified noticing the complex nature of the roots of the characteristic equation (3.5) in the entire range of prestrain under consideration. These roots are designated as [5]

$$\Gamma_1=X-iY \quad \Gamma_2=X+iY \tag{4.8}$$

where the real and the imaginary parts have an elegant expressions

$$X^2=\frac{1}{2}\left(\sqrt{\frac{b\beta}{a\alpha}}+\frac{d}{2a\alpha}\right) \quad Y^2=\frac{1}{2}\left(\sqrt{\frac{b\beta}{a\alpha}}-\frac{d}{2a\alpha}\right) \tag{4.9}$$

For BK solid, these expressions have a simple form

$$X^2=\frac{1}{2}\left(1+\frac{8R-1-R^2}{6R}\right) \quad Y^2=\frac{1}{2}\left(1-\frac{8R-1-R^2}{6R}\right) \tag{4.10}$$

where $R\equiv\lambda^{-8/3}$.

Using expressions (4.8) in (4.7) lead to

$$X(Q_1-Q_2)\sinh\left(2X\frac{\pi k}{2}\right)\pm iY(Q_1+Q_2)\sinh\left(i2Y\frac{\pi k}{2}\right)=0 \tag{4.11}$$

The real part of that expression, representing the slope of the curve, will be zero for any k provided

$$Q_1-Q_2=0 \tag{4.12}$$

Substituting expression (4.9) and the instantaneous moduli a, b, c, μ in expression (3.9), condition (4.12) takes the form

$$\frac{3R^2-12R+1}{R(1+R)}=0 \tag{4.13}$$

Solution of that quadratic equation leads to the prestrain for BK rubber

$$R=2-\sqrt{\tfrac{11}{3}}\Rightarrow\lambda_p\approx0.5994 \tag{4.14}$$

at which the real part of the symmetric eigenvalue has a peak value and the imaginary part reaches zero. That value is in a good agreement with the numerical result in Fig. 2.

The peak value of the eigenvalue at this prestretch λ_p can be obtained directly from the transcendental equation (4.5). Under the requirement expressed by (4.12), Eq. (4.5) is reduced to

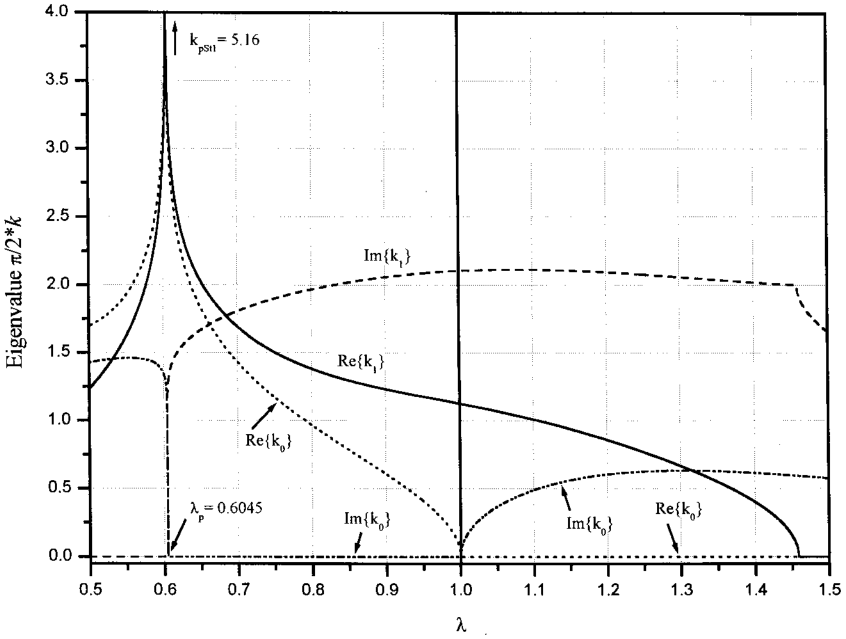


Fig. 3 Two lowest eigenvalues (multiplied by factor $\pi/2$) in tension ($\lambda>1$) and in compression ($\lambda<1$) for the St1 rubber (2.10). k_0 and k_1 are the antisymmetric and the symmetric eigenvalues, respectively.

$$\sinh\left[(\Gamma_1-\Gamma_2)\frac{\pi k}{2}\right]=0 \tag{4.15}$$

Using the complex notation of (4.8) in Eq. (4.15) gives a simple condition

$$\sinh\left(i2Y\frac{\pi k}{2}\right)=0 \tag{4.16}$$

Observing that at the critical prestrain given by (4.14) the eigenvalue is real $k=k_p$, condition (4.16) reduces to

$$\sin\left(2Y\frac{\pi k_p}{2}\right)=0 \tag{4.17}$$

which yields[1]

$$\frac{\pi}{2}k_p=\frac{n\pi}{2Y}\quad n=1,2,3\ldots \tag{4.18}$$

For the appropriate value of Y at the prestrain given by (4.14) calculated from (4.9*b*), we find that the peak value for BK solid is

$$\frac{\pi}{2}k_p=3.694 \tag{4.19}$$

It is conceivable from Eq. (4.5), as well as from Fig. 2, that the critical prestrain λ_p and the peak value k_p are both common to symmetric and antisymmetric modes. The observable deviation of the peak value in Fig. 2 from the correct one given by (4.19) is due to numerical sensitivity to conversion criteria in the vicinity of that singular point. Identical peak values (4.14) and (4.19) for BK model are obtained by Durban and Stronge [5] based on an assumption that the symmetric and the antisymmetric modes intersect at that point. Beyond the peak value, the symmetric eigenvalue k_1 is purely real while the antisymmetric eigenvalue k_0 becomes complex.

Region C. At high tension, the symmetric eigenvalue k_1 reaches zero (both real and imaginary parts) at a prestrain $\lambda_u=3^{3/4}\approx 2.28$ corresponding to necking under uniaxial plane-strain tension reported previously by Durban and Karp [4]. That necking phenomenon is absent in St1, St2, and OG rubbers (see [4] for St1 and St2 results) and given here for BK rubber for the sake of completeness.

4.2 St1 Rubber. The first two eigenvalues for St1 rubber for the prestrain range $0.5<\lambda<1.5$ (regions A and B) are plotted in Fig. 3. It can be observed that the St1 rubber shares with the BK rubber the pattern of the eigenvalue dependence on the prestrain in both regions A and B. The eigenvalues at the stress-free state are, again, identical to those obtained in Linear Elasticity and agree with the asymptotical relation (4.3).

Detailed analysis reveals that for St1 rubber, the roots Γ_1, Γ_2 are complex for $\lambda<0.734$, making the peak analysis given above for the BK rubber applicable here as well. Numerical evaluation of Eqs. (4.12) and (4.18) yields the critical prestrain and the peak values for the St1 solid to be

$$\lambda_p\approx 0.6045\quad \frac{\pi}{2}k_p\approx 5.16 \tag{4.20}$$

4.3 St2 and OG Rubbers. The first two eigenvalues for the nearly and completely incompressible solids, St2 and OG, for the prestrain range $0.5<\lambda<1.5$, are given in Figs. 4 and 5, respectively. The incompressibility of the OG solid in (2.8) is approximated by taking $n=200$. As expected, the eigenvalues at the stress-free state, $\lambda=1$, are identical to those obtained for compressible solids discussed above. That constitutive insensitivity coincides with asymptotic relations (4.3) and with the numerical results obtained in Linear Elasticity (the Fadle-Papkovich equation does not comprise any material parameters).

---

[1] The integer n used here and in the sequel should impose no confusion with the material constant used in Eqs. (2.8)–(2.12).

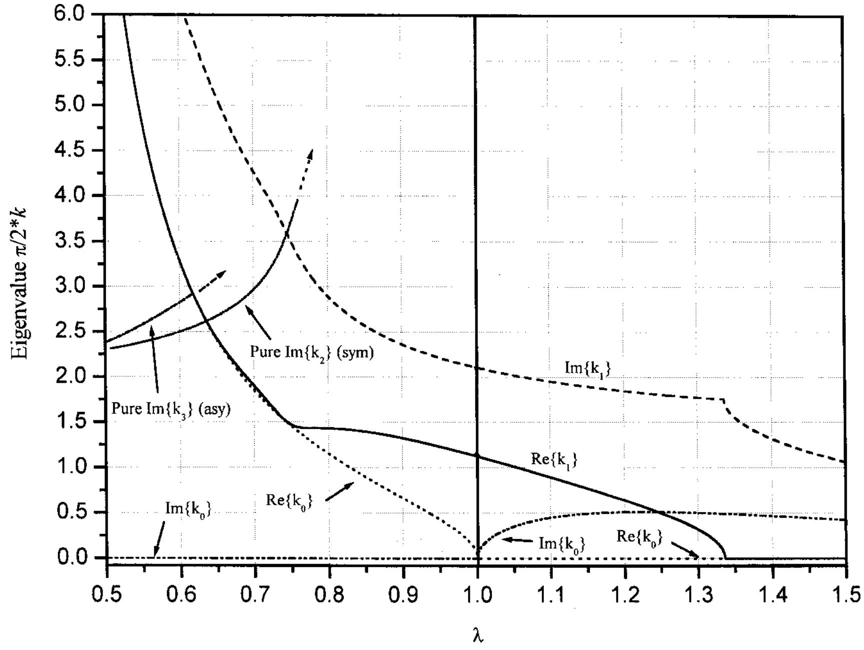


Fig. 4 Two lowest eigenvalues (multiplied by factor $\pi/2$) in tension ($\lambda>1$) and in compression ($\lambda<1$) for the St2 rubber (2.11). k_0 and k_1 are the antisymmetric and the symmetric eigenvalues, respectively.

The pattern of the branches of the eigenvalues for both St2 and OG solids are similar, though differs from the pattern obtained for the compressible solids (Figs. 2 and 3). Both solids do not exhibit a peak at high compression in the range examined. Indeed, the roots Γ_1, Γ_2 for these nearly and completely incompressible solids are real, making the analysis of the peak value not valid. Both materials lack a purely real symmetric eigenvalue at high compressive prestrain levels, exhibited by the symmetric eigenvalue for BK and St1 solids beyond the peak prestrain λ_p. Moreover, the lowest symmetric eigenvalue in region B is purely imaginary and intersects the imaginary part of k_1 at $\lambda \approx 0.75$. That means that beyond the stretch $\lambda \approx 0.75$, the decay rate for St2 rubber is governed by an eigenvalue that does not appear to be the first in the stress-free state. For OG rubber such intersection occurs at lower stretch, not shown in Fig. 5. It should be noted that the higher eigenvalues, not shown in figures here, behave qualitatively in the same manner as the k_1 eigenvalue for each solid.

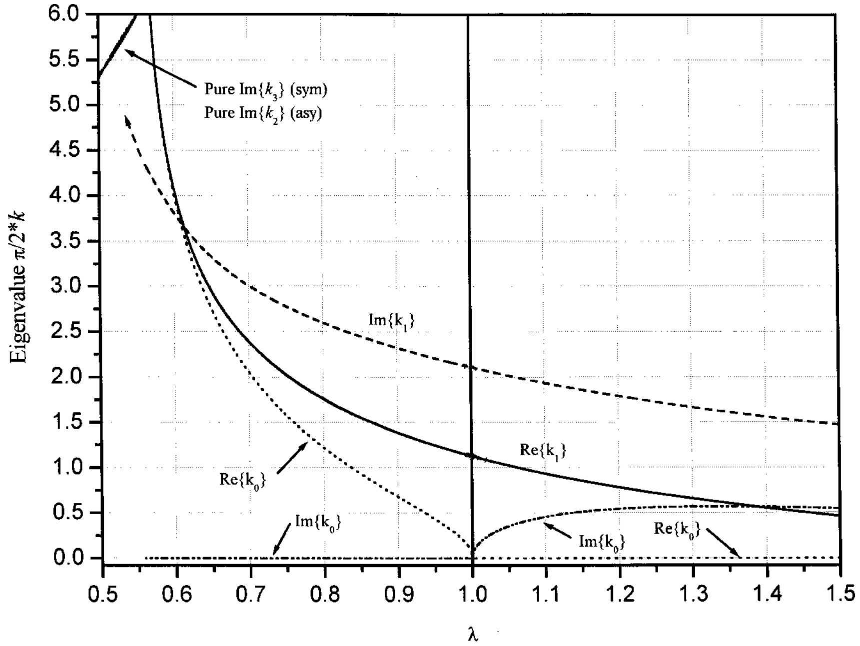


Fig. 5 Two lowest eigenvalues (multiplied by factor $\pi/2$) in tension ($\lambda>1$) and in compression ($\lambda<1$) for the OG rubber (2.12). k_0 and k_1 are the antisymmetric and the symmetric eigenvalues, respectively.

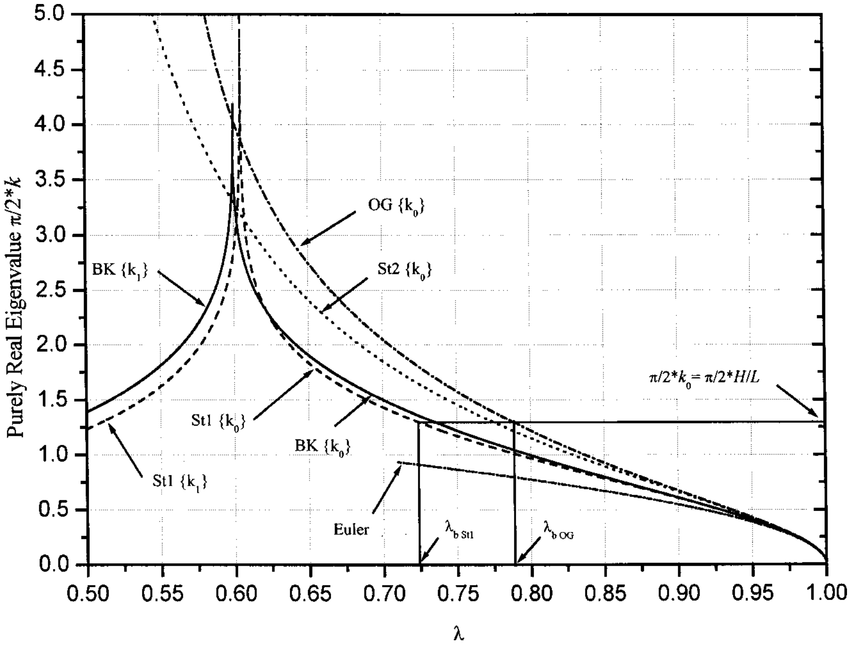


Fig. 6 Purely real eigenvalue in compression for the four rubbers BK, St1, St2, OG, and the classical Euler formula for buckling of a pinned-pinned column

5 Discussion

The characteristic decay rate of end effects is governed by the eigenvalue with the lowest imaginary part[2] (e.g., Refs. [20] and [19]) and associated with the lower bound for validity of Saint-Venant's principle. However, the eigenvalue with the lowest imaginary part is k_0, which is purely real in the compression range of region A. Therefore, before the question of validity of Saint-Venant's principle in compression can be addressed, an interpretation of that purely real eigenvalue in the present study is suggested in the sequel.

5.1 Buckling. Incremental analysis of a prestrained state, posed in Sec. 2, is commonly employed as a bifurcation analysis. The objective of such analysis is to obtain the loci of a critical prestrain level λ_b, at which an adjacent equilibrium first becomes possible, as a function of the aspect ratio of a column H/L. It turns out that the real eigenvalue k_0 in the analysis given here is equivalent to the aspect ratio of the column H/L in the bifurcation analysis. Indeed, some versions of the real branch of k_0, and the purely real k_1 beyond λ_p, depicted in Figs. 2–5, have been exposed in several previous papers, in the context of bifurcation analysis (e.g., Refs. [15], [21–23] and recently Ref. [5]).

The equivalence between the real eigenvalue k_0 and the aspect ratio of a plate H/L can be unveiled by simple consideration. The purely real eigenvalue in solution (3.1) represents displacement, harmonic in the axial direction, with no axial decay. Boundary conditions for sliding ends (no tangential traction and no axial displacement) lead to a solution of the equilibrium equation in the form

$$\sin\left(\pi\frac{L}{H}k_0\right)=0 \tag{5.1}$$

which is reduced to

$$k_0=n\frac{H}{L}\quad n=1,2,3\ldots \tag{5.2}$$

On the other hand, for small prestrain levels, k_0 is given by (4.3) and can be written in the form

$$k_0\approx i\frac{2}{\pi}\sqrt{3\frac{\sigma}{E}} \tag{5.3}$$

which by substitution in (5.2) leads to

$$\sigma_{cr}\approx n^2\frac{\pi^2}{(2L)^2}EI\quad n=1,2,3\ldots \tag{5.4}$$

where E is Young's modulus and I is the inertia of the plate for bending. Expression (5.4) is the Euler formula for the buckling of columns of length $2L$ with pinned ends. The possibility to recover the Euler formula from the eigenfunction analysis with the same boundary conditions at the ends has been demonstrated by Levinson [21] (as well as by many following works) along somewhat different arguments.

An illuminating observation in the present context is an equivalent derivation of the beam theory out of the eigenfunction analysis for a plate in tension [4]. It was demonstrated there that the relation (4.3) is derivable from beam theory for a cantilever to which a self equilibrated system of loads and moments is applied at the free end.

According to these lines, the real eigenvalue can be interpreted as a criterion for the onset of antisymmetric instability (buckling) for a given geometry [represented here by k_0 through (5.2)] with appropriate boundary conditions at the ends. The purely real k_0 for the four solids is replotted jointly in Fig. 6 along with the classical Euler buckling condition. If one wishes to find the stretch λ_b, at which instability first occurs, the aspect ratio of the column—H/L first should be fixed. Assuming sliding ends, the appropriate k_0 is calculated from expression (5.2). Entering Fig. 6 with that value returns the prestrain at which the first mode buckling will occur. It is interesting to note that the more compressible solids sustain higher levels of compression before buckling can occur.

Following the reasoning given above, it can be argued that for some material properties and geometry H/L, only complex eigenvalues are possible in the range $\lambda_b<\lambda<1$, where λ_b is the buck-

---

[2] Recall the exchange of the imaginary and the real parts here due to different notation in (3.1).

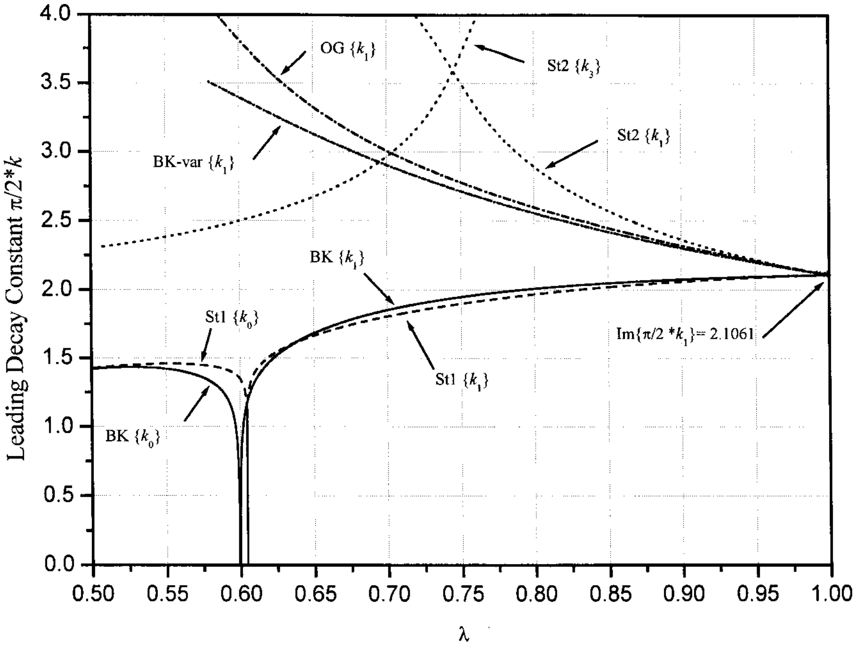


Fig. 7 Leading decay exponent in compression for the four rubbers BK, St1, St2 OG, and for a variation of Bk rubber (BK-Var) with compressibility constant $n=25$

ling stretch corresponding to that geometry and material properties. In other words, the eigenfunction expansion shows that, for a plate of infinite length, a real eigenvalue is possible for any compressive load. In a finite plate, on the other hand, that real eigenvalue is excluded in the range $\lambda_b<\lambda<1$ by virtue of the boundary conditions at the perturbed ends and becomes possible only for $\lambda<\lambda_b$. Thus, the region $\lambda_b<\lambda<1$ is considered as a prebuckling region to which the question of validity of Saint-Venant's principle might be relevant.

5.2 Saint-Venant's Principle. By excluding the sole real antisymmetric eigenvalue from the set of eigenfunctions in the prebuckled state, an infinite number of complex eigenfunctions are left, representing axial decay of end disturbance. Therefore, the lower bounds on the decay rate are estimated by the first symmetric Im$\{k_1\}$ and the antisymmetric Im$\{k_2\}$ eigenvalues in region A and B, respectively. These two eigenvalues are recapitulated in Fig. 7 for the four solids defined by (2.9)–(2.12). It can be concluded from Fig. 7 that for BK and St1 rubbers, the lowest imaginary part in the range $\lambda_p<\lambda<1$ belongs to the symmetric eigenvalue k_1. These solids are characterized by an approximately constant decay rate up to moderate compression levels ($\lambda\approx 0.8$) with an abrupt drop of the decay rate, reaching zero at a critical prestrain λ_p.

The nearly incompressible solids, St2 and OG, on the other hand, exhibit an increase of decay rate in compression. For the St2 solid, the rate of decay drops with further compression ($\lambda<0.75$) to the decay rate of the order of the decay at the stress-free state ($k\approx 2$) due to interchange of modes. Similar interchange of modes is exhibited by OG solid at higher compression, not shown in the figures.

The difference between the patterns of compressible solids and nearly incompressible solids is interesting on two grounds. First is the very exposure of different behavior. Second is the absence of such substantial difference in tension. As a further examination of the constitutive sensitivity of that pattern, a calculation has been made for a modified BK solid with a compressibility constant $n=25$ (identical to the compressibility of the St2 solid) with other constants in (2.9) unaltered. The decay constant for that hypothetical solid is plotted on Fig. 7, designated by BK-var. The pattern of that solid resembles the pattern characterizing the nearly incompressible solids (St2, OG) suggesting that compressibility is a major property in pattern determination. Nevertheless, it is conceivable that additional material properties might have a nonnegligible effect as well.

For Saint-Venant's principle to be applicable to a finite plate, it is necessary to verify the smallness of the decay length in comparison to the length of the plate. One possible way to assess the applicability of Saint-Venant's principle is by evaluating the distance $l_{0.01}$ at which the end disturbance decreases by 99%. From (3.1) that condition is given by

$$\exp\left(i\,\frac{\pi}{2}\,\frac{l_{0.01}}{H}\,k_1\right)=0.01 \tag{5.5}$$

where k_1 has been chosen as a lower bound for the decay length. Substituting a typical value Im$\{\pi/2k_1\}\approx 2.106$ for k_1 in the region $0.8<\lambda<1$ [from (4.2) or Fig. 7)], we find that

$$l_{0.01}=\frac{4.6}{2.1061}H\approx 2.2H \tag{5.6}$$

By limiting the affected region to be not larger than 1/10 of the plate length from each end, the minimal length of the plate will be

$$2L>22H \tag{5.7}$$

According to (5.2), in order to avoid buckling of such plate, the prestrain should be smaller than the critical value corresponding to

$$k_0=\frac{H}{L}=\frac{1}{11} \tag{5.8}$$

Inserting (5.8) into (4.3) results in

$$\Delta\approx\frac{1}{3}\left(\frac{1}{11}\right)^2\left(\frac{\pi}{2}\right)^2 \tag{5.9}$$

which corresponds to stretch $\lambda\approx 0.993$.

Any attempt to avoid buckling by increasing the aspect ratio H/L will be followed by decreasing the region in the plate, that is not affected by end disturbances, loosening the validity of Saint-Venant's principle to that geometry. It can be noticed from (5.6), that the increase of the decay rate in the nearly incompressible solids (St2, OG) observed in Fig. 7, has only a small effect on the prestrain level obtained in (5.9) for which Saint-Venant's principle is valid.

6 Summary

The above discussion can be summarized as follows: Saint-Venant's principle can be applied to plates in compression with the limitation on the plate being ‘‘long enough’’ ($H/L < 1/10$), a requirement essentially identical to the one imposed in cases of tension, bending, or torsion. This requirement puts relatively severe limits on compressive loads in order to pertain the relevance of the semi-inverse method solutions (the outer solutions). It turns out that the limits of Euler buckling of column with pinned ends and the limits of applicability of Saint-Venant's principle are of the same order of magnitude. Thus, as long as slender plates do not buckle, Saint-Venant's principle can be applied safely. For shorter plates, which sustain higher compressive loads without buckling, Saint-Venant's principle might be too rough an assumption.

Finally, it should be emphasized that the relations (5.2), (5.8), and (5.9) were derived for plates with ideal boundary conditions of smooth ends. It is conceivable that different conditions will end up with modified relations for (5.2), (5.8), and (5.9). Nevertheless, the estimation of decay distance in (5.6) as an upper bound is valid for any boundary data.

Acknowledgments

Part of this study has been performed while on sabbatical leave to the University of British Columbia, Vancouver, Canada. The author wishes to thank Professor Gary Schajer of the University of British Columbia, Vancouver, for helpful discussions on the subject of the present paper. The hospitality of the Mechanical Engineering Department of UBC, Vancouver, is gratefully acknowledged.

References

[1] Abeyaratne, R., Horgan, C. O., and Chung, D.-T., 1985, ‘‘Saint-Venant End Effects for Incremental Plane Deformations of Incompressible Nonlinearly Elastic Materials,’’ ASME J. Appl. Mech., **52**, pp. 847–852.
[2] Durban, D., and Stronge, W. J., 1988, ‘‘On the Validity of Saint Venant's Principle in Finite Strain Plasticity,’’ ASME J. Appl. Mech., **55**, pp. 11–16.
[3] Durban, D., and Stronge, W. J., 1988, ‘‘Diffusion of Self-Equilibrating End Loads in Plane Strain Plasticity,’’ J. Mech. Phys. Solids, **36**, pp. 459–476.
[4] Durban, D., and Karp, B., 1992, ‘‘Axial Decay of Self-Equilibrating End Loads in Compressible Solids,’’ ASME J. Appl. Mech., **59**, pp. 738–743.
[5] Durban, D., and Stronge, W. J., 1995, ‘‘Plane-Strain Incremental Response and Sensitivity of Stretched Plates,’’ Eur. J. Mech. A/Solids, **14**, pp. 553–575.
[6] Blatz, P. J., and Ko, W. L., 1962, ‘‘Application of Finite Elastic Theory to the Deformation of Rubbery Materials,’’ Trans. Soc. Rheol., **6**, pp. 223–251.
[7] Storåkers, B., 1986, ‘‘On Material representation and Constitutive Branching in Finite Compressible Elasticity,’’ J. Mech. Phys. Solids, **34**, 125–145.
[8] Ogden, R. W., 1986, ‘‘Recent Advances in the Phenomenological Theory of Rubber Elasticity,’’ Rubber Chem. Technol., **59**, 361–383.
[9] Hill, R., 1979, ‘‘On the Theory of Plane Strain in Finitely Deformed Compressible Materials,’’ Math. Proc. Cambridge Philos. Soc., **86**, pp. 161–178.
[10] Durban, D., and Stronge, W. J., 1992, ‘‘Diffusion of Incremental Loads in Prestrained Bars,’’ Proc. R. Soc. London, Ser. A, **439**, pp. 583–600.
[11] Ling, Y., Engel, P. A., and Geer, J. A., 1994, ‘‘The End Problem of Incompressible Elastic Cylinders,’’ ASME J. Appl. Mech., **61**, pp. 30–37.
[12] Hill, R., 1978, ‘‘Aspects of Invariance in Solid Mechanics,’’ Adv. Appl. Mech., **18**, pp. 1–75.
[13] Knowles, J. K., and Sternberg, E., 1977, ‘‘On the Failure of Ellipticity of the Equations for Finite Elastostatic Plane Strain,’’ Arch. Ration. Mech. Anal., **63**, pp. 321–336.
[14] Ogden, R. W., and Roxburgh, D. G., 1993, ‘‘The Effect of Pre-Stress on the Vibration and Stability of Elastic Plates,’’ Int. J. Eng. Sci., **31**, pp. 1611–1639.
[15] Roxburgh, D. G., and Ogden, R. W., 1994, ‘‘Stability and Vibration of Pre-Stressed Compressible Elastic Plates,’’ Int. J. Eng. Sci., **32**, pp. 427–454.
[16] Buchwald, V. T., 1964, ‘‘Eigenfunctions of Plane Elastostatics I. The Strip,’’ Proc. R. Soc. London, Ser. A, **277**, pp. 385–400.
[17] Gregory, R. D., 1980, ‘‘The Traction Boundary Value Problem for the Elastostatis Semi-Infinite Strip; Existence of Solution, and Completeness of the Papkovich-Fadle Eigenfunctions,’’ J. Elast., **10**, pp. 295–327.
[18] Karp, B., and Durban, D., 2002, ‘‘Influence of Boundary Conditions on Decay Rates in a Prestrained Plate,’’ ASME J. Appl. Mech., **69**, pp. 515–520.
[19] Timoshenko, S. R., and Goodier, J. N., 1982, *Theory of Elasticity*, Third ed., McGraw-Hill.
[20] Horgan, C. O., and Knowles, J. K., 1983, ‘‘Recent Developments Concerning Saint-Venant's Principle,’’ Adv. Appl. Mech., **23**, pp. 179–269.
[21] Levinson, M., 1968, ‘‘Stability of a Compressed Neo-Hookean Rectangular Parallelepiped,’’ J. Mech. Phys. Solids, **16**, pp. 403–415.
[22] Sawyers, K. N., and Rivlin, R. S., 1974, ‘‘Bifurcation Conditions for a Thick Elastic Plate Under Thrust,’’ Int. J. Solids Struct., **10**, pp. 483–501.
[23] Ogden, R. W., 1984, *Non-Linear Elastic Deformations*, Dover Pub., New York.

Numerical Prediction of Cavitating MHD Flow of Electrically Conducting Magnetic Fluid in a Converging-Diverging Nozzle

Jun Ishimoto
Associate Professor,
Department of Intelligent Machines and System Engineering,
Hirosaki University,
3, Bunkyo-cho,
Hirosaki 036-8561, Japan
e-mail: ishimoto@cc.hirosaki-u.ac.jp

The fundamental characteristics of the two-dimensional cavitating MHD flow of an electrically conducting magnetic fluid in a vertical converging-diverging nozzle under a strong nonuniform magnetic field are numerically predicted to realize the further development and high performance of a two-phase liquid-metal MHD power generation system using electrically conducting magnetic fluids. First, the governing equations of the cavitating flow of a mercury-based magnetic fluid based on the unsteady thermal nonequilibrium multifluid model are presented, and several flow characteristics are numerically calculated taking into account the effect of the strong nonuniform magnetic field. Based on the numerical results, the two-dimensional structure of the cavitating flow and cavitation inception phenomena of the mercury-based magnetic fluid through a converging-diverging nozzle are shown in detail. The numerical results demonstrate that effective two-phase magnetic driving force, fluid acceleration, and high power density are obtained by the practical use of the magnetization of the working fluid. Also clarified is the precise control of the cavitating flow of magnetic fluid that is possible by effective use of the magnetic body force that acts on cavitation bubbles. [DOI: 10.1115/1.1794164]

1 Introduction

The fundamental investigation of cavitating flow or two-phase flow phenomena of magnetic fluid with electrical conductivity is very interesting and important, not only for the basic study of hydrodynamics of magnetic fluids, but also for finding solutions to problems related to the development of practical engineering applications of two-phase electromagnetic fluids, such as the two-phase liquid metal MHD (LMMHD) power generation system [1–6]. In this regard, a fluid-driving system using two-phase flows or cavitating flows of magnetic fluid has been proposed by one of the authors [7–9].

The idea of using a two-phase flow system originated from the two-phase LMMHD power generation system, which was proposed and developed by Petrick and Branover [1]. After their proposal, the present authors reported the results of a theoretical study that demonstrated the possibility of using an electrically conducting magnetic fluid (ECMF) [10–13] as a working fluid in a boiling two-phase LMMHD power generation system [14], where it was shown that a better driving force or pressure rise than that of the conventional LMMHD system was obtained by using ECMF as the working fluid due to the practical application of the magnetization of the fluid.

Furthermore, theoretical and experimental studies on the basic characteristics of two-phase flow of magnetic fluid were conducted, and the possibility of flow control or effective driving-force generation by magnetic force in the new energy conversion system using boiling two-phase flow was confirmed [7,8]. It was also concluded that stabilization of two-phase flow is possible by effective application of the magnetic force of the fluid [8,15]. According to these previous studies, it is likely that high performance of power generation systems is possible by applying the ECMF to a working fluid in the two-phase LMMHD [14] power generation system.

In the application of the two-phase flow of the magnetic fluid to an actual fluid transport apparatus, it is important to determine a simple and effective method to generate the two-phase flow state for the improvement of the total performance of a fluid driving system using multiphase flow. However, conventional two-phase flow systems essentially require a powerful heat source or gas-injection equipment to produce the boiling two-phase flow or gas-liquid two-phase flow state. Additionally, research on methods for the production of the two-phase magnetic fluid-flow state have not been precisely focused, and only a few studies have so far been made on the basic mechanism of the cavitating flow due to the difficulty of confirming for experimental and theoretical results in high-speed two-phase magnetic fluid flow with phase change.

In order to overcome these difficulties, we contrived a new type of LMMHD power generation system with a two-phase fluid driving and acceleration system by using cavitating flow of ECMF as a working fluid. This system is characterized by its utilization of a two-phase magnetic driving force and no-heat sources or no-additional gas-injection devices are required, except for a converging-diverging nozzle. Based on an advanced mathematical model, which takes the effect of two-phase electromagnetic body force acting on cavitating magnetic fluid flow state into consideration, we herein develop a new method for analyzing cavitating flow. Such an electrically conducing magnetic fluid is usually prepared by dispersing fine iron (Fe) particles in a liquid metal, such as mercury [10,11]. To prevent solidification of particles and to maintain homogeneous dispersion, the particles' surfaces 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 13, 2003; final revision, March 1, 2004. Associate Editor: B. A. Younis. 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.

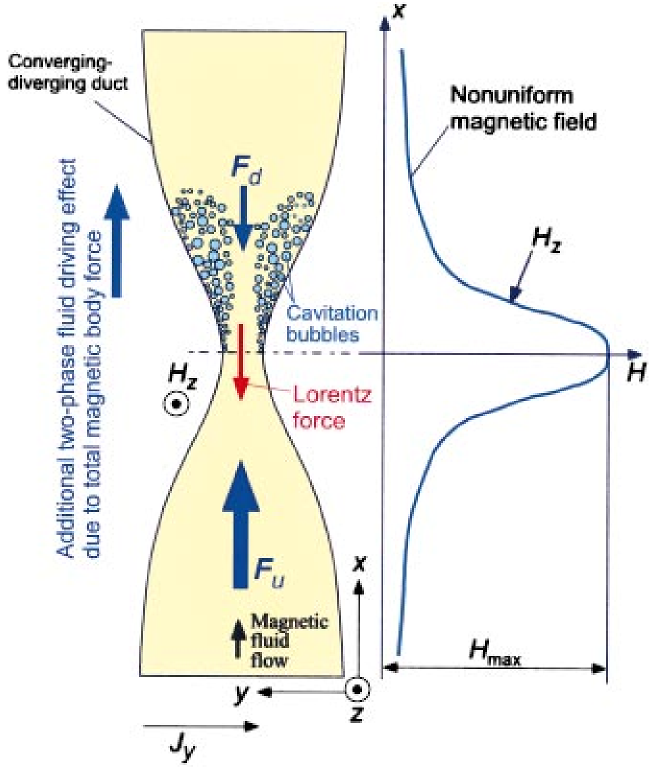


Fig. 1 Principle of cavitating MHD power generation system using electrically conducting magnetic fluid (ECMF) flow. Magnetic body force $F_u = \mu_0 M \cdot \nabla H = F_d$ in the case without cavitation, and $F_d = (1-\alpha_g)\mu_0 M \cdot \nabla H < F_u$ with cavitation (H is the vector of magnetic field and M is the vector of magnetization).

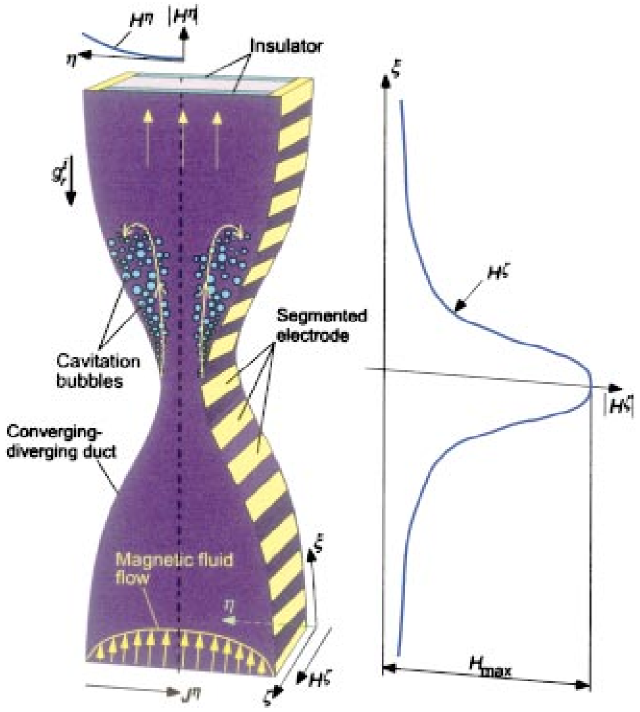


Fig. 2 Schematic of model for numerical analysis

coated with a thin film of tin. Thus, the ECMF behaves as fluid having magnetization.

The principle of such a two-phase LMMHD fluid-driving system proposed by one of the authors is schematically depicted in Fig. 1. In this system, the flow is accelerated in the region of the converging nozzle, although the Lorentz force acts in the opposite direction of the mainstream. Cavitation inception is induced in the downstream throat of a diverging nozzle due to a pressure decrease. Furthermore, the flow is additionally accelerated not only by the pumping effect of the cavitation bubbles, but also by the rise of magnetic pressure induced by the unbalance of magnetic body forces that act in the single- and two-phase flow regions under a nonuniform magnetic field.

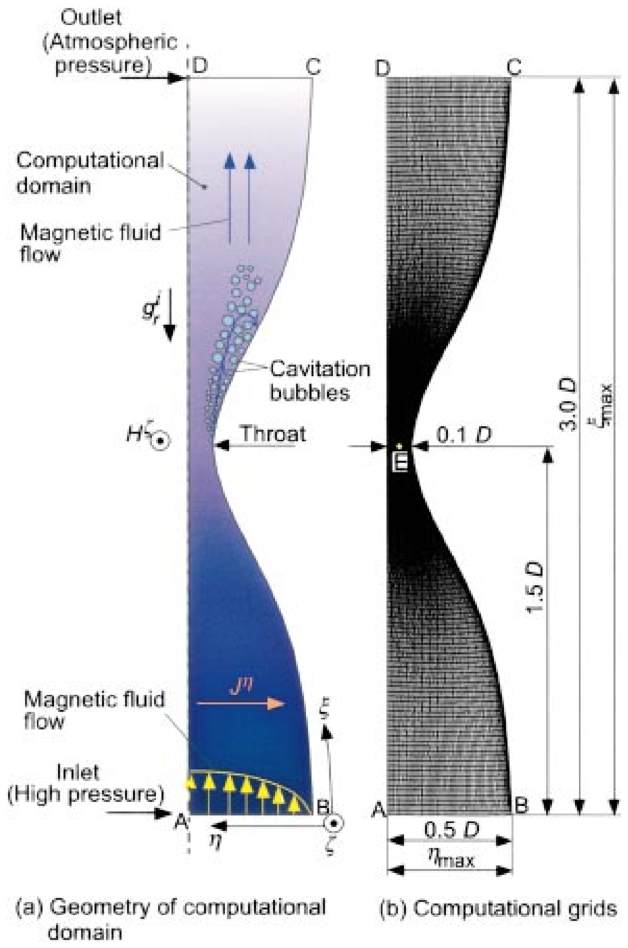
Cavitation is usually found in high-speed liquid flows around obstacles, such as the impellers of fluid machinery. It is well known that cavitating flow causes many adverse effects including material erosion, noise, and performance degradation in turbomachinery. In the magnetohydrodynamic field, the experimental study on cavitation of mercury flow in a horizontal venturi channel under magnetic field has been conducted [16]. In this research, the effect of the magnetic field on mercury cavitation or two-phase pressure loss has been precisely investigated. The main finding of the experiment is that the cavitation inception can be easily generated by applying the magnetic field especially in case of the abrupt diverging channel flow. Nevertheless, the effective method for magnetic control in regard to the cavitation inception or the improvement of pressure loss has not been sufficiently clarified.

In the usual fluids engineering field, the cavitation is something that people like to avoid in the fluid machinery implementation. However, the unique feature in the present research is that the cavitation phenomenon is positively utilized to improve the fluid-driving effect of conventional two-phase LMMHD power generation system. It seems reasonable to suppose that the presently proposed cavitating MHD-ECMF driving system is most applicable to utilize in a pre- or post-power generation system for the conventional LMMHD systems. Accordingly, to combine the present ECMF system with the conventional system, the total two-phase MHD performance of the hybrid power generation system will be drastically improved.

In the present study, numerical analysis is extended to the case of mercury-based electrically conducting magnetic fluid as a basic study to demonstrate the possibility of application of ECMF to working fluid in a two-phase LMMHD power generation system. Specifically, the two-dimensional high-speed cavitating flow characteristics of ECMF in a vertical converging-diverging nozzle under a strong nonuniform traverse magnetic field are numerically predicted to realize the further development and high performance of the two-phase fluid-driving system and to realize high power density in the application of ECMF to two-phase LMMHD power generation systems. The governing equations for cavitating flow of ECMF based on the unsteady multifluid model in the generalized curvilinear coordinate system are presented, and then several cavitating flow characteristics are numerically calculated, taking into account the effect of the strong nonuniform magnetic field. The numerical results of the cavitating flow characteristics of the ECMF are compared to those of mercury (electrically conducting nonmagnetic fluid) and to those of the boiling two-phase flow of ECMF.

2 Numerical Method

The numerical model used in the analysis is schematically depicted in Fig. 2, and the total system of the computational domain and numerical grids used in the present analysis is depicted in Fig. 3. The ECMF flows between parallel insulated plates under the applied nonuniform magnetic field H and homogeneous electric field E. The mainstream is in a vertically upward direction. The magnetic field, the electric field, and the mainstream of the work-



(a) Geometry of computational domain (b) Computational grids

Fig. 3 Schematic of computational system used in numerical analysis

ing fluid are orthogonal to one another. A minutely segmented electrode is installed on the duct side. In the initial stationary state, the flow duct is filled with pressurized mercury-based magnetic fluid, and flow immediately occurs when the outlet D-C is opened. Magnetic fluid is continuously introduced at high speed via the inlet section A-B, the flow is accelerated at the point of the converging-diverging nozzle, and the inception of cavitation is induced by a pressure decrease. The model for analysis simulates the high-speed cavitating flow of magnetic fluid passing through the converging-diverging nozzle in a vertical duct.

2.1 Governing Equations. In the present numerical formulation of the cavitating flow characteristics of magnetic fluid, we extend the general two-fluid model to the new vapor-liquid multiphase fluid model taking into account the effect of the strong nonuniform magnetic field for analysis, which is based on the unsteady thermal nonequilibrium multifluid model by Kataoka [17], and Harlow and Amsden [18]. In the numerical model, the condition of the working fluid with the cavitating magnetic fluid flow structure can be approximated to form a homogeneously dispersed bubbly flow. In the process of modeling, to consider the effects of the rapid evaporation and condensation of magnetic fluid, we apply the rapid phase-change model of Yamamoto et al. [19] and Young [20] to the cavitating flow of magnetic fluid.

The calculation is carried out using the two-dimensional generalized curvilinear coordinate system (ξ, η) as shown in Fig. 3; ξ and η denote the longitudinal and transverse coordinate, respectively. ζ denotes the orthogonal coordinate in relation to the ξ and η. It is assumed that the flow field is homogeneous in the ζ direction and is symmetric to the central axis D-A as shown in Fig. 2. The nonuniform magnetic field is applied in the ζ direction, which is transverse to the mainstream of working fluid flow. In numerical modeling under this condition, the following assumptions are employed to formulate the governing equations:

1. The cavitating flow is a two-dimensional unsteady internal flow.
2. The magnetic field and electric field are not influenced by the existence of the gas phase.
3. The applied electric field is homogeneous and steady.
4. The polarization of the liquid and gas phase is neglected.
5. The induced magnetic field in the flow is much smaller than the applied magnetic field ($\mathrm{R_m} \ll 1$).
6. The energy exchange between the liquid and gas phases is taken into account.
7. The liquid phase is incompressible fluid.

Here, we mention additional explanations concerning the above assumptions. The magnetic Reynolds number $[\mathrm{R_m} = \sigma \mu_l u_{l(\mathrm{in})}^{\xi} D]$ in the present analysis is $\mathrm{R_m} \approx 0.075$. Accordingly, the induced magnetic field in the flow can be neglected because the typical MHD approximation of $\mathrm{R_m} \ll 1$ is satisfied in the numerical condition. The liquid-phase compressibility does not so strongly influence the two-phase MHD flow characteristics compared to the gas-phase compressibility. It seems reasonable to suppose that the compressibility of the cavitating flow is mainly dominated by the gas-phase compressibility. Also, the total cavitating flow is regarded as weak compressible flow because of the dispersed gas phase, which has strong compressibility. Consequently, we assumed that the liquid phase is an incompressible fluid of constant density. Furthermore, we assumed that the aspect ratio (height/width) of the flow duct is small, the electrical conductivity of the segmented electrode is sufficiently large, and the boundary layer of insulator side wall is sufficiently thin, which does not influence the mainstream. These assumptions introduce the validity of numerical condition for the two-dimensional flow [21,22]. If the aspect ratio is large or the applied magnetic field is an unsteady traveling induction field, then three-dimensionality in the MHD flow should be considered.

In general, it is known that the electric current in the conducting fluid has to go around the nonconducting particles, such as vapor bubbles [23–25]. The deformation of the electric field by the nonconducting particle generates inhomogeneity of the electromagnetic force. According to this effect, the additional flow fields are created surrounding the particles. As a result, it has been previously studied and determined that electromagnetic expulsive force is generated and that this force contributes to the migration of nonconducting particles [23–25]. The hydrodynamic velocity dealt with in the electromagnetic separation of the metallurgy process represented in the previous research [23] is on the order of several (μm/s), and the handling particle diameter is less than 30 μm. Thus, it may be possible to suppose that the nonconducting particle migration by the electromagnetic expulsive force is dominant in the limited condition of the extremely slow velocity field and small particle diameter. However, the gas- and liquid-phase mean velocity, which deal with the present analysis, is on the order of 1.0–7.0 (m/s), and the velocity magnitude is quite large compared to the condition for research on electromagnetic separation, which requires the consideration of electromagnetic expulsive force. Also, the cavitation bubble expands to the diameter of about 0.30 mm in the present two-phase flow analysis. Accordingly, the electromagnetic expulsive force and the induced migration of bubbles by the force are neglected in the present analysis.

Under the above conditions, the governing equations of the cavitating MHD flow of ECMF, taking into account the effect of a nonuniform magnetic field based on the unsteady two-dimensional multifluid model, are derived as follows.

The mass conservation equation for a gas phase is

$$\frac{\partial}{\partial t}(\alpha_g\rho_g)+\nabla_j(\alpha_g\rho_g u_g^j)=\Gamma_g \tag{1}$$

The mass conservation equation for a liquid phase is

$$\frac{\partial}{\partial t}(\alpha_l\rho_l)+\nabla_j(\alpha_l\rho_l u_l^j)=\Gamma_l \tag{2}$$

where α_g and α_l are the gas- and liquid-phase volume fraction, respectively. The relationship $\alpha_g+\alpha_l=1$ is assumed. u_g^j and u_l^j are the gas-and liquid-phase contravariant velocity, respectively.

Equations for the electromagnetic field are

$$\frac{1}{2}e^{ijk}(\nabla_j H_k-\nabla_k H_j)=J_T^i \tag{3}$$

$$M_T^i=\alpha_l M^i \tag{4}$$

$$B_T^i=\mu_0(H^i+M_T^i) \tag{5}$$

$$J_T^i=\sigma_T(E^i+e^{ijk}u_{lj}B_{Tk}) \tag{6}$$

where the subscript T denotes the two-phase flow. σ_T denotes the electrical conductivity in the two-phase region and is defined by the following equation [26]:

$$\sigma_T=\sigma_l\frac{2(1-\alpha_g)}{2+\alpha_g} \tag{7}$$

where σ_l is the electrical conductivity of the base liquid. The high-void fraction region is especially found in the limited region of wall vicinity where the cavitation actively generates. The region is constructed of closely aggregated small-bubble clouds and not by large single bubbles. Also the two-phase magnetic body force effectively performs compared to the Lorentz force in the region. Therefore, the Eq. (7) is generally applicable to approximate the electrical conductivity in the whole two-phase flow field except for the limited region of high α_g.

The strength of the electric field E^η is determined from a load factor K of the outside electric circuit. K is defined by the following equation:

$$K=\frac{E^\eta}{\overline{u_l^\xi}\cdot B_T^\zeta} \tag{8}$$

where $\overline{u_l^\xi}$ is the cross-sectional mean value of the longitudinal liquid-phase velocity component u_l^ξ. In the present analysis, it is assumed that the load factor K of the electrode is always in the condition of $K=-0.5$. It is also assumed that the direction of current J^i is a negative transverse coordinate of η. Therefore, the Lorentz force always acts as a flow resistance. In this numerical model, the case of $-1<K<0$ corresponds to the MHD power generation system, and the case of $K\leqslant -1$ corresponds to the electromagnetic pump.

The momentum equation for single-phase ECMF flow by Shizawa and Tanahashi [27,28] is extended to the two-phase flow case, and the combined equation of motion for a total gas and liquid phase is derived by the following equation:

$$\begin{aligned}\frac{\partial}{\partial t}(\alpha_g\rho_g u_g^i+\alpha_l\rho_l u_l^i)+\nabla_j(\alpha_g\rho_g u_g^i u_g^j+\alpha_l\rho_l u_l^i u_l^j)\\ =-g^{ij}\nabla_j p_l+\mu_0 M_T^j\nabla_j H^i+\mu_0 e^{ijk}J_{Tj}H_k+\beta_T g^{jk}\nabla_j\nabla_k u_l^i\\ +\tfrac{1}{3}(\beta_T\nabla_j\nabla_k u_l^k)g^{ij}+\alpha_l\rho_l g_r^i\end{aligned} \tag{9}$$

where the second and third terms on the right-hand side of Eq. (9) represent the magnetic body force term and the Lorentz force term in the two-phase flow, respectively. Especially in the case of the ECMF flow, the Lorentz force is given by the cross product of J^i and H^i [27,28].

Additionally, β_T in Eq. (9) denotes the viscosity of the two-phase mixture flow that includes small dispersed bubbles. β_T was evaluated using the following formula for the viscosity of a suspension [29,30]:

$$\beta_T=\left[1-\left(\frac{\alpha_g}{0.680}\right)\right]^{-2}\cdot\beta_l,\quad \alpha_g<0.5 \tag{10}$$

To consider the effects of additional forces that act on the bubbles and the effects of radial expansion of the bubbles, the equation of motion for the gas phase is replaced with the translational motion of a single bubble [31]. Therefore, the Eulerian-Lagrangian two-way coupling model [32,33] is applied to predict the two-dimensional cavitating flow characteristics.

The equation of motion for the gas phase is

$$\frac{4}{3}\pi\rho_g R_g^3\frac{\mathrm{d}u_g^i}{\mathrm{d}t}=-F_p^i+F_g^i-F_D^i-F_{VM}^i-F_B^i+F_{LM}^i+F_{LS}^i \tag{11}$$

where each additional force term is derived as follows:

$$F_p^i=\frac{4}{3}\pi R_g^3 g^{ij}\nabla_j p_l \tag{12}$$

$$F_g^i=\frac{4}{3}\pi R_g^3\rho_g g_r^i \tag{13}$$

$$F_D^i=\frac{1}{2}\rho_l \mathrm{C_D}|u_g^i-u_l^i|(u_g^i-u_l^i)\pi R_g^2 \tag{14}$$

$$F_{VM}^i=\mathrm{C_{VM}}\cdot\rho_l\frac{4}{3}\pi R_g^3\left[\frac{\mathrm{d}}{\mathrm{d}t}(u_g^i-u_l^i)+\frac{3}{R_g}(u_g^i-u_l^i)\frac{\mathrm{d}R_g}{\mathrm{d}t}\right] \tag{15}$$

$$F_B^i=6R_g^2\sqrt{\pi\rho_l\beta_l}\int_0^t\frac{\frac{\mathrm{d}}{\mathrm{d}\tau}(u_g^i-u_l^i)}{\sqrt{t-\tau}}d\tau \tag{16}$$

$$F_{LM}^i=\pi R_g^3\rho_l e^{ijk}(\Omega_{gj}-\Omega_{lj})(u_{gk}-u_{lk}) \tag{17}$$

$$F_{LS}^i=6.46\frac{\beta_l R_g^2}{\sqrt{|(\Omega_g^i-\Omega_l^i)|\nu_l}}e^{ijk}(\Omega_{gj}-\Omega_{lj})(u_{gk}-u_{lk}) \tag{18}$$

$$\Omega_l^i=\frac{1}{2}\omega_l^i=\frac{1}{4}e^{ijk}(\nabla_j u_{lk}-\nabla_k u_{lj}) \tag{19}$$

where F_p^i is the force due to the liquid-phase pressure gradient, F_g^i is the gravitational acceleration force, F_D^i is the drag force, F_{VM}^i is the virtual mass force considering the expansion of a bubble, and F_B^i is the Basset history term, which takes into account the effect of the deviation in flow pattern from steady state. F_{LM}^i is the Magnus lift force caused by the rotation of the bubble as reported by Auton et al. [34]. F_{LS}^i is Saffman's lift force [35] caused by the velocity gradient of the liquid phase. $\mathrm{C_D}$ is the drag coefficient, $\mathrm{C_{VM}}$ is the virtual mass coefficient, R_g is the equivalent bubble diameter, Ω^i is the contravariant angular velocity, and ω^i is the contravariant vorticity. d/dt denotes the substantial derivative.

The equation for the angular velocity of a bubble is derived as follows [35]:

$$\frac{d\Omega_g^i}{dt}=\frac{15\beta_l}{R_g^2\cdot\rho_g}(\Omega_l^i-\Omega_g^i) \tag{20}$$

The energy equation for the gas and liquid phases is

$$\frac{\partial}{\partial t}(\alpha_m\rho_m e_m)+\nabla_j(\alpha_m\rho_m e_m u_m^j)$$
$$=-p_m\frac{\partial\alpha_m}{\partial t}-\nabla_j(\alpha_m p_m u_m^j)+\Gamma_m h_m^{(i)}+q_m^{(i)}a^{(i)}$$
$$-\nabla_j(\alpha_m q_m^j)+\alpha_m\Phi_m \tag{21}$$

In the above equation, the subscript m denotes the gas phase ($m=g$) or liquid phase ($m=l$). $h_g^{(i)}$ and $h_l^{(i)}$ are the enthalpy of the gas phase and the liquid phase at the interface, respectively. $a^{(i)}$ is the gas-liquid interfacial area concentration per unit volume. $\Gamma_g h_g^{(i)}$ and $\Gamma_l h_l^{(i)}$ are the interfacial energy transfer terms due to the liquid-vapor phase change. $q_g^{(i)}$ and $q_l^{(i)}$ are the heat transfer terms of mutual interaction between the vapor and liquid interface. q^j is the contravariant heat flow vector, and Φ is the energy dissipation function, as described below:

$$\begin{cases} q_m^i=-\lambda_m g^{ij}\nabla_j T_m \\ \Phi_m=-\dfrac{2}{3}\beta_m(\nabla_i u_m^i)^2+2\beta_m s_{jm}^i s_{im}^j \\ s_{jm}^i=\dfrac{1}{2}(\nabla_j u_m^i+\nabla_i u_m^j) \end{cases} \tag{22}$$

where the present sufficient conditions of two-dimensional flow with small magnetic Reynolds number ($\mathrm{R_m}\ll 1$) contribute to decrease the effect of energy loss due to Joule heating. The induced current is not so large as to increase the Joule heating because of the condition for load factor of $K=-0.5$. Also, since the initial fluid temperature (573.15 K) given is sufficiently higher than room temperature, the effect of temperature increase caused by the Joule heating is small compare to the total unsteady temperature profiles influenced by the convection, conduction, dissipation, and energy exchange between gas and liquid phases. Therefore, the effect of Joule heating in the energy equation (21) is neglected.

Assuming that the mass of each vapor bubble and of the condensed liquid droplet in each computational location is constant results in the following mass conservation equation for number density N_k:

$$\frac{\partial}{\partial t}\left(\frac{4}{3}\pi R_k^3 N_k\rho_k\right)+\nabla_j\left(\frac{4}{3}\pi R_k^3 N_k\rho_k u_k^j\right)=\Gamma_k$$
$$\begin{cases} k=e:\ R_k=R_g,\ N_k=N_g,\ \rho_k=\rho_g,\ u_k^i=u_g^i,\ \Gamma_k=\Gamma_g \\ k=c:\ R_k=R_l,\ N_k=N_l,\ \rho_k=\rho_l,\ u_k^i=u_l^i,\ \Gamma_k=\Gamma_l \end{cases} \tag{23}$$

where subscript k denotes evaporation ($k=e$) or condensation ($k=c$).

The governing equations of cavitating flow mentioned above are constructed by Eulerian-type equations for the liquid phase and by Lagrangian-type equations for the gas phase.

2.2 Constitutive Equations. The drag coefficient $\mathrm{C_D}$, and the virtual mass coefficient $\mathrm{C_{VM}}$, are defined as follows [35]:

$$\begin{cases} \mathrm{C_D}=\dfrac{24}{\mathrm{Re_B}}(1+0.15\,\mathrm{Re_B^{0.687}})+\dfrac{0.42}{1+42500\,\mathrm{Re_B^{-1.16}}} \\ \mathrm{C_{VM}}=0.5 \\ \mathrm{Re_B}=\dfrac{\rho_l|u_g^i-u_l^i|D}{\beta_l} \end{cases}. \tag{24}$$

The energy balance condition through the interface of the gas and liquid phases is expressed by the following equation:

$$\begin{cases} \Gamma_g h_g^{(i)}+\Gamma_l h_l^{(i)}=0 \\ q_g^{(i)}+q_l^{(i)}=0 \end{cases} \tag{25}$$

where $h_k=c_{pk}T_k$; $k=g,l$. The detailed constitutive equations for interfacial energy transfer terms in Eq. (25) are given by following extended empirical formulas [36,37]:

$$q_g^{(i)}=k^{(i)}(T_g-T_s) \tag{26}$$

where $k^{(i)}$ is the interfacial heat transfer rate between gas and liquid phases and is given by following equations [36]:

$$\begin{cases} k^{(i)}=\alpha k_g^{(i)}+\alpha_l k_l^{(i)} \\ k_g^{(i)}=\dfrac{8.067\cdot\lambda_g}{R_g} \\ k_l^{(i)}=\dfrac{1.0+0.37\,\mathrm{Re_V^{0.50}}\cdot\mathrm{Pr_V^{0.35}}}{R_g} \\ \mathrm{Re_V}=\dfrac{2.0R_g|u_g^i-u_l^i|}{\nu_l} \\ \mathrm{Pr_V}=\dfrac{c_{pl}\cdot\beta_l}{\lambda_l} \end{cases} \tag{27}$$

It is assumed that the energy transfer is caused by the heat transfer between the isothermal spherical bubble and the surrounding liquid. Assuming a spherical bubble with equivalent radius R_g, the expression of the interfacial area concentration per unit volume $a^{(i)}$ is obtained by the following equation [17]:

$$a^{(i)}=\frac{3\alpha_g}{R_g} \tag{28}$$

In general, the interfacial transfer terms are proportional to the interfacial area concentration, $a^{(i)}$. Therefore, $a^{(i)}$ is one of the most important parameters in the two-fluid model. Assuming that the vapor gas phase follows an ideal gas law and that the relationship between gas-phase pressure p_g and density ρ_g obeys polytropic change, the following equation by Hirt and Romero [38] results:

$$\rho_g(\kappa_g-1)e_g=[p_g-c_{l0}^2\rho_l(\alpha_g^*-\alpha_g)]\alpha_g^*$$
$$\begin{cases} \alpha_g\geqslant\alpha_{gc}: & \alpha_g^*=\alpha_g \\ \alpha_g<\alpha_{gc}: & \alpha_g^*=\alpha_{gc} \end{cases} \tag{29}$$

where c_{l0} is the sound velocity in the mercury at the initial state ($c_{l0}=1321.0$ m/s), and α_{gc} denotes the threshold of the void fraction ($\alpha_{gc}=0.005$) [38].

The constitutive equation for gas-phase generation density Γ_g is defined by the following equation:

$$\Gamma_g=\Gamma_{ge}-\Gamma_{gc} \tag{30}$$

where Γ_{ge} and Γ_{gc} denote the gas-phase evaporation density and gas-phase condensation density, respectively. By introducing constitutive equations for Γ_{ge} and Γ_{gc}, also to consider the effect of the surface tension γ_l in the cavitation inception process, we extend the classical nucleation theory for water droplets from subcooled vapor to the mercury-based magnetic fluid. Namely, Γ_{ge} and Γ_{gc} are assumed to be proportional to the degree of subcooling and superheat. Furthermore, if Γ_{gk}, ($k=e,c$) is expressed by the sum of the nucleation rate of the evaporated bubble or the condensed liquid droplet and also by the increase in mass due to the growth of vapor bubbles and condensed droplets, the following equations for Γ_{gk} are derived [19,20]:

$$\Gamma_{gk}=\frac{4}{3}\pi\rho_k I_k R_{k(\mathrm{cr})}^3+4\pi\rho_k\sum_{i=1}^{i_{\max}}N_{ki}R_{ki}^2\frac{dR_{ki}}{dt}$$
$$\begin{cases} I_k=\dfrac{A_c}{1+\Theta}\left(\dfrac{2\gamma_l}{\pi m_a^3}\right)^{1/2}\dfrac{\rho_g^2}{\rho_l}\exp\left(-\dfrac{4\pi R_{k(\mathrm{cr})}^2\gamma_l}{3k_B T_k}\right) \\ \Theta=\dfrac{2(\kappa_g-1)}{\kappa_g+1}\dfrac{\Delta h}{\Re T_g}\left(\dfrac{\Delta h}{\Re T_g}-0.5\right) \\ R_{k(\mathrm{cr})}\approx\dfrac{2\gamma_l T_s}{\rho_k\Delta h\Delta T} \end{cases} \tag{31}$$

In Eq. (31), subscript k has the same definition as that used in Eq. (23), R_k is the radius of a bubble or droplet, $R_{k(\mathrm{cr})}$ is the Kelvin-Helmholtz critical nucleate radius, k_B is Boltzmann's constant, I_k is the nuclei generation rate of vapor bubbles or liquid droplets, A_c is the condensation coefficient, Θ is the nonisothermal correction factor, m_a is the mass of a single molecule of mercury, γ is the surface tension, $\mathfrak{R}$ is the gas constant, T_s is the saturation temperature, and subscript i is the value at each calculation cell. Δh denotes the latent heat, which is described by the difference in specific enthalpy between the liquid and gas phases, and is defined as $\Delta h = h_l - h_g$. The temperature difference between saturation temperature and gas-phase temperature ΔT is defined as $\Delta T = |T_s - T_k|$. N_{ki} denotes the number density of the generated vapor bubbles or condensed liquid droplets at each calculation cell i.

By introducing the formulation of the growth process for bubbles and condensed droplets, we assume that the growth rate of a bubble or droplet is controlled by the rate at which the enthalpy of vaporization or condensation can be conducted away from the bubble and droplets to the bulk liquid [39]. Under that assumption, the equation of the growth process for a single-vapor bubble and a condensed droplet is derived as

$$\Delta h \rho_k \frac{dR_{ki}}{dt} = \frac{p_k}{\sqrt{2\pi \mathfrak{R} T_g}} \frac{\kappa_k + 1}{2\kappa_k} c_{pk} \Delta T^{(i)} \tag{32}$$

where $\Delta T^{(i)}$ denotes the interfacial temperature between the vapor phase and the condensed droplet and is derived by the following equation:

$$\Delta T^{(i)} = \left(1 - \frac{R_{k(\mathrm{cr})}}{R_{ki}}\right) |T_s - T_k| \tag{33}$$

3 Numerical Conditions and Procedure

3.1 Numerical Conditions. As a practical example, we use the fluid properties of a mercury-based magnetic fluid with dispersed iron (Fe) particles [10]. The relation between temperature T and the normalized saturation magnetization $[(M_s/M_{s0}) = g(T)]$ is approximated by Brillouin function based on the measured value of the magnetization of Fe particles [40] and is defined by the following equation:

$$\begin{cases} \dfrac{M_s}{M_{s0}} = g(T) = A_0 \left\{ \left(\dfrac{2A_1+1}{2A_1}\right) \coth\left(\dfrac{2A_1+1}{2A_1}\right) \left[6 - 6\left(\dfrac{T}{T_c}\right)\right] \right. \\ \qquad \left. - \dfrac{1}{2A_1} \coth\left[\dfrac{6-6(T/T_c)}{2A_1}\right] \right\} \\ A_0 = 1, \quad A_1 = 1, \quad T_c = 1040.2 \end{cases} \tag{34}$$

where M_s is the saturation magnetization, M_{s0} is the spontaneous magnetization, and T_c is the Curie temperature of the ferromagnetic particles. The strength of magnetization M is expressed as a combined function of temperature $g(T)$ in Eq. (34) and the strength of magnetic field $f(H)$ as

$$\begin{cases} M = f(H) \cdot g(T) \\ f(H) = M_s \left[\coth(C_h) - \dfrac{1}{C_h}\right] \\ C_h = \dfrac{\mu_0 m H}{k_B T_{l(\mathrm{in})}} \end{cases} \tag{35}$$

where $f(H)$ has a type of Langevin function, $T_{l(\mathrm{in})}$ is the liquid-phase inlet temperature, and m is the magnetic moment of a ferromagnetic particle. Next, in order to consider the effect of nonuniform magnetic field as depicted in Fig. 2, the longitudinal distributions of magnetic field component H^{ζ} and the transverse distributions of magnetic field component H^{η} are derived by the following equations. In introducing H^{η} and H^{ζ}, we referred to the analytical solution of the nonuniform magnetic field distribution of the Helmholtz coil [41] and the measurement results of the magnetic field of the electromagnet, which were used in the previous experimental study [7,8],

$$\begin{cases} H^{\zeta} = H_{\max} \cdot \exp(-\xi_h^{*2}) \\ H^{\eta} = H_{\max} \cdot \eta^* \cdot |\xi_h^*| \cdot \exp(-\xi_h^{*2}) \\ \xi_h^* = \xi^* - \dfrac{1}{2}\xi_{\max}^* \end{cases} \tag{36}$$

where $H_{\max}$ is the maximum magnetic field strength; ξ^* denotes the normalized longitudinal coordinate, which is defined as $\xi^* = \xi/\xi_{\max}$; and η^* denotes the normalized transverse coordinate, which is defined as $\eta^* = \eta/\eta_{\max}$. The $H_{\max}$ on the electromagnet is installed in the position of the nozzle throat.

Furthermore, in order to consider the special characteristics of power generation of the MHD system, the power density G is defined by the following equation:

$$G = -\sigma_T K(K+1) \overline{\mu_l^{\xi}}^{\,2} B_T^{\zeta 2} \tag{37}$$

where $\overline{u_l^{\xi}}$ is the cross-sectional mean value of u_l^{ξ}.

3.2 Numerical Procedure. The finite difference method is used to solve the set of governing equations mentioned above. In the present calculation, the discrete forms of these equations are semi-implicitly obtained using a staggered grid. The grid is concentrated at the nozzle wall to capture the cavitation inception precisely.

The convective terms are discretized with a third-order QUICK scheme [42]. Also, the implicit fractional-step method [43] is used for time integration. Then the modified SOLA (numerical SOLution Algorithm for transient fluid flow) method of Tomiyama et al. [44], which is superior for the formulation and solution of a gas liquid two-phase flow problem, is applied for the numerical calculation. The numerical method can take into account the Neumann-type boundary condition in the iteration process of the pressure-correction equation. Also the effect of void fraction is implicitly taken into account in each iteration process [33]. The solution procedure for the pressure correction term δp_l is derived as follows. δp_l includes the effect of electromagnetic body force,

$$\delta p_l^{(l+1)} = \frac{A_r(\alpha_l^{(n)} - \Delta t \cdot D_T^{(l)} - 1)}{\left[\dfrac{1}{\rho_l c_{l0}^2}(\alpha_l^{(n)} - \Delta t \cdot D_T^{(l)}) - \dfrac{\Delta t^2}{\rho_l}\left(\dfrac{c_{l\xi}^{TP}}{\Delta \xi^2} + \dfrac{c_{l\eta}^{TP}}{\Delta \eta^2}\right)\right]} \tag{38}$$

where A_r is the relaxation factor and the optimized value of ($A_r = 1.965$) is employed [44]. c_{l0} is the sound velocity in liquid phase. The superscript (l) is the iteration number, (n) is the nth time level, Δt is the time increment, $\Delta\xi$, $\Delta\eta$ are the cell size in the ξ and η directions, respectively. $D_T^{(l)}$ and c_{lm}^{TP}, $(m = \xi, \eta)$ are derived as following equations:

$$D_T^{(l)} = \nabla_j(\alpha_l^{(l)} u_l^{j(l)}) + \frac{\alpha_g^{(l)}}{N_g} \nabla_j(N_g^{(l)} u_g^{j(l)}) + \frac{\alpha_g^{(l)}}{p_l^{(l)}} \frac{\partial p_l^{(l)}}{\partial t} \tag{39}$$

$$c_{lm}^{TP} = c_{lm+} \alpha_{lm+}^{(n)} + c_{lm-} \alpha_{lm-}^{(n)} \quad (m = \xi, \eta) \tag{40}$$

$$c_{lm\pm} = \begin{cases} 0 & \text{for } m\text{directional} \pm \text{side-cell boundary prescribes} \\ & \text{in the } m\text{directional velocity component} \\ 1 & \text{for the other cell boundaries (same order} \\ & \text{of double sign)} \end{cases}$$

where the subscripts + means the cell boundary at the positive directional side of the cell, − means the cell boundary at the negative directional side of the cell.

The liquid phase velocity u_l^i at the location of bubbles is calculated using an area-weighting interpolation method, which was used in the SMAC algorithm by Amsden and Harlow [45]. To

Table 1 Numerical Conditions

Liquid phase density	ρ_l	1.287×10^4	kg/m^3
Inlet pressure	$p_{l(\mathrm{in})}$	0.350	MPa
Outlet pressure	$p_{l(\mathrm{ex})}$	0.101	MPa
Inlet temperature	$T_{l(\mathrm{in})}$	573.15	K
Inlet width of duct	D	10.0	mm
Gas constant	$\mathfrak{R}$	41.45	J/(kg·K)
Maximum magnetic field strength	$H_{\max}$	600	kA/m
Magnetic moment of particle	m	7.96×10^{-19}	A·m^2
Permeability in vacuum	μ_0	$4\pi\times10^{-7}$	H/m
Load factor	K	-0.5	

determine the boundary conditions, the free-slip condition for prescribed liquid phase-velocity is applied to the central axis D-A, and the nonslip condition for prescribed liquid-phase velocity is applied to the sidewall C-B in Fig. 3. Also, a fully developed velocity profile is applied for liquid-phase velocities to the inlet cross-sectional area of the flow duct A-B. A convective outflow condition is applied for liquid-phase velocities to the exit section of the duct D-C. Adiabatic conditions are applied for thermal boundary conditions at the duct wall surface. The initial stationary condition of the liquid phase is assumed to be the pressurized fluid state.

The conditions for numerical analysis are listed in Table 1. For other physical properties used in constitutive equations, β_l, σ_l, λ_l, and γ_l are given as functions of temperature. The saturation temperature T_s is given as a function of pressure. The required physical properties of the liquid phase are given by the tables of the thermophysical properties of mercury [46].

The interval of each time step is automatically adjusted during the computation to satisfy the CFL condition [44]. We actually calculated solutions on three different grid densities: 50×120, 60×150, and 100×220 nodes.

Figure 4 shows the convergence histories for three sets of computational grids. For all cases, the magnitude of the residuals decrease by at least two orders of magnitude within 5000 iterations. For a finer grid, convergences slow down, as expected. For all three grids, the iteration errors and uncertainties are assumed to be negligible in comparison with the grid errors. Since iterative errors are negligible, correction of solutions for iterative error is not required. As a result, we found that each numerical result shows almost the same profile; the grid independence of the numerical results was confirmed. Thus, as a compromise between computer memory and accuracy, we chose to use the 60×150 structured grid in the ξ and η directions for the calculations. The simulations have been performed on a single Pentium IV 2.7 GHz processor of an IBM-compatible computer, and the required memory size for each time step is about 1.0 GB with a grid density of 60 ×150. The required total CPU time has been about 120 min for one operating condition. During the execution of the unsteady calculation, no significant differences in the mean-flow profiles are found in the 1200–1500 time steps. We determined that the cavitating flow almost reaches steady state when such flow profiles are obtained.

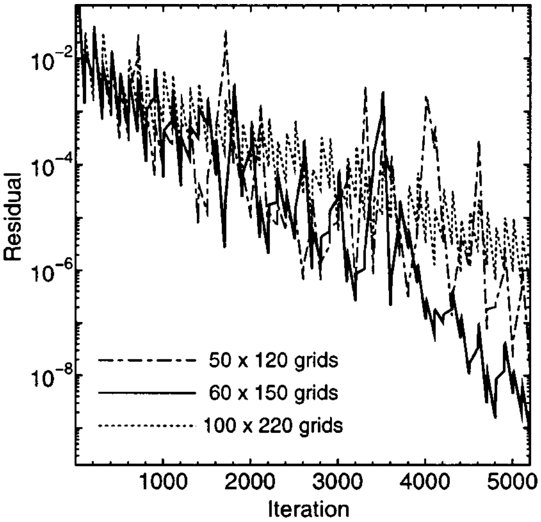


Fig. 4 Convergence histories for three sets of grids

The present calculation procedure is specifically employed as follows:

1. The translational velocity of bubbles is calculated by Eqs. (11) and (24).
2. The liquid-phase velocity is calculated by Eq. (9).
3. The phase generation density and bubble radius are calculated by Eqs. (31) and (32).
4. The void fraction is calculated by Eqs. (23) and (1).
5. The pressure correction term δp_l is calculated by Eq. (38) in application with modified SOLA method. When the residual becomes sufficiently small value, the calculation for obtaining the new time-level solution is executed by returning to step 1 of the procedure until a steady-state flow field is obtained.

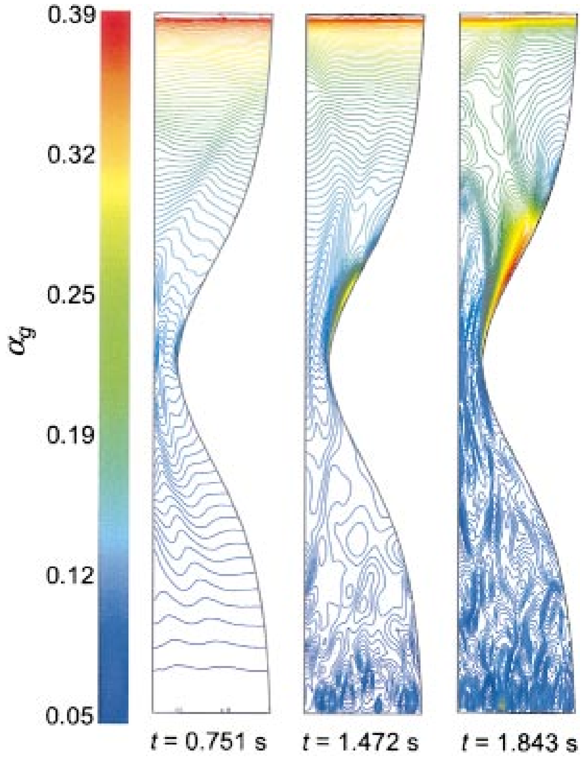
4 Results and Discussion

Figure 5 shows the numerical results of the transient evolution of the void fraction α_g contour, and Fig. 6 shows the instantaneous liquid-phase pressure p_l contour. According to Fig. 5, the cavitation inception effectively occurs at the position of the divergent nozzle throat. The cavitation inception or formation of cloud cavity in ECMF flow at the diverging nozzle throat is suppressed compared with that of mercury flow due to the magnetic body force under a sharp magnetic field gradient, compare to that of mercury flow. It was found that the growth rate of the volume fraction of the cavitation bubble is lower than that of mercury flow. Focusing on ECMF flow, immediately after the flow is initially induced, taking note of the primary feature of the void fraction profile under a strong magnetic field, the void fraction profile elongated in the longitudinal direction of negative magnetic field gradient because the bubbles are accelerated and migrate due to the two-phase magnetic body force in the direction of the negative magnetic field gradient. The effect of two-phase magnetic body force is characterized by the second right-hand term in the momentum equation, Eq. (9).

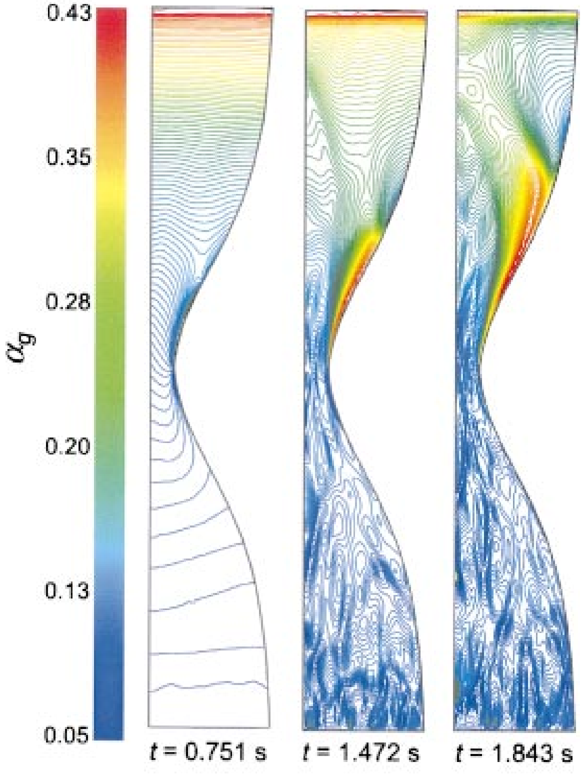
With time elapses, because the magnetic body force based on the transverse magnetic field gradient acts so that the bubbles migrate from the wall into the center of the duct, it becomes clear that the void fraction is locally increasing near the central axis. Downstream of the nozzle throat, although both magnetic body force and the Lorentz force act as a flow resistance, the forces are decreased due to the gas-phase inclusion. In the two-phase flow region, not only the decreasing effect of flow resistance due to the Lorentz force, but also the pressure rise effect caused by the unbalance of the magnetic body force between the single- and two-phase flow region is obtained.

Furthermore, additional lift force operates in the direction, which causes the bubbles to migrate in the duct-wall direction. However, as the two-phase magnetic body force begins to dominate the bubbles, it is found that the large volume fraction region of gas phase moves to the central axis of the duct. According to this result, the gas-phase motion is controlled not only by the buoyancy force, but also due to the liquid-phase pressure gradient, additional lift forces, and especially due to the two-phase magnetic body force that acts on the cavitation bubbles.

If the phenomenon applicable to the conditions of both ECMF flow and mercury flow is as explained here, then the gas-phase

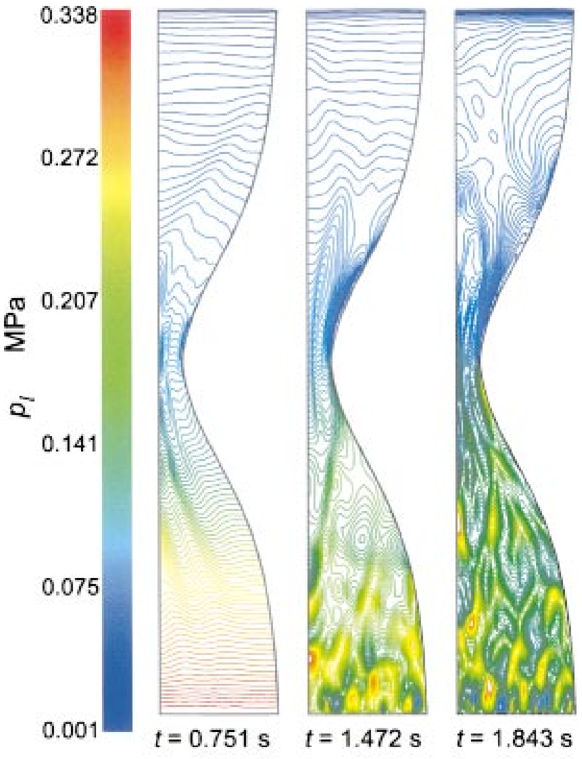


(a) ECMF

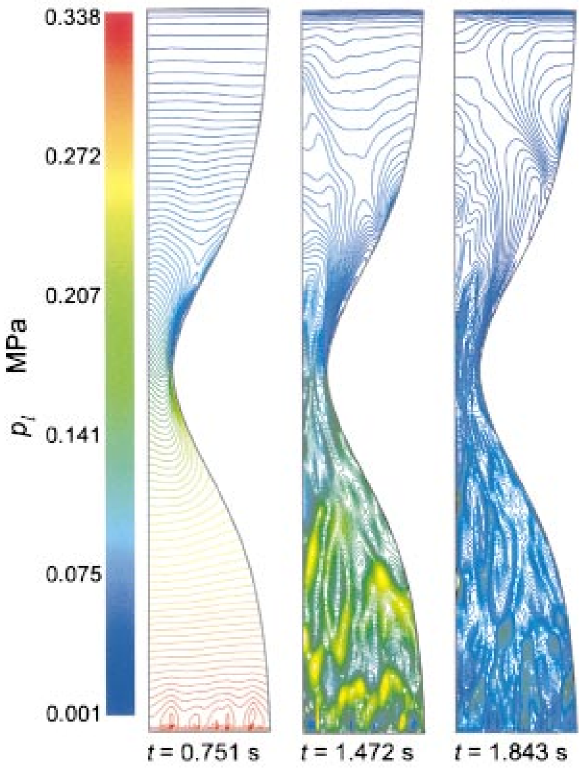


(b) Mercury

Fig. 5 Time evolution of void fraction distributions



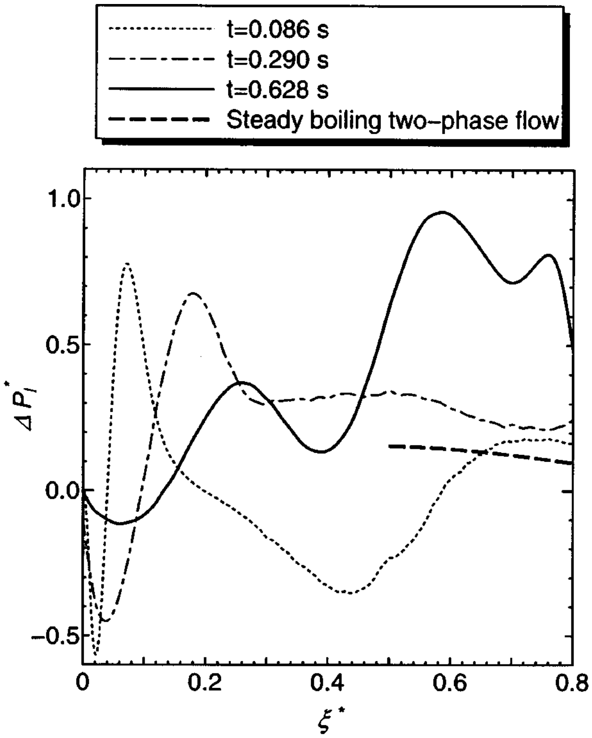
(a) ECMF



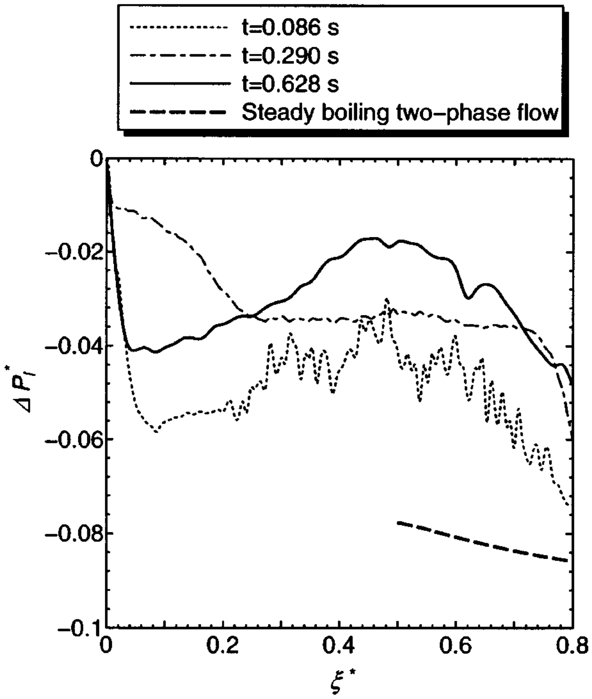
(b) Mercury

Fig. 6 Instantaneous liquid-phase pressure contours

volume fraction increases at the throat position and is concentrated to form a small cavity cloud downstream of the divergent nozzle due to the small vortex induced by the wake passing through the nozzle throat, which is based on the effect of the negative pressure gradient. When the magnitude of the cloud cavity is above a certain size, due to the reentrant jet resulting from the boundary layer separation, the cavity becomes detached from the cloud and then remains in the high-volume fraction region as



(a) ECMF



(b) Mercury

Fig. 7 Pressure rise effect of ECMF flow and mercury flow in the longitudinal direction

the gas phase moves downstream. Especially in the case of ECMF flow, because the two-phase magnetic ejection effect acts on the cloud cavity in the negative magnetic field gradient region, the separation of the cloud cavity is enhanced by the two-phase magnetic body force acting on the cavitation bubbles. Due to the suppression effect and magnetic ejection effect on the cavitation bubbles, the magnitude of the cloud cavity in ECMF flow becomes smaller than that of mercury flow.

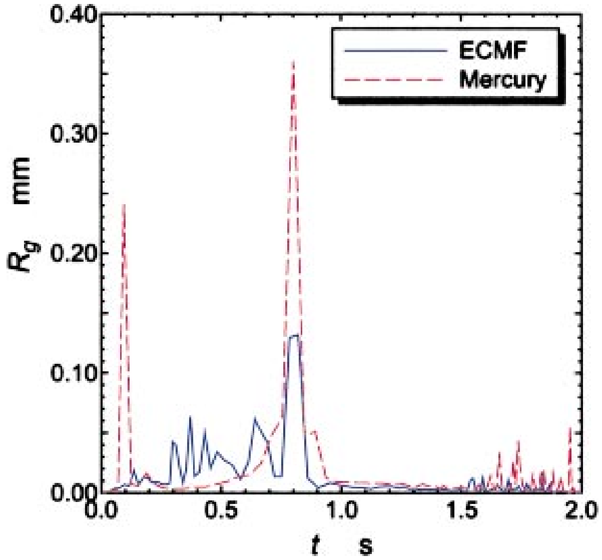


Fig. 8 Fluctuations of bubble radius as a function of time

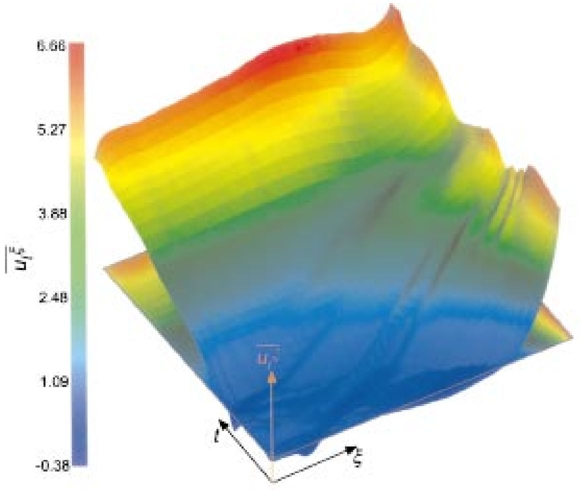
The pressure around the inlet section shows a little nonuniformity with time due to the propagation of the pressure fluctuation from the nozzle throat section to the inlet section of the duct. The pressure fluctuation at the nozzle throat is caused by the effects of sudden cavitation generation and rapid bubble growth with time, also by the high-speed inflow of liquid phase into the throat section.

Figure 7 shows the present numerical results of the liquid-phase pressure rise effect of ECMF flow and mercury flow in the longitudinal direction, in comparison with the previous one-dimensional numerical result of its effect for the steady boiling two-phase pipe flow of ECMF [14]. In this figure, the axis of the ordinate denotes the normalized pressure-rise parameter Δp_l^* and is derived by the following equation:

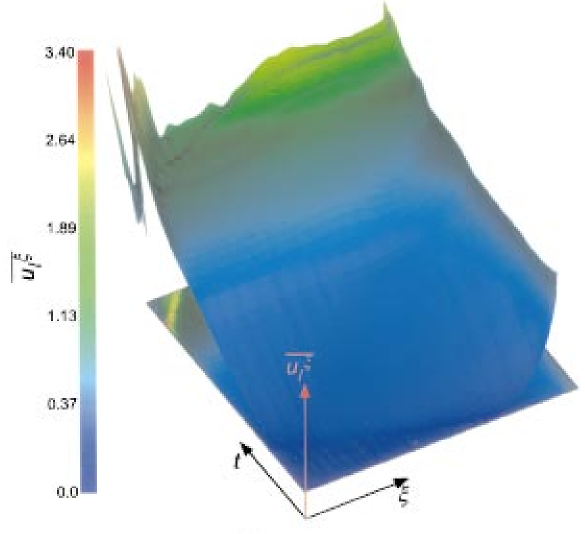
$$\Delta p_l^* = \frac{\overline{p_{l(\mathrm{eff})}} - p_{l(\mathrm{in})}}{p_{l(\mathrm{in})}} \tag{41}$$

where $\overline{p_{l(\mathrm{eff})}}$ is the cross-sectional mean-effective driving pressure from which the influence on the prudence of the liquid-phase fluid can be deducted. $p_{l(\mathrm{in})}$ is the inlet pressure. The axis of the abscissa in Fig. 7 denotes the normalized longitudinal coordinate $\xi^*(=\xi/\xi_{\max})$. Focusing on the present numerical result of ECMF cavitating flow, the pressure rise effect based on the two-phase magnetic body force under an applied magnetic field region acts effectively with the progress of time. However, as time elapses and the flow approaches steady state, it is found that the pressure rise effect, which is based on cavitating two-phase magnetic body force, decreases around the exit section of the duct. The magnitude of the pressure rise effect in ECMF flow is indicated to have a larger value than that in mercury flow. Especially in the case of mercury flow, as for the limited portion of the cross-sectional area in a duct, there is a possibility of the pressure rise effect being obtained, however, when cross-sectional mean value of Δp_l^* is taken into account, the Δp_l^* in mercury flow exhibits negative value because of the strong effect of the Lorentz force as the flow resistance. Also, it may be possible to suppose that under the optimized condition possible to obtain the effective pressure rise of cavitating ECMF flow, it is difficult to realize the sufficient pressure rise effect making use of the cavitating mercury flow.

Comparing our previous results regarding the pressure distribution of the boiling two-phase ECMF flow [14] with the present



(a) ECMF

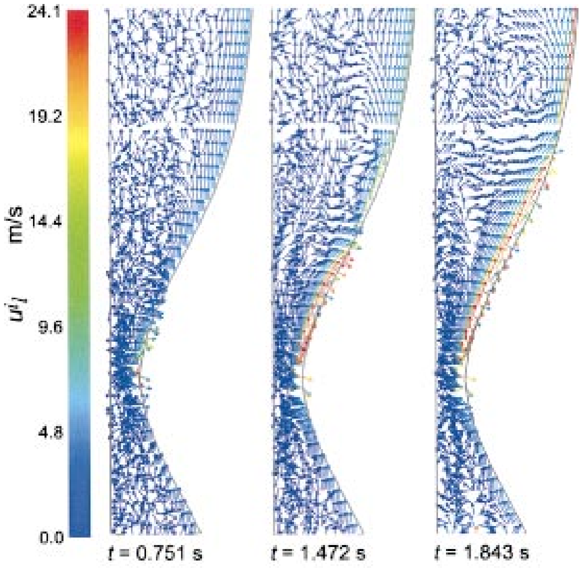


(b) Mercury

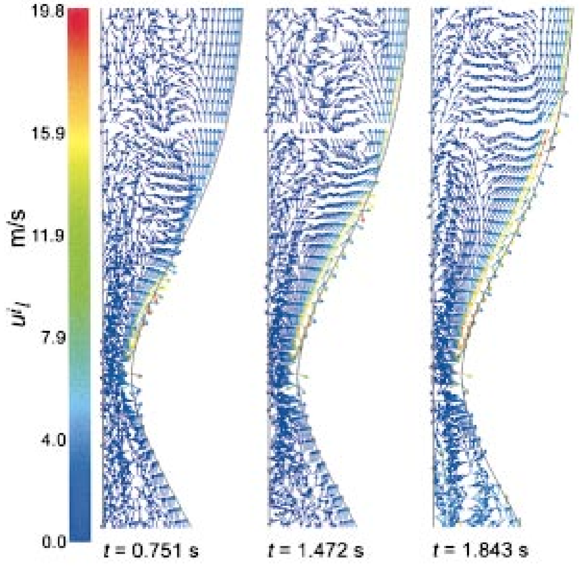
Fig. 9 Effect of magnetic acceleration on the longitudinal liquid-phase cross-sectional mean velocity as a function of time

numerical results, it is found that the present ECMF acceleration system, which uses cavitation, can achieve a greater pressure rise effect. Therefore, when the LMMHD power generation system, which uses two-phase magnetic fluid, is employed, rather high pressure rise effect resulting from the use of cavitation for two-phase flow production can be obtained.

Figure 8 shows the fluctuation of bubble radius (R_g) as a function of the time at position E (as depicted in Fig. 3, $\xi = 0.5\xi_{\max}$), which is just a quarter of the nozzle throat width, where the cavitation actively occurs. It is found that the magnitude of transient displacement of R_g in ECMF flow shows a smaller value and that the size of R_g becomes homogeneous compared with that of mercury flow because the magnetic body force acts to suppress the expanding cavitation bubbles. From Figs. 5–8, it is clarified that the decrease of p_l induces an increase of α_g and that the expansion or contraction of bubble radius R_g corresponds to the change of p_l. Since the displacement magnitude of R_g has a small value, it is also clarified that the generated cavitation bubbles maintain a small size in the vaporization process and in the initial cavitating flow state.



(a) ECMF

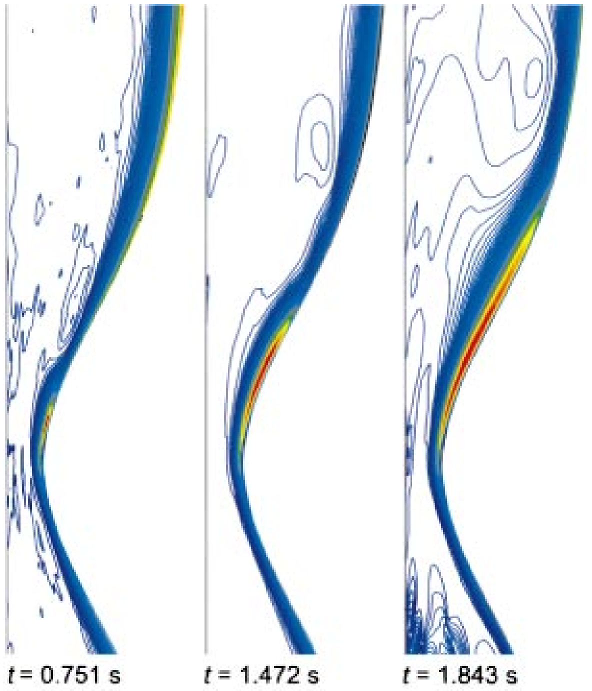


(b) Mercury

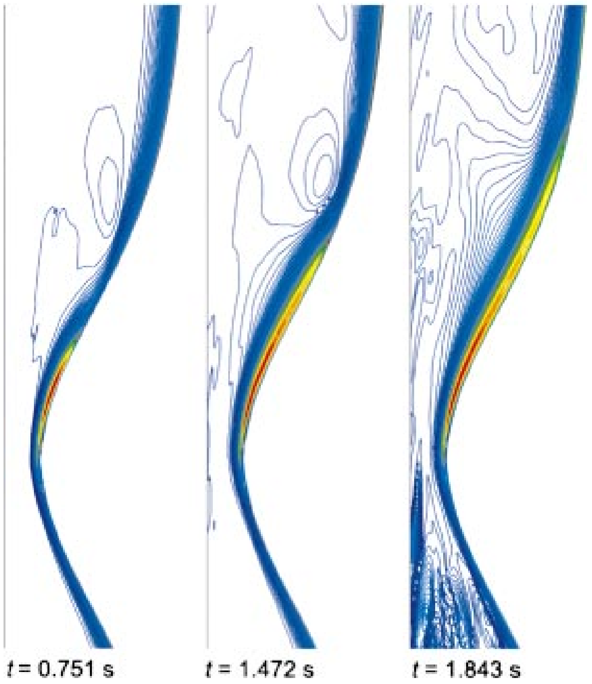
Fig. 10 Instantaneous liquid-phase velocity vector (enlarged view of nozzle throat section)

Figure 9 shows the effect of magnetic acceleration on the longitudinal liquid phase cross-sectional mean velocity, $\overline{u_l^\xi}$, as a function of time t. In the case of ECMF flow, the fluid acceleration effect, which is based on the two-phase magnetic body force, increases with the lapse of time. It is especially found that $\overline{u_l^\xi}$ sharply increases in the flow field and becomes two-phase flow state, and that $\overline{u_l^\xi}$ tends to express the maximum value at the exit section of the duct. In the case of mercury, it is found that $\overline{u_l^\xi}$ increases with increase in t. However, the magnitude of $\overline{u_l^\xi}$ exhibits a lower value than that of $\overline{u_l^\xi}$ in ECMF flow. Furthermore, it is found that the rate of increase on $\overline{u_l^\xi}$ in the ξ direction of ECMF flow shows a greater value than that of mercury flow.

It is found that the fluctuation of $\overline{u_l^\xi}$ in mercury flow in the vicinity of the inlet section sharply increases with maximum time step because of the propagation of the velocity and pressure fluc-



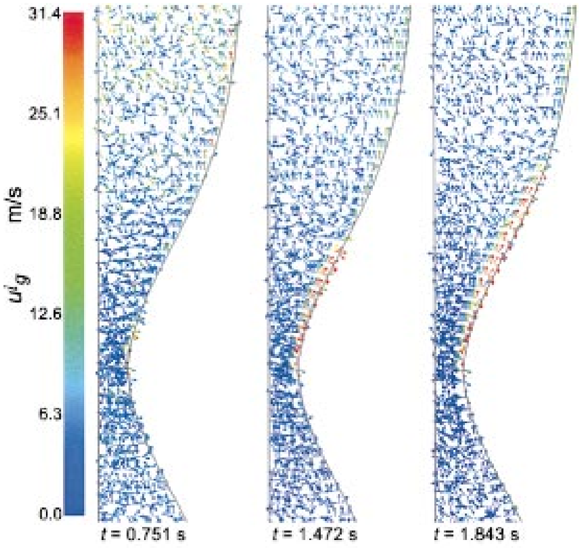
(a) ECMF



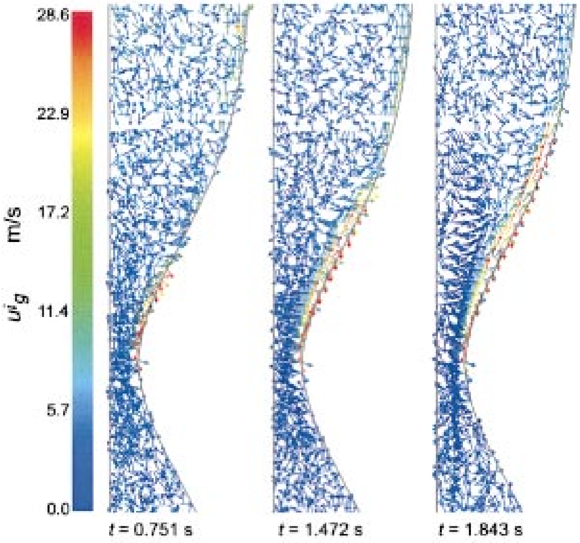
(b) Mercury

Fig. 11 Instantaneous liquid-phase stream lines (enlarged view of nozzle throat section)

tuations from the nozzle throat section to the inlet section of the duct for a similar reason as the generation of inhomogeneity in inlet pressure. Contrarily, in the case of ECMF flow, the sharp increase in the fluctuation of $\overline{u_l^{\xi}}$ cannot be found at the inlet section with time progress because of the two-phase magnetic stabilizing effect due to the magnetization of the fluid [15].



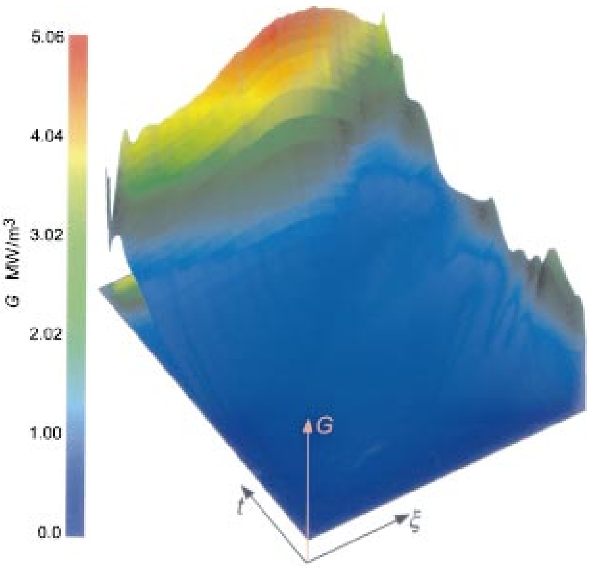
(a) ECMF



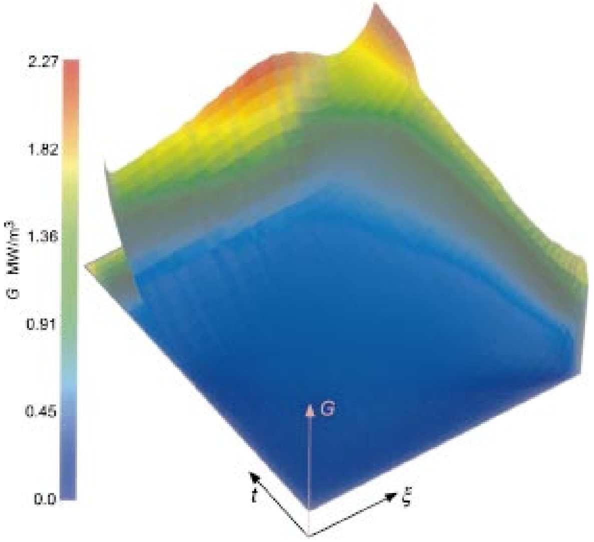
(b) Mercury

Fig. 12 Instantaneous gas-phase velocity vector (enlarged view of nozzle throat section)

Figure 10 shows the profiles of the liquid-phase velocity u_l^i, and Fig. 11 shows instantaneous liquid-phase stream lines around the throat section in ECMF and mercury flow. The flow separation and backward flow of u_l^i locally occur in the vicinity of the wall of the throat section upstream of the point of cavitation inception. In the case of ECMF flow, the liquid phase is locally accelerated around the throat wall section, and the magnitude of u_l^i has a larger value compared with the case of mercury flow because of the two-phase magnetic acceleration effect, which acts on the liquid phase due to the strong magnetic field gradient. Furthermore, because of the increase in momentum exchange between the gas and liquid phases, the magnitude of u_l^i locally increases in the region where the cavitation is actively generated. According to these numerical results on velocity profile, it is found that rather than utilizing only the pumping effect of the bubbles when employing the two-phase LMMHD fluid-driving system, the method



(a) ECMF



(b) Mercury

Fig. 13 Power density profile in the longitudinal direction as a function of time

that utilizes the two-phase magnetic body force generated by cavitating magnetic fluid flow can obtain the enhanced fluid acceleration effect.

Taking focus around the central axis region at the nozzle throat, the liquid-phase longitudinal velocity component is decelerated because of the Lorentz force extensively acting as a flow resistance due to the strong magnetic field strength and small void fraction at the throat section. The backward liquid-phase flow at the central axis induced by the cavity separation from the throat wall is the another reason for longitudinal velocity deceleration. Especially in the case of ECMF, the magnetic body force also locally acts as a flow resistance in the central axis region at the nozzle throat due to the strong magnetic field gradient and small void fraction. Therefore, the liquid-phase longitudinal velocity component at the region is decelerated, and the two-phase magnetic driving force cannot effectively act on the fluid due to small void fraction. However, in the throat wall vicinity where cavitation actively generates, the liquid-phase velocity is strongly accelerated because of the two-phase magnetic driving force and bubbles' pumping effect acting on the liquid phase. The slip ratio u_g^i/u_l^i in the vicinity of the throat wall section tends to have a large value.

Figure 12 shows the instantaneous gas-phase velocity u_g^i around the throat section in ECMF flow and mercury flow. In the initial flow state, it is found that the backward flow of u_g^i is generated upstream of the throat due to the effect of the separation wake of liquid phase in the vicinity of the throat wall and that the gas phase is dispersed throughout the downstream region of the throat section. In the case of ECMF flow, the gas phase is locally accelerated in the direction of the negative magnetic field gradient, and the magnitude of u_g^i is indicated to have a larger value compared to the case of mercury flow because of the magnetic ejection effect, which acts on the bubbles. Considering Figs. 12 and 5 together, in the strong magnetic field gradient region, the void fraction α_g locally decreases due to application of the nonuniform magnetic field because the gas phase is locally accelerated and the slip ratio increases due to the magnetic ejection effect on bubbles in the two-phase region.

In addition to the formation of the cavity vortex and its growth in ECMF, the advection of the cavity cloud is enhanced by the magnetic ejection effect in the downstream region of the throat. With time, the gas-phase motion exhibits diffusing behavior, and the u_g^i profile takes on a different aspect from the liquid-phase velocity profiles u_l^i. The characteristic gas-phase behavior in the magnetic fluid is not only due to the several additional forces that appear in Eq. (11), but also to the two-phase magnetic body force and other forces that act on the bubbles due to the two-phase Lorentz force that are included in the momentum terms in Eq. (9). According to the numerical results on gas phase behavior, it is clarified that the precise control of bubble motion and control of the two-phase MHD flow is possible by practical use of the characteristic magnetization of fluid inherent ECMF.

Figure 13 shows the profiles of power density G in the ξ direction as a function of time in ECMF flow and mercury flow. The magnitude of the power density G of ECMF flow has a larger value than that of mercury flow. Especially if the flow field becomes two-phase flow state, the rate of increase of G in ECMF flow with time is greater than that in the mercury flow. The higherpower density in ECMF flow is due to the fluid acceleration resulting from the two-phase magnetic body force under the applied magnetic field region. According to these results, it is found that the present ECMF system, which uses cavitation, can obtain higher power density than the mercury system in the two-phase LMMHD power generation systems because of the practical use of magnetization in the working fluid.

5 Conclusions

The two-dimensional cavitating flow characteristics of electrically conducting magnetic fluid (ECMF) in a converging-diverging nozzle under a strong nonuniform traverse magnetic field were numerically predicted to realize the further development and high performance of the two-phase LMMHD power generation systems. The main results obtained can be summarized as follows:

1. Inception of cavitation or formation of a cloud cavity in ECMF flow in the diverging nozzle throat is more greatly suppressed than in the case of mercury flow due to the magnetic body force with a sharp magnetic field gradient. Especially in the case of ECMF flow, because the two-phase magnetic ejection effect acts on the cloud cavity in the negative magnetic field gradient region, separation of the cloud cavity is enhanced by the two-phase magnetic body force acting on the cavitation bubbles.
2. The pressure rise effect due to the two-phase magnetic body force under an applied magnetic field region effectively acts with the progress of time in the two-phase region. The magnitude of

the pressure rise effect in cavitating ECMF flow has a larger value than that in cavitating mercury flow or boiling two-phase magnetic fluid flow.

3. It is found that rather than utilizing only the pumping effect of bubbles when employing the LMMHD two-phase fluid driving system, the method that utilizes the two-phase magnetic body force generated by cavitating magnetic fluid flow can obtain the enhanced fluid acceleration effect.

4. The present ECMF system, which uses cavitation, can obtain a power density greater than that with the mercury system in the two-phase LMMHD power generation systems because of the practical use of the magnetization in the working fluid.

Acknowledgments

This research was supported by a Grant-in-Aid for Scientific Research (Contract No. 15760099) from the Ministry of Education, Science and Culture, Japan; Tohoku Intelligent Cosmos Promotion Council Secretariat Foundation, Japan; and Scientific International Promotion Foundation at Hirosaki University (Contract No. A-1-1), Japan.

Nomenclature

$a^{(i)}$ = interfacial area concentration per unit volume
B = magnetic flux density
B^i = contravariant vector of magnetic flux density
$\mathrm{C_D}$ = drag coefficient
$\mathrm{C_V}$ = virtual mass coefficient
c_p = specific heat at constant pressure
c_0 = sound velocity
D = inlet width of duct
E = strength of electric field
E^i = contravariant vector of electric field
e = specific internal energy
e^{ijk} = permutation symbol
g_r^i = contravariant vector of gravitational acceleration
g^{ij} = fundamental metric tensor
H = strength of magnetic field
H^i = contravariant vector of magnetic field
$H_{\max}$ = maximum strength of magnetic field
h = specific enthalpy
J = current density
J^i = contravariant vector of current density
K = load factor
k = heat transfer rate
k_B = Boltzmann's constant
M = strength of magnetization
M^i = contravariant vector of magnetization
M_s = saturation magnetization
M_{s0} = spontaneous magnetization
m = magnetic moment
m_a = mass of a single molecule
N = number density
p = absolute pressure
q = heat flux
q^i = contravariant heat flux vector
R = radius
$\mathrm{R_m}$ = magnetic Reynolds number
$\mathfrak{R}$ = gas constant
T = absolute temperature
T_c = Curie point temperature
T_s = saturation temperature
t = time
Δt = time increment
u^i, u^j, u^k = contravariant velocity
u_i, u_j, u_k = covariant velocity

Greek Symbols

α = volume fraction
β = dynamic viscosity
Γ = phase generation density
γ = surface tension
η = transverse coordinate
κ = ratio of specific heat
λ = thermal conductivity
μ_l = magnetic permeability in a fluid
μ_0 = magnetic permeability in vacuum
ν = kinematic viscosity
ξ = longitudinal coordinate
ρ = density
σ = electrical conductivity
Ω^i = contravariant angular velocity vector
ω^i = contravariant vorticity vector
∇_j = covariant differential

Subscripts

c = condensation
e = evaporation
(ex) = exit section of the duct
g = gas phase
i, j, k = covariant component
(in) = inlet section of the duct
l = liquid phase
s = saturation
T = two-phase

Superscripts

i, j, k = contravariant component
(i) = interface
(l) = iteration number
(n) = nth time level
ξ = contravariant component in the ξ-direction
η = contravariant component in the η-direction
ζ = contravariant component in the ζ-direction
$\overline{(\ )}$ = mean value

References

[1] Petrick, M., and Branover, H., 1985, ''Liquid Metal MHD Power Generation: Its Evolution and Status,'' Prog. Astronaut. Aeronaut., **100**, pp. 371–400.
[2] Alexander, Y., and Branover, H., 1982, ''An Analytical Model of a Two-Phase Liquid Metal Magnetohydrodynamic Generator,'' Phys. Fluids, **25**, pp. 446–451.
[3] Dobran, F., 1981, ''Analysis of Two-Phase Flow Magnetohydrodynamic Generator Performance in Terms of Flow and Electrical Conductivity Distribution Parameters,'' Int. J. Multiphase Flow, **7**, pp. 595–617.
[4] Eckert, S., Gerbeth, G., and Lielausis, O., 2000, ''The Behavior of Gas Bubbles in a Turbulent Liquid Metal Magnetohydrodynamic Flow, Part I: Dispersion in Quasi-Two-Dimensional Magnetohydrodynamic Turbulence,'' Int. J. Multiphase Flow, **26**, pp. 45–66.
[5] Eckert, S., Gerbeth, G., and Lielausis, O., 2000, ''The Behavior of Gas Bubbles in a Turbulent Liquid Metal Magnetohydrodynamic Flow, Part II: Magnetic Field Influence on the Slip Ratio,'' Int. J. Multiphase Flow, **26**, pp. 67–82.
[6] Eckert, S., Gerbeth, G., Mihalache, G., and Thibault, J.-P., 1997, ''Influence of External Magnetic Fields on Slip Ratio in LMMHD Two-Phase Flow,'' Magnetohydrodynamics (N.Y.), **33**, pp. 239–247.
[7] Ishimoto, J., Okubo, M., Nishiyama, H., and Kamiyama, S., 1995, ''Basic Study on an Energy Conversion System Using Gas-Liquid Two-Phase Flows of Magnetic Fluid (Analysis on the Mechanism of Pressure Rise),'' JSME Int. J., Ser. B, **39**, pp. 72–79.
[8] Kamiyama, S., and Ishimoto, J., 1995, ''Boiling Two-Phase Flows of Magnetic Fluid in a Nonuniform Magnetic Field,'' J. Magn. Magn. Mater., **149**, pp. 125–131.
[9] Ishimoto, J., 2002, ''Cavitating Flow of Magnetic Fluid in Converging-Diverging Nozzle Under Nonuniform Magnetic Field,'' Magnetohydrodynamics, **38**, pp. 307–317.
[10] Charles, S. W., and Popplewell, J., 1980, ''Progress in the Development of Ferromagnetic Liquids,'' IEEE Trans. Magn., **16**, pp. 172–177.
[11] Fedonenko, A. I., and Smirnov, V. I., 1983, ''Particle Interaction and Clumping an Electrically Conducting Magnetic Fluid,'' J. Magn. Magn. Mater., **19**, pp. 388–391.

[12] Shepherd, P. G., and Popplewell, J., 1971, ''Ferrofluids Containing Ni-Fe Alloy Particles,'' Philos. Mag., **23**, pp. 239.

[13] Alekseev, V. A., 1991, ''Structural Transformations in an Electrically Conducting Ferrocolloid,'' Magnetohydrodynamics (N.Y.), **27**, pp. 18–22.

[14] Okubo, M., Ishimoto, J., Nishiyama, H., and Kamiyama, S., 1993, ''Analytical Study on Two-Phase MHD Flow of Electrically Conducting Magnetic Fluid,'' Magnetohydrodynamics (N.Y.), **29**, pp. 291–297.

[15] Ishimoto, J., Okubo, M., and Kamiyama, S., 1995, ''Effect of Magnetic Field on the Stability of Boiling Two-Phase Flows of Magnetic Fluid,'' *Proc. 2nd Int. Conf. Multiphase Flow*, A. Serizawa, ed., Kyoto, Japan, Vol. 4, pp. FC17–FC24.

[16] Kamiyama, S., and Yamasaki, T., 1975, ''Study on Magnetohydrodynamic Flow, Part I (the Effect of Magnetic Field on Cavitation of Mercury Flow),'' Report of the Institute of High Speed Mechanics, Tohoku Univ., **36**(344), pp. 1–16.

[17] Kataoka, I., and Serizawa, A., 1989, ''Basic Equations of Turbulence in Gas-Liquid Two-Phase Flow,'' Int. J. Multiphase Flow, **15**, pp. 843–855.

[18] Harlow, F. H., and Amsden, A. A., 1975, ''Numerical Calculation of Multiphase Fluid Flow,'' J. Comput. Phys., **17**, pp. 19–52.

[19] Yamamoto, S., Hagari, H., and Murayama, M., 1997, ''Numerical Simulation of Condensation around the 3-d Wing,'' Trans. Jpn. Soc. Aero. Space Sci., **138**, pp. 182–189.

[20] Young, J. B., 1992, ''Two-Dimensional, Nonequilibrium, Wet-Stream Calculations for Nozzles and Turbine Cascades,'' ASME J. Turbomach., **114**, pp. 569–579.

[21] Ueno, K., 1991, ''Inertia Effect in Two-Dimensional MHD Channel Flow Under a Traveling Sine Wave Magnetic Field,'' Phys. Fluids A, **3**, pp. 3107–3116.

[22] Ueno, K., 1993, ''Effect of Turnaround Lines of Magnetic Flux in Two-Dimensional MHD Channel Flow Under a Traveling Sine Wave Magnetic Field,'' Phys. Fluids A, **5**, pp. 490–492.

[23] Shu, D., Sun, B., Wang, J., Li, T., Xu, Z., and Zhou, Y., 2000, ''Numerical Calculation of the Electromagnetic Expulsive Force Upon Nonmetallic Inclusions in an Aluminum Melt,'' Metall. Mater. Trans. B, **31B**(6), pp. 1527–1533.

[24] Meir, A. J., and Schmidt, P. G., 1999, *Fluid Flow Phenomena in Metals Processing*, N. El-Kaddah et al., ed., TMS, San Diego.

[25] Leenov, D., and Kolin, A., 1954, ''Theory of Electromagnetophoresis. I. Magnetohydrodynamic Forces Experienced by Spherical and Symmetrically Oriented Cylindrical Particles,'' J. Chem. Phys., **22**(4), pp. 683–688.

[26] Tanatugu, N., Fujii-E, Y., and Suita, T., 1972, ''Electrical Conductivity of Liquid Metal Two-Phase Mixture in Bubbly and Slug Flow Regime,'' J. Nucl. Sci. Technol., **9**, pp. 753–755.

[27] Shizawa, K., and Tanahashi, T., 1987, ''A New Complete Set of Basic Equations for Conducting Magnetic Fluids With Internal Rotation (Derivation by Thermodynamical Method),'' J. Magn. Magn. Mater., **65**, pp. 181–184.

[28] Shizawa, K., and Tanahashi, T., 1986, ''New Constitutive Equations for Conducting Magnetic Fluids With Internal Rotation (Thermodynamical Discussions),'' Bull. JSME, **29**, pp. 2878–2884.

[29] Cross, M. M., 1975, ''Viscosity-Concentration-Shear Rate Relations for Suspensions,'' Rheol. Acta, **14**, pp. 402–403.

[30] Tomiyama, A., Zun, I., Higaki, H., Makino, Y., and Sakaguchi, T., 1997, ''A Three-Dimensional Particle Tracking Method for Bubbly Flow Simulation,'' Nucl. Eng. Des., **175**, pp. 77–86.

[31] Fan, L. S., and Zhu, C., 1998, *Principles of Gas-Solid Flows*, Cambridge University Press, New York.

[32] Patankar, N. A., and Joseph, D. D., 2001, ''Modeling and Numerical Simulation of Particle Flows by the Eulerian-Lagrangian Approach,'' Int. J. Multiphase Flow, **27**, pp. 1659–1684.

[33] Murai, Y., and Matsumoto, Y., 2000, ''Numerical Study of the Three-Dimensional Structure of a Bubble Plume,'' ASME J. Fluids Eng., **122**, pp. 754–760.

[34] Auton, T. R., 1988, ''The Force Exerted on a Body in Invisid Unsteady Nonuniform Rotational Flow,'' J. Fluid Mech., **197**, pp. 241–257.

[35] Clift, R., Grace, J. R., and Weber, M. E., 1978, *Bubbles, Drops, and Particles*. Academic Press, San Diego.

[36] Dobran, F., 1988, ''Liquid and Gas-Phase Distributions in a Jet With Phase Change,'' ASME J. Heat Transfer, **110**, pp. 955–960.

[37] Solbrig, C. W., McFadden, J. H., Lyczkowski, R. W., and Hughes, E. D., 1978, ''Heat Transfer and Friction Correlations Required to Describe Steam-Water Behavior in Nuclear Safety Studies,'' AIChE Symp. Ser., **174**, pp. 100–128.

[38] Hirt, C. W., and Romero, N. C., 1975, Application of a Drift Flux Model to Flashing in Straight Pipes, LA-6005-MS, Los Alamos Scientific Laboratory Report.

[39] Moses, C. A., and Stein, G. D., 1978, ''On the Growth of Steam Droplets Formed in Laval Nozzle Using Both Static Pressure and Light Scattering Measurements,'' ASME J. Fluids Eng., **100**, pp. 311–322.

[40] Crangle, J., and Hallam, G. C., 1963, ''The Magnetization of Face-centred Cubic and Body-centred Cubic Iron+Nickel Alloys,'' Proc. Phys. Soc., London, Sect. A, **272**, pp. 119.

[41] Otis, D. R., 1966, ''Computation and Measurement of Hall Potentials and Flow-field Perturbations in Magnetogasdynamic Flow of an Axisymmetric Free Jet,'' J. Fluid Mech., **24**, pp. 41–63.

[42] Leonard, B. P., 1979, ''A Stable and Accurate Convection Modelling Procedure Based on Quadratic Upstream Interpolation,'' Comput. Methods Appl. Mech. Eng., **19**, pp. 59–98.

[43] Rosenfeld, M., Kwak, D., and Vinokur, M., 1991, ''A Fractional Step Solution Method for the Unsteady Incompressible Navier-Stokes Equations in Generalized Coordinate Systems,'' J. Comput. Phys., **94**, pp. 102–137.

[44] Tomiyama, A., and Hirano, M., 1994, ''An Improvement of the Computational Efficiency of the SOLA Method,'' JSME Int. J., Ser. B, **37**, pp. 821–826.

[45] Amsden, A. A., and Harlow, F. H., 1970, The SMAC Method: A Numerical Technique for Calculating Incompressible Fluid Flows, LA-4370, Los Alamos Scientific Laboratory Report.

[46] JSME, 1997, *Thermophysical Properties of Fluids*, Japan Soc. Mech. Eng., Tokyo.

K. M. Liew
Professor,
Nanyang Centre for Supercomputing and Visualisation,
School of Mechanical and Production Engineering,
Nanyang Technological University,
Singapore 639798
Fellow ASME

J. Yang
Department of Civil Engineering,
The University of Queensland,
St. Lucia, Brisbane,
Queensland 4072, Australia

S. Kitipornchai
Professor,
Department of Building and Construction,
City University of Hong Kong,
Kowloon, Hong Kong

Thermal Post-Buckling of Laminated Plates Comprising Functionally Graded Materials With Temperature-Dependent Properties

This paper presents thermal buckling and post-buckling analyses for moderately thick laminated rectangular plates that contain functionally graded materials (FGMs) and subjected to a uniform temperature change. The theoretical formulation employs the first-order shear deformation theory and accounts for the effect of temperature-dependent thermoelastic properties of the constituent materials and initial geometric imperfection. The principle of minimum total potential energy, the differential quadrature method, and iterative algorithms are used to obtain critical buckling temperatures and the post-buckling temperature-deflection curves. The results are presented for both symmetrically and unsymmetrically laminated plates with ceramic/metal functionally graded layers, showing the effects of temperature-dependent properties, layup scheme, material composition, initial imperfection, geometric parameters, and boundary conditions on buckling temperature and thermal post-buckling behavior. [DOI: 10.1115/1.1795220]

1 Introduction

Laminated composite plate structures, including those made of inhomogeneous functionally graded materials (FGMs) and characterized by a smooth and continuous change in both composition profile and material properties in certain spatial coordinates, are widely used in many industries. Quite often they are required to operate in varying temperature environments where, due to the in-plane displacement constraints on plate edges, significant strains and in-plane stresses are induced even when there is no mechanical load, therefore, initiating buckling and post-buckling phenomena involving geometric nonlinearity at high thermal load levels. In addition, significant variations in both the thermal and elastic properties of plate materials are to be expected with temperature fluctuations; for example, Young's modulus and the shear modulus usually decrease, while Poisson's ratio and the thermal expansion coefficient usually increase at elevated temperatures. Hence, knowledge of the thermal buckling and post-buckling characteristics of laminated plates that accounts for the effects of temperature-dependent material properties is of utmost importance for reliable and economic engineering design.

Many investigations of the buckling and post-buckling responses of composite laminated plates in thermal environments are available in the literature. Tauchert [1], Noor and Burton [2], and Argyris and Tenek [3] presented comprehensive reviews in state-of-the-art development in this field. However, studies of the thermal buckling and post-buckling behavior of laminated plates with temperature-dependent material properties are few. Chen and Chen [4,5] and Singha et al. [6] obtained finite element solutions for the thermal buckling and post-buckling of composite laminated thin plates. Significant influence of temperature-dependent properties on the critical buckling temperature and post-buckling load-deflection curves was reported in [4,5]. By using the natural mode method, Argyris and Tenek [7] assumed the material properties to be in cubic variation with temperature and examined the post-buckling response of composite laminates under compressive load and a temperature gradient. In the framework of the first-order shear deformation theory, Srikanth and Kumar [8] discussed the effect of temperature-dependent properties on the post-buckling behavior and failure modes of symmetric laminates through a direct application of the principle of minimum total potential energy. Shen [9] employed Reddy's higher-order shear deformation theory and an iterative perturbation technique to obtain asymptotic post-buckling solutions for both perfect and imperfect antisymmetric angle-ply and symmetric cross-ply laminated plates that were resting on an elastic foundation. All of the foregoing studies dealt with homogeneous materials and a thermal load due to a uniform temperature change, except [6,7], which considered nonuniform temperature distribution.

In the past few years, the buckling and post-buckling of FGM structures has drawn the attention of the scientific community. Shahsiah and Eslami gave the thermal buckling results for simply supported FGM cylindrical shells [10,11]. Shen [12–14] and Shen and Leung [15] examined the post-buckling behavior of pressure-loaded and axially loaded cylindrical shells and panels in thermal environments. Javaheri, Najafizadeh, and Eslami conducted a series of investigations into the compressive and thermal buckling of FGM plates with simply supported and clamped edges. They presented successively closed solutions for the buckling loads of compressed rectangular plates [16] and circular plates [17], and the critical buckling temperatures of rectangular plates [18,19] and circular plates [20] under different temperature distributions. Their work [16–20], however, did not include post-buckling analysis and neglected the existence conditions of bifurcation-type buckling in composite plates with the stretching-bending coupling effect, which were clearly brought out by Qatu and Leissa [21]. Hence, the results they provided for pure FGM plates that were not fully clamped do not exist in actual situations. To address the buckling and post-buckling problems of FGM plate structures on a sound physical basis, Yang and Shen [22] developed a semi-

---

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, April 9, 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 of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

analytical approach to determine the post-buckling paths for FGM rectangular thin plates that are subjected to transverse and in-plane loads. Thermal loads due to temperature change were not considered. Most recently, the present authors [23] discussed the existence of bifurcational instability and carried out a post-buckling analysis for shear deformable FGM rectangular plates with surface-mounted piezoelectric actuators under thermo-electro-mechanical loads. Buckling load parameters for clamped plates were calculated and load-deflection curves were traced by using Reddy's higher-order shear deformation theory. All of the above analyses concerning FGM plates, except [22], did not account for the variation in material properties with temperature change and the effect of unavoidable initial geometric imperfections. There is obviously a need to gain a comprehensive insight into the thermal post-buckling behavior of plate structures that are comprised of FGMs with actual temperature-dependent thermoelastic properties, possible geometric imperfections, and under general edge boundary conditions.

This paper investigates the post-buckling of moderately thick FGM laminated rectangular plates with temperature-varying properties and undergoing a uniform temperature change. The mathematical formulation is based on the first-order shear deformation theory and geometric nonlinearity due to moderately large deformation of the von Kármán type. The initial imperfection of the plate is taken into consideration but, for simplicity, is assumed to be of the same form as the buckling mode. A system of nonlinear ordinary differential equations is obtained by the principle of minimum total potential energy and is then solved by the differential quadrature method and iterative approaches to determine the critical buckling temperature and the nonlinear temperature-deflection curves of plates with different boundary conditions. Numerical results are presented for laminated plates with various symmetric and unsymmetric layup schemes.

2 Theoretical Formulations

2.1 Preliminaries. We consider an imperfect N_L-layer laminated rectangular plate $(a\times b\times h)$ that consists of an isotropic, homogeneous metal layer and inhomogeneous ceramic/metal FGM layers. The plate is defined in a Cartesian coordinate system (x,y,z), where z is the coordinate along the normal direction to the midplane of the plate and (x,y) are the coordinates of a generic point along the in-plane directions. In accordance with the first-order shear deformation theory, the displacement field (U,V,W) of the plate takes the following form:

$$\begin{Bmatrix} U(x,y,z) \\ V(x,y,z) \\ W(x,y,z) \end{Bmatrix} = \begin{Bmatrix} u(x,y) \\ v(x,y) \\ w(x,y) \end{Bmatrix} + z\begin{Bmatrix} \varphi_x(x,y) \\ \varphi_y(x,y) \\ 0 \end{Bmatrix} + \begin{Bmatrix} 0 \\ 0 \\ w^*(x,y) \end{Bmatrix} \tag{1}$$

where (u,v,w) denotes the midplane $(z=0)$ displacements of a point, (φ_x,φ_y) are the midplane rotations of normal about the y- and x-axes, respectively, and w^* is the initial geometric imperfection, which is assumed to be relatively small and have the same shape as w. Let η be the imperfection parameter

$$\eta = 1 + 2\frac{w^*}{w} \tag{2}$$

Based on von Kármán's assumptions, geometric nonlinear strains are obtained in the sense of moderately large deflections and small rotations by strain-displacement relations as

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_0 + z\boldsymbol{\varepsilon}_1 \tag{3}$$

where

$$\boldsymbol{\varepsilon}_0 = \mathbf{L}_0\mathbf{U}, \quad \boldsymbol{\varepsilon}_1 = \mathbf{L}_1\mathbf{U}, \quad \mathbf{U} = [u \;\; v \;\; w \;\; \varphi_x \;\; \varphi_y]^{\mathrm{T}}$$

$$\boldsymbol{\varepsilon} = [\varepsilon_x \;\; \varepsilon_y \;\; \gamma_{xy} \;\; \gamma_{yz} \;\; \gamma_{zx}]^{\mathrm{T}}$$

and the nonlinear differential operator $\mathbf{L}_0$ and linear differential operator $\mathbf{L}_1$ take the forms of

$$\mathbf{L}_0 = \begin{bmatrix} \dfrac{\partial}{\partial x} & 0 & \dfrac{1}{2}\eta\left(\dfrac{\partial}{\partial x}\right)^2 & 0 & 0 \\ 0 & \dfrac{\partial}{\partial y} & \dfrac{1}{2}\eta\left(\dfrac{\partial}{\partial y}\right)^2 & 0 & 0 \\ \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial x} & \eta\dfrac{\partial}{\partial x}\dfrac{\partial}{\partial y} & 0 & 0 \\ 0 & 0 & \dfrac{1+\eta}{2}\dfrac{\partial}{\partial y} & 0 & 1 \\ 0 & 0 & \dfrac{1+\eta}{2}\dfrac{\partial}{\partial x} & 1 & 0 \end{bmatrix},$$

$$\mathbf{L}_1 = \begin{bmatrix} 0 & 0 & 0 & \dfrac{\partial}{\partial x} & 0 \\ 0 & 0 & 0 & 0 & \dfrac{\partial}{\partial y} \\ 0 & 0 & 0 & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial x} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

Neglecting the transverse normal stress σ_z, the Duhamel-Neumann form of the linear elastic stress-strain relation is written as

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{Bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} & 0 & 0 \\ Q_{12} & Q_{22} & Q_{23} & 0 & 0 \\ Q_{13} & Q_{23} & Q_{33} & 0 & 0 \\ 0 & 0 & 0 & k_4^2 Q_{44} & 0 \\ 0 & 0 & 0 & 0 & k_5^2 Q_{55} \end{bmatrix} \left(\begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{zx} \end{Bmatrix} - \begin{Bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ 0 \\ 0 \end{Bmatrix} \Delta T \right) \tag{5}$$

or in a short form

$$\boldsymbol{\sigma} = \mathbf{Q}(\boldsymbol{\varepsilon} - \boldsymbol{\alpha}\Delta T) \tag{6}$$

where σ_x, σ_y, τ_{xy} are in-plane normal and shear stresses, τ_{yz}, τ_{zx} are the transverse shear stresses, k_4^2 and k_5^2 are shear correction factors and taken as $k_4^2 = k_5^2 = 5/6$, Q_{ij} are the stiffness coefficients, and α_i are the linear thermal expansion coefficients. Note that, for the materials considered herein, $Q_{13} = Q_{23} = 0$, $\alpha_1 = \alpha_2 = \alpha$, $\alpha_3 = 0$. As both the elastic and thermal properties of the materials are dependent on the environment temperature, we have

$$\mathbf{Q} = \mathbf{Q}(T), \quad \boldsymbol{\alpha} = \boldsymbol{\alpha}(T) \tag{7}$$

where the temperature $T = T_0 + \Delta T$, T_0 is the reference temperature at which the plate is stress free, and ΔT is the uniform temperature increment.

The stress resultants are

$$\begin{bmatrix} N_x & M_x \\ N_y & M_y \\ N_{xy} & M_{xy} \end{bmatrix} = \sum_{k=1}^{N_L} \int_{z_k}^{z_{k+1}} \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix}^{(k)} (1,z)\,dz,$$

$$\begin{Bmatrix} Q_y \\ Q_x \end{Bmatrix} = \sum_{k=1}^{N_L} \int_{z_k}^{z_{k+1}} \begin{Bmatrix} \tau_{yz} \\ \tau_{zx} \end{Bmatrix}^{(k)} dz \tag{8}$$

Thermal stress resultants N_x^*, N_y^*, N_{xy}^* and moment resultants M_x^*, M_y^*, M_{xy}^* are

$$\begin{bmatrix} N_x^* & M_x^* \\ N_y^* & M_y^* \\ N_{xy}^* & M_{xy}^* \end{bmatrix} = \sum_{k=1}^{N_L} \int_{z_k}^{z_{k+1}} \begin{Bmatrix} (Q_{11}+Q_{12})\alpha \\ (Q_{12}+Q_{22})\alpha \\ 0 \end{Bmatrix}^{(k)} \Delta T(1,z)dz \tag{9}$$

The general boundary conditions at plate edges, i.e., simply supported (S), clamped (C), free (F), or a combination of these three, are considered and take the form of

$$\text{S:} \quad u_n = N_{ns} = w = M_n = \varphi_s = 0 \tag{10a}$$

$$\text{C:} \quad u_n = u_s = w = \varphi_n = \varphi_s = 0 \tag{10b}$$

$$\text{F:} \quad N_n = N_{ns} = Q_n = M_n = M_{ns} = 0 \tag{10c}$$

where the subscripts n and s refer to the normal and tangential directions of the plate edge.

2.2 Variational Formulation. In the absence of body force and applied mechanical load, the total potential energy of the laminated plate due to thermal deformation is

$$\Pi = \tfrac{1}{2} \int_V (\boldsymbol{\varepsilon} - \boldsymbol{\alpha} \Delta T)^{\mathrm{T}} \boldsymbol{\sigma} dV \tag{11}$$

We assume that the displacement vector $\mathbf{U}$ can be represented in a series form as

$$\begin{aligned} \mathbf{U} &= \sum_{m=1}^{M} \tilde{\mathbf{U}}_m(y) \hat{\mathbf{U}}_m(x) \\ &= \sum_{m=1}^{M} \begin{bmatrix} \tilde{u}_m(y) & 0 & 0 & 0 & 0 \\ 0 & \tilde{v}_m(y) & 0 & 0 & 0 \\ 0 & 0 & \tilde{w}_m(y) & 0 & 0 \\ 0 & 0 & 0 & \tilde{\varphi}_{xm}(y) & 0 \\ 0 & 0 & 0 & 0 & \tilde{\varphi}_{ym}(y) \end{bmatrix} \\ &\quad \times \begin{Bmatrix} \hat{u}_m(x) \\ \hat{v}_m(x) \\ \hat{w}_m(x) \\ \hat{\varphi}_{xm}(x) \\ \hat{\varphi}_{ym}(x) \end{Bmatrix} \end{aligned} \tag{12}$$

where the components in $\hat{\mathbf{U}}_m(x)$ are unknown functions to be determined, and those in $\tilde{\mathbf{U}}_m(y)$ are chosen to be the analytical functions which satisfy at least the geometric boundary conditions at edges $y=0$, b, as given in Appendix A.

Substituting Eqs. (3), (6), and (12) into Eq. (11), integrating over $z=-h/2$ to $z=h/2$ and $x=0$ to $x=a$, introducing dimensionless parameters

$$\xi = x/a, \quad \varsigma = y/b, \quad \beta = a/b, \quad \gamma = h/a,$$

$$(\hat{u}_m, \hat{v}_m, \hat{w}_m) = (\hat{\hat{u}}_m, \hat{\hat{v}}_m, \hat{\hat{w}}_m)/h \tag{13}$$

and then applying the principle of minimum total potential energy results in an ordinary differential equation system in terms of $\hat{\mathbf{U}}_m$

$$\mathbf{L}_m \hat{\mathbf{U}}_m = \mathbf{0} \tag{14}$$

or

$$\begin{bmatrix} L_{11} & L_{12} & \gamma\eta L_{13} & L_{14} & L_{15} \\ L_{21} & L_{22} & \gamma\eta L_{23} & L_{24} & L_{25} \\ L_{31} & L_{32} & \gamma(L_{331}+\eta L_{332}) & L_{34} & L_{35} \\ L_{41} & L_{42} & \gamma(L_{431}+\eta L_{432}) & L_{44} & L_{45} \\ L_{51} & L_{52} & \gamma(L_{531}+\eta L_{532}) & L_{54} & L_{55} \end{bmatrix} \begin{Bmatrix} \hat{u}_m \\ \hat{v}_m \\ \hat{w}_m \\ \hat{\varphi}_{xm} \\ \hat{\varphi}_{ym} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix} \tag{15}$$

in which the details of the ordinary differential operators are defined in Appendix B.

3 Solution Procedures

Equation (15) is a nonlinear ordinary differential system that is also a nonlinear function of temperature. To solve the problem, the differential quadrature method (DQM) is used to convert the equations and the associated boundary conditions at $\xi=0$, 1 into a set of nonlinear algebraic equations, the solutions of which are then sought by an iterative algorithm. According to DQM, the unknown function vector $\hat{\mathbf{U}}_m$ and its kth partial derivatives with respect to ξ at a discrete point are approximated as the linear weighted sums of its values at all of discrete points along the ξ-axis

$$\hat{\mathbf{U}}_m = \sum_{j=1}^{N} P_j(\xi) \hat{\mathbf{U}}_{mj}, \quad \left. \frac{\partial^k \hat{\mathbf{U}}_m}{\partial \xi^k} \right|_{\xi=\xi_i} = \sum_{j=1}^{N} C_{ij}^{(k)} \hat{\mathbf{U}}_{mj} \tag{16}$$

where $\hat{\mathbf{U}}_{mj} = \hat{\mathbf{U}}_m|_{\xi=\xi_j}$ is the value of $\hat{\mathbf{U}}_m$ at the jth point and N is the number of discrete points whose ξ-coordinates are given by either of the two spacing patterns generally used in DQM:

1. Equal spacing pattern

$$\xi_j = \frac{j-1}{N-1} \quad (j=1,2,\ldots,N) \tag{17}$$

2. Cosine spacing pattern

$$\xi_j = \frac{1}{2} - \frac{1}{2}\cos\left[\frac{\pi(j-1)}{N-1}\right] \quad (j=1,2,\ldots,N) \tag{18}$$

The cosine spacing pattern is employed in this paper because it provides better convergence for the present problem. The interpolation functions P_j are Lagrange polynomials, and the weighting coefficients $C_{ij}^{(k)}$ are determined through recursive formulas given in [24,25].

By applying approximation relation (16) to the ordinary differential equation (14) at each of the interior grid points, we have

$$\overline{\mathbf{L}}_{mi} \boldsymbol{\Delta}_m = \mathbf{0} \quad (i=2,\ldots,N-1) \tag{19}$$

where $\overline{\mathbf{L}}_{mi}$ is the discretized form of differential operator matrix $\mathbf{L}_m$ at ξ_i, and $\boldsymbol{\Delta}_m$ is the unknown vector composed of displacement components at all grid points, which takes the form

$$\boldsymbol{\Delta}_m = [\hat{\mathbf{U}}_{m1}^{\mathrm{T}}, \ldots, \hat{\mathbf{U}}_{mi}^{\mathrm{T}}, \ldots, \hat{\mathbf{U}}_{mN}^{\mathrm{T}}]^{\mathrm{T}} \tag{20}$$

Treating the boundary conditions (9) at $\xi_1=0$, $\xi_N=1$ by relation (16) in a similar way and then combining the resulted equations with Eq. (19) yields $5\times N$ nonlinear algebraic equations for each m,

$$\mathbf{G}_m \boldsymbol{\Delta}_m = \mathbf{R}_m \tag{21}$$

where $\mathbf{G}_m$ is the nonlinear matrix that is dependent on both temperature and unknown displacement vector $\boldsymbol{\Delta}_m$.

It should be noted that the right-hand-side term $\mathbf{R}_m$ comes from thermally induced stress resultants and bending moments in simply supported or free boundary conditions at edges $\xi=0$, 1, and will automatically vanish when the plate is clamped where only displacement boundary conditions are involved or when the plate is symmetrically laminated where the stretching-bending stiffness elements B_{ij} are absent. It is obvious that the bifurcation-type

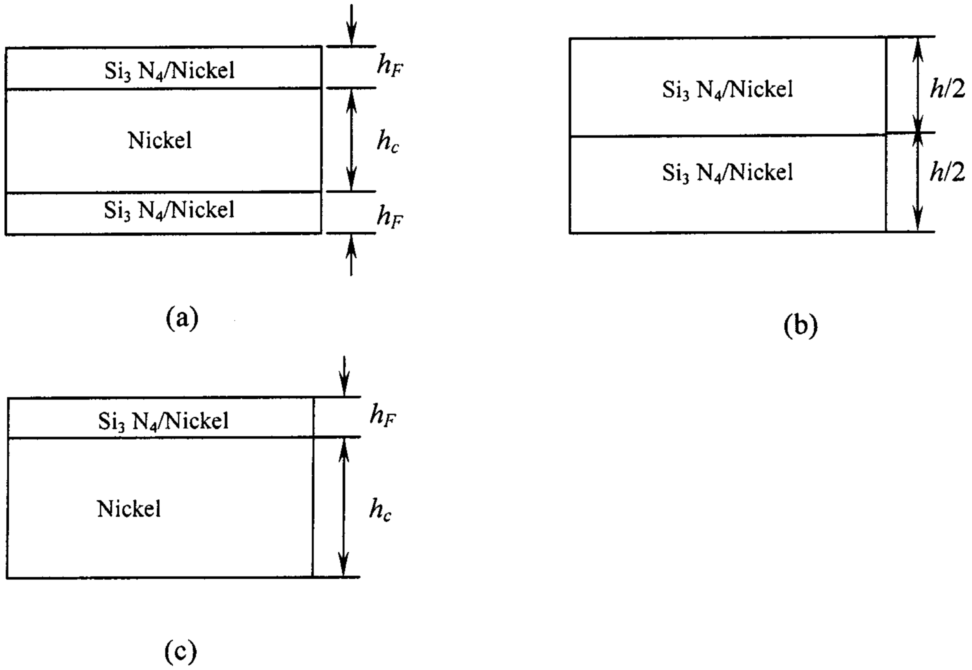


Fig. 1 FGM laminated plates with different layup schemes: (*a*) FGM/Nickel/FGM, (*b*) FGM/FGM, and (*c*) FGM/Nickel

thermal buckling will occur only when $\mathbf{R}_m=\mathbf{0}$, otherwise transverse deflection will take place when $\mathbf{R}_m\neq\mathbf{0}$, irrespective of the magnitude of the temperature change.

The thermal buckling temperature, when it exists, is determined from the nonlinear homogeneous equation

$$\mathbf{G}_m\boldsymbol{\Delta}_m=\mathbf{0} \tag{22}$$

by an iterative numerical procedure with the following steps:

1.1. Assume that $\Delta T=0$, the buckling temperature ΔT_{cr} is solved by using the thermoelastic properties at reference temperature T_0.
1.2. Calculate the thermoelastic properties at $T=T_0+\Delta T_{cr}$ and update $\mathbf{G}_m$, and a new buckling temperature is obtained.
1.3. Repeat step 1.2 until the thermal buckling temperature converges.

The nonlinear temperature-deflection curve, also known as the post-buckling equilibrium path, is traced by two different iterative schemes, depending on the presence of $\mathbf{R}_m$. When $\mathbf{R}_m=\mathbf{0}$, the following iteration process is applicable

2.1. Begin with the dimensionless central deflection $w_c/h=0$.
2.2. Use the iterative procedures 1.1–1.3.
2.3. Specify a new value of w_c/h.
2.4. Calculate the thermo-elastic properties at $T=T_0+\Delta T_{cr}$, and scale up the buckling mode that is obtained in step 2.2 to form a new $\mathbf{G}_m$ to determine the post-buckling temperature.
2.5. Repeat step 2.4 until the post-buckling temperature converges.
2.6. Repeat steps 2.3–2.5 to obtain the post-buckling equilibrium path.

The modified Newton-Raphson technique is used if $\mathbf{R}_m\neq\mathbf{0}$, but the process is omitted here for brevity.

The convergence criterion adopted in this study is of the form

$$\left|\frac{\Delta T_{cr}^{(i+1)}-\Delta T_{cr}^{(i)}}{\Delta T_{cr}^{(i)}}\right|\leqslant\delta \tag{23}$$

That is, the iteration process is to be continued until the relative difference between the solutions obtained from two consecutive iterations (i) and $(i+1)$ reduces to a prescribed error tolerance δ.

4 Results and Discussion

In this section, thermal buckling and post-buckling analyses are carried out for laminated plates that are comprised of FGM layers, which are a mixture of nickel and silicon nitride (Si_3N_4). Four different layup schemes are considered, including a) a symmetric three-layer laminate that consists of a pure nickel middle layer and two FGM thin layers (FGM/Nickel/FGM); b) a laminate with two FGM layers that are symmetrically bonded together (FGM/FGM); c) an unsymmetrically laminated plate with an FGM thin layer perfectly attached to the top of a thick, isotropic nickel layer (FGM/Nickel); and d) a pure FGM plate. Schematic configurations of the first three types of laminated plates are given in Fig. 1. The volume fraction $V_{ceramic}$ for Si_3N_4 in the FGM layer follows a simple power law:

$$\text{FGM/Nickel/FGM:}\quad V_{ceramic}(z)=\left(\frac{2z-h_c}{2h_{\mathrm{F}}}\right)^n\quad\left(\frac{h_c}{2}\leqslant z\leqslant\frac{h}{2}\right) \tag{24a}$$

$$V_{ceramic}(z)=\left(-\frac{2z+h_c}{2h_{\mathrm{F}}}\right)^n\quad\left(-\frac{h}{2}\leqslant z\leqslant-\frac{h_c}{2}\right) \tag{24b}$$

$$\text{FGM/FGM:}\quad V_{ceramic}(z)=\left(\frac{2z}{h}\right)^n\quad\left(0\leqslant z\leqslant\frac{h}{2}\right) \tag{25a}$$

$$V_{ceramic}(z)=\left(-\frac{2z}{h}\right)^n\quad\left(-\frac{h}{2}\leqslant z\leqslant 0\right) \tag{25b}$$

Table 1 Temperature-dependent coefficients for silicon nitride and nickel

Properties	Material	P_{-1}	P_0	P_1	P_2	P_3
E (GPa)	Silicon nitride	0	348.43e9	-3.070e-4	2.160e-7	-8.946e-11
	Nickel	0	223.95e9	-2.794e-4	3.998e-9	0
ν	Silicon nitride	0	0.2400	0	0	0
	Nickel	0	0.3100	0	0	0
α (1/K)	Silicon nitride	0	5.8723e-6	9.095e-4	0	0
	Nickel	0	9.9209e-6	8.705e-4	0	0

$$\text{FGM/Nickel:}\quad V_{ceramic}(z)=\left(\frac{2z-h+2h_F}{2h_F}\right)^n \quad \left(\frac{h}{2}-h_F\leqslant z\leqslant\frac{h}{2}\right) \tag{26}$$

$$\text{Pure FGM:}\quad V_{ceramic}(z)=\left(\frac{2z+h}{2h}\right)^n \quad \left(-\frac{h}{2}\leqslant z\leqslant\frac{h}{2}\right) \tag{27}$$

where n is the volume fraction index ($0\leqslant n\leqslant\infty$). The effective properties P_{eff} at an arbitrary point within the FGM layer can be estimated by the rule of mixture as

$$P_{\text{eff}}=P_{metal}+(P_{ceramic}-P_{metal})V_{ceramic} \tag{28}$$

in which $P_{ceramic}$ and P_{metal} stand for the property parameters of ceramic and metal, respectively. These properties, as discussed earlier, are also the functions of environment temperature T (K), as described by [26]

$$P=P_0(P_{-1}T^{-1}+1+P_1T+P_2T^2+P_3T^3) \tag{29}$$

The coefficients P_0, P_{-1}, P_1, P_2, and P_3 for Young's modulus E, the Poisson ratio ν, and the thermal expansion coefficient α of nickel and Si_3N_4 are given in Table 1.

To evaluate the effect of temperature-dependence of constituent materials, the results are compared with those obtained by using the following temperature-independent material constants [27]

$$\text{Nickel:}\quad E=204\ \text{GPa},\quad \nu=0.31,\quad \alpha=13.2\times10^{-6}\ \text{1/K}$$

$$Si_3N_4:\quad E=310\ \text{GPa},\quad \nu=0.24,\quad \alpha=3.4\times10^{-6}\ \text{1/K}$$

4.1 Validation and Convergence Studies. Due to the lack of appropriate results of functionally graded plates for direct comparison, validation of the present formulation is conducted in two ways. We first study the accuracy and convergence of the proposed method in linear buckling problems where no iteration process is involved. For this purpose, the buckling temperatures $\lambda_{cr}=\alpha\Delta T_{cr}\times10^3$ of simply supported isotropic square plates ($\nu=0.21$) are compared in Table 2 with the analytical solutions of Noor and Burton [28] using the three-dimensional thermoelasticity theory and Shen [29] based on the first-order shear deformation plate theory. Material properties are assumed to be temperature independent. Our results are given in the form of convergence studies with respect to N, the number of discrete points distributed along the ξ-axis, and to M, the number of truncated series expansions in Eq. (12). The present method converges to solutions that are in excellent agreement with both analytical solutions when $N\geqslant15$, $M\geqslant5$. In the following computation, $N\geqslant15$, $M\geqslant5$ will be used, otherwise, they will be stated.

To validate the present formulation in thermal post-buckling of plates with temperature-dependent material properties, the post-buckling of a $(0°/90°)_s$ symmetric cross-ply square plate ($a/h=30$) under uniform temperature rise is considered, which was also analyzed by Shen [9] using the perturbation asymptotic method and Reddy's higher-order shear deformation theory. The plate is simply supported with all edges restrained against in-plane normal displacements, but free to move in the tangential directions. Except the Poisson' ratio, which is considered to be independent of temperature, the thermoelastic constants of each ply are assumed to be linear functions of temperature as [9]

$$E_{11}=E_{110}(1+E_{111}T),\quad E_{22}=E_{220}(1+E_{221}T)$$
$$G_{12}=G_{120}(1+G_{121}T),\quad G_{13}=G_{130}(1+G_{131}T),$$
$$G_{23}=G_{230}(1+G_{231}T)$$
$$\alpha_{11}=\alpha_{110}(1+\alpha_{111}T),\quad \alpha_{22}=\alpha_{220}(1+\alpha_{221}T) \tag{30}$$

where $E_{110}/E_{220}=40$, $G_{120}/E_{220}=G_{130}/E_{220}=0.5$, $G_{230}/E_{220}=0.2$, $\alpha_{110}/\alpha_0=1$, $\alpha_{220}/\alpha_0=10$, $\nu_{12}=0.25$, $E_{111}=-5\times10^{-4}$, $E_{221}=G_{121}=G_{131}=G_{231}=-2\times10^{-4}$, $\alpha_{111}=\alpha_{221}=5\times10^{-3}$, $\alpha_0=10^{-6}/°C$. The post-buckling temperature-deflection curves of the plate with and without temperature-dependent material properties are compared in Fig. 2 with Shen's results. In temperature-independent cases, the constants E_{111}, E_{221}, G_{121}, G_{131}, G_{231}, α_{111}, and α_{221} are set equal to zero. It is evident that good agreement is achieved in this comparison study.

Convergence study of the iterative process is also conducted. Error tolerance $\delta=10^{-3}$ is used throughout the paper. The convergence of the iteration scheme is demonstrated in Table 3, where the total numbers of iteration (TNI) required to obtain convergent results for thermal post-buckling curves of plates ($a/h=30$, $n=2.0$) with temperature-dependent material properties are recorded for FGM/Nickel/FGM, FGM/FGM, and FGM/Nickel square plates with SSSS, SCSC, and CCCC boundary conditions.

Table 2 Comparison of critical buckling temperature $\lambda_{cr}=\alpha\Delta T_{cr}\times10^3$ for SSSS isotropic square plates

	Present				Noor & Burton	Shen
b/h	$N=9, M=3$	$N=11, M=5$	$N=15, M=5$	$N=19, M=7$	[28]	[29]
5	40.037	40.433	41.232	41.237	39.90	41.297
10	11.472	11.772	11.962	11.962	11.83	11.978
20	3.0217	3.0516	3.1156	3.1156	3.109	3.1194
100	0.1135	0.1235	0.1265	0.1265	0.1264	0.1265

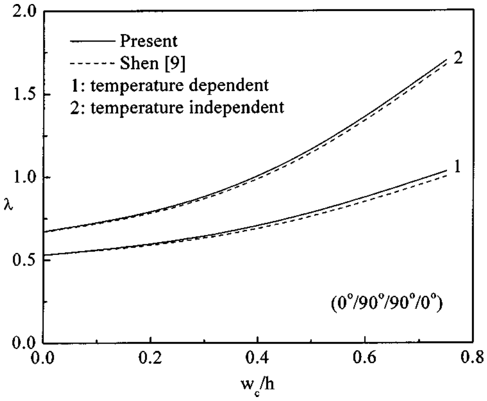


Fig. 2 Comparisons of post-buckling paths for a simply supported, symmetric cross-ply square plate

Note that for FGM/Nickel/FGM, FGM/FGM, and CCCC FGM/Nickel plates, the numbers at $w_c/h=0$ are in fact the TNIs to achieve the convergent critical buckling temperatures. The convergence rate varies mainly with boundary conditions and post-buckling deformations. As it can be expected, the process takes a longer time at greater deflections w_c/h, i.e., as the nonlinearity increases. Rapid convergence is observed for both SSSS and SCSC plates while the CCCC plate needs more iterative computation when $w_c/h \geqslant 0.6-0.8$.

4.2 Parametric Studies. We now present critical buckling temperature parameters $\lambda_{cr}=\alpha_0 \Delta T_{cr} \times 10^3$ for various FGM laminated rectangular plates under uniform temperature change, where α_0 is the thermal expansion coefficient of nickel at reference temperature $T_0=300\,\mathrm{K}$. All of the four plate edges are immovable in the midplane in the SSSS, SCSC, and CCCC plates, but are free to move along the x-axis in the SFSF and CFCF plates. Out-of-plane boundary conditions are denoted by symbolic notations; for example, SFSF refers to a plate free at $x=0$, a and simply supported at $y=0$, b. Thickness ratio h_c/h_F between the nickel layer and the FGM layer is 8, 9, and 0 for the FGM/Nickel/FGM, FGM/Nickel, and FGM/FGM laminated plates.

Table 4 gives λ_{cr} for SSSS, SCSC, CCCC, SFSF, and CFCF symmetrically laminated FGM square plates with side-to-thickness ratio $a/h=30$ and different volume fraction indices ($n=0.5,2.0,10$). Table 5 shows the results for CCCC FGM rectangular plates with $a/h=30$, 50, $a/b=1.0$, 1.5, and various material compositions, among which Si_3N_4 and Nickel represent isotropic silicon nitride plate and pure nickel plate, respectively. The value of a/b varies with a remaining constant.

It can be seen that a plate with higher a/b and lower values of n and a/h has the greater thermal buckling temperature, and a plate with symmetric FGM/Nickel/FGM lamination has the lowest thermal loading capacity among all of the layup schemes under consideration.

Note that the critical buckling temperature is greatly overestimated when the temperature-dependence of material properties is not taken into account. The discrepancy between temperature-

Table 3 Total numbers of iteration (TNI) in thermal post-buckling analysis for laminated FGM square plates

w_c/h	FGM / Nickel / FGM			FGM / FGM			FGM / Nickel		
	SSSS	SCSC	CCCC	SSSS	SCSC	CCCC	SSSS	SCSC	CCCC
0.0	4	5	5	4	5	5	4	5	5
0.2	5	5	5	5	5	5	5	5	5
0.4	5	5	5	5	5	5	5	5	5
0.6	5	5	6	5	7	7	5	5	5
0.8	5	6	9	5	7	8	5	5	7
1.0	5	6	9	6	7	9	5	5	8

Table 4 Thermal buckling temperature $\lambda_{cr}=\alpha_0 \Delta T_{cr} \times 10^3$ for symmetric laminated square plates

Boundary conditions	Temperature dependence	FGM / Nickel / FGM			FGM / FGM		
		$n=0.5$	$n=2.0$	$n=10$	$n=0.5$	$n=2.0$	$n=10$
SSSS	dependent	1.5218	1.4134	1.3347	1.9266	1.6984	1.4852
	independent	1.6395	1.4748	1.3669	2.9334	2.0302	1.5783
SCSC	dependent	2.7289	2.5408	2.4035	3.4103	3.0293	2.6658
	independent	3.1096	2.7986	2.5947	5.5658	3.8492	2.9940
CCCC	dependent	3.6081	3.3627	3.1830	4.4706	3.9919	3.5264
	independent	4.4271	3.8443	3.5647	7.6435	5.2857	4.1127
CFCF	dependent	3.6721	3.4336	3.2574	4.4178	4.0244	3.5859
	independent	4.4362	3.9363	3.6598	7.5336	5.3318	4.1903
SFSF	dependent	0.9849	0.9140	0.8631	1.2360	1.1002	0.9554
	independent	1.0325	0.9294	0.8647	1.8173	1.2760	0.9880

Table 5 Thermal buckling temperature $\lambda_{cr}=\alpha_0\Delta T_{cr}\times 10^3$ for clamped FGM rectangular plates

Side-to-thickness ratio	Temperature dependence	Material composition				
		Si_3N_4	$n=0.5$	$n=2.0$	$n=10.0$	Nickel
				$a/b=1.0$		
$a/h=30$	dependent	4.9867	4.0583	3.6595	3.4825	3.0894
	independent	14.087	6.9458	4.8657	4.0680	3.4271
$a/h=50$	dependent	2.0828	1.6609	1.4798	1.3976	1.2328
	independent	5.1235	2.5260	1.7712	1.4822	1.2480
				$a/b=1.5$		
$a/h=30$	dependent	7.6920	5.7121	5.3330	5.0544	4.9083
	independent	24.342	7.3584	6.6275	6.1479	5.9129
$a/h=50$	dependent	3.3981	2.4112	2.2433	2.1208	2.0572
	independent	8.9186	2.7082	2.4366	2.2584	2.1713

dependent solutions and temperature-independent solutions increases dramatically as side-to-thickness ratio a/h and volume fraction index n decrease. In terms of out-of-plane boundary conditions and lamination scheme, the discrepancy is maximum for a CCCC FGM plate and minimum for an SFSF plate with FGM/Nickel/FGM symmetric configuration. The difference reaches as high as approximately 316% for a clamped, isotropic Si_3N_4 rectangular plate with $a/h=30$ and $a/b=1.5$, but is less than 0.2% for an SFSF FGM/Nickel/FGM square plate with volume fraction index $n=10$ and $a/h=30$.

It should also be mentioned that the presence of stretching-bending coupling effect in pure FGM plates and FGM laminated plates with unsymmetric layup, irrespective of the magnitude of the temperature, gives rise to lateral deflection, and hence, no bifurcation-type thermal buckling exists in actual situations for FGM/Nickel and FGM plates if they are not fully clamped at all edges.

Typical thermal post-buckling results for FGM laminated plates under uniform temperature change are plotted in Figs. 3–8 in terms of the dimensionless temperature load $\lambda=\alpha_0\Delta T\times 10^3$ versus dimensionless central deflection w_c/h. Solid curves and dashed curves represent post-buckling equilibrium paths for temperature-dependent and temperature-independent plates, respectively.

Figure 3 gives the post-buckling equilibrium paths for simply supported, symmetrically laminated FGM/Nickel/FGM square plates with $n=0.5$, 2.0, and 10.0. The thermal post-buckling curve becomes lower as the value of n increases and when the thermal dependent properties are considered. The difference between solutions by using thermal-dependent properties and thermal-independent properties increases steadily with a decrease in the volume fraction index.

Thermal post-buckling behavior for plates with various lamination schemes is compared in Fig. 4, where curves 1, 2, 3, and 4 are the post-buckling paths for FGM/Nickel/FGM, FGM/FGM, FGM/Nickel and pure FGM plates, respectively. The plate edges are simply supported and the volume fraction index in FGM layer is $n=2.0$. In such a case, the post-buckling behavior of unsymmetrically laminated plates is different from those of symmetric plates in the early post-buckling region. Curves 3 and 4 originate from coordinate origin because no bifurcation buckling point exists for the FGM/Nickel and FGM plates. The results show that the FGM/FGM plate has the highest post-buckling thermal-loading capacity, while the FGM/Nickel plate possesses the weakest heat resistance. This is due to the fact that the FGM/FGM plate, among the four types of plates considered, contains the greatest amount of silicon nitride with a higher Young's modulus. Note that the temperature-independent results are significantly overpredicted, especially for FGM/FGM and FGM plates.

Figure 5 presents a comparison between the equilibrium paths

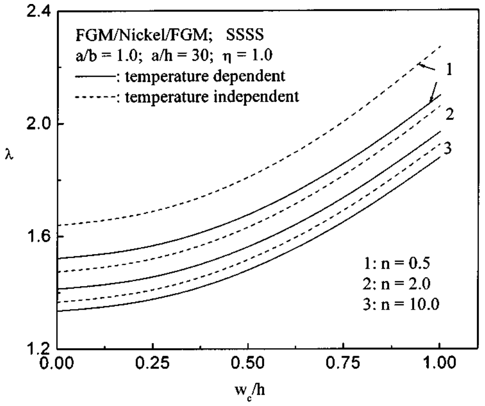


Fig. 3 Thermal post-buckling paths for simply supported FGM/Nickel/FGM square plates with different volume fraction indices

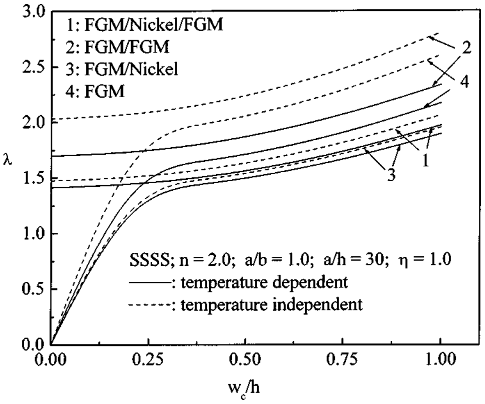


Fig. 4 Thermal post-buckling paths for SSSS square plates with different lamination schemes

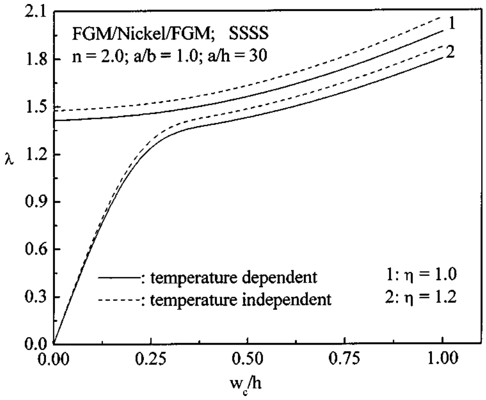


Fig. 5 Effect of geometric imperfection on the thermal post-buckling paths of SSSS FGM/Nickel/FGM square plates

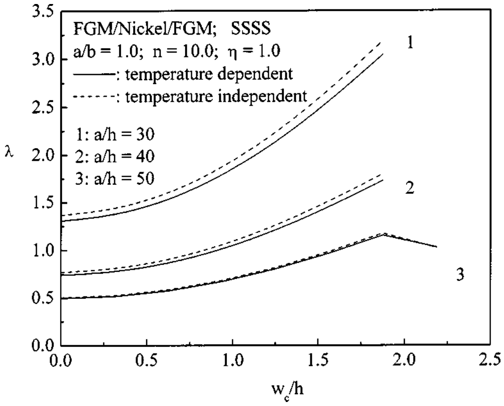


Fig. 7 Effect of side-to-thickness ratio on the thermal post-buckling paths of SSSS FGM/Nickel/FGM square plates

of SSSS FGM/Nickel/FGM square plates with and without geometric imperfection. $\eta=1.0$ and $\eta=1.2$ correspond to perfect and imperfect plates, respectively. The effect of initial imperfection is seen to weaken the thermal post-buckling strength of the plate. Quite similar to the case that is shown in Fig. 4, the curves for imperfect plates differ greatly from those for perfect plates at small deflections ($w_c/h<0.3$), but all curves tend to be quite similar at large deflections ($w_c/h>0.3$).

Figure 6 depicts the thermal post-buckling paths for symmetric FGM/Nickel/FGM laminated square plates with SCSC, CFCF, and SFSF boundary conditions. Note that both CFCF and SFSF plates under temperature change are free to expand or contract in the *x*-axis and only the in-plane displacement in the *y*-axis is constrained. Numerical illustration shows that quite different from the results for the SSSS and SCSC plates, only a small gain in the post-buckling temperature load of the SFSF and CFCF plates is observed as the deflection increases. Because of its weak edge support and movability of in-plane displacement in the *x*-axis, the thermal effect is relatively small in the SFSF plate, and therefore, the difference between temperature-dependent and temperature-independent solutions of the buckling temperature and thermal post-buckling path of the SFSF plate is insignificant.

Figure 7 deals with the effect of the side-to-thickness ratio on the thermal post-buckling response. To this end, thermal post-buckling paths are displayed for SSSS FGM/Nickel/FGM square plates with $n=10.0$ and $a/h=30$, 40, and 50. As expected, the dimensionless post-buckling temperature load becomes much larger as the plate becomes thicker. The discrepancy between temperature-dependent solutions and temperature-independent solutions decreases with an increase in a/h and even seems to be negligible when $a/h=50$. It is noteworthy that the post-buckling curves cannot be extended beyond $w_c/h\approx 1.875$ due to the convergence problem, and a sudden drop is observed in the post-buckling paths for the plate with $a/h=50$ and at $w_c/h>1.875$, thus indicating the possibility of a secondary instability phenomenon for thinner FGM laminated plates at larger post-buckling deflections.

Figure 8 shows the effect of plate aspect ratio, where curves 1, 2, and 3 are the results for simply supported, symmetrically laminated FGM/Nickel/FGM plates with $a/b=0.75$, 1.0, and 1.5 (a is kept constant). A lower value of a/b reduces the critical buckling temperature and, in turn, lowers the post-buckling temperature-load-carrying capacity. The influence of the temperature dependence of the material properties is much more pronounced for plates with higher values of a/b.

5 Conclusions

Thermal buckling and post-buckling analyses of symmetrically and unsymmetrically laminated rectangular plates that are comprised of functionally graded materials with temperature-dependent material properties and subjected to uniform temperature change have been conducted by using the first-order shear deformation plate theory. Numerical illustration has demonstrated that the critical bucking temperature and thermal post-buckling equilibrium path are remarkably overestimated, especially for thicker plates, when temperature-dependence of material properties is not taken into account, thus leading to possible unsafe engineering design. Unsymmetric FGM laminated plates do not have bifurcation-type thermal buckling due to the presence of

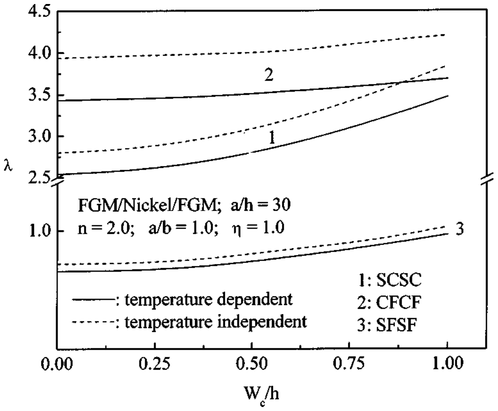


Fig. 6 Thermal post-buckling paths for SCSC, SFSF, and CFCF FGM/Nickel/FGM square plates

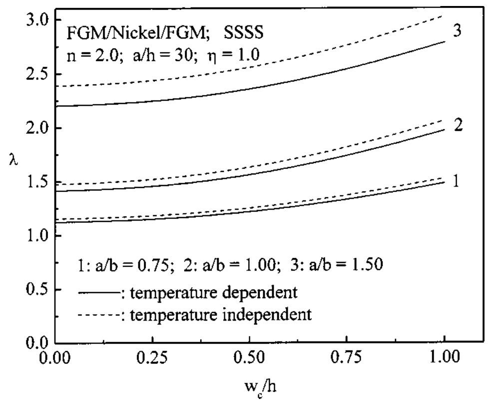


Fig. 8 Effect of plate aspect ratio on the thermal post-buckling paths of SSSS FGM/Nickel/FGM rectangular plates

stretching-bending coupling, and they exhibit post-buckling characteristics that are different from those of their symmetric counterparts. Certain factors, such as the layup scheme, volume fraction index in FGM layer, the out-of-boundary condition, initial geometric imperfection, side-to-thickness ratio, and plate aspect ratio, also significantly influence the buckling temperature and post-buckling behavior of the FGM laminated plates.

Acknowledgments

This work was supported by research grants from the Australian Research Council (A00104534) and the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 1024/01E). The authors are grateful for this financial support.

Appendix A

The functions $\tilde{u}_m$, $\tilde{v}_m$, $\tilde{w}_m$, $\tilde{\varphi}_{xm}$, and $\tilde{\varphi}_{ym}$, for different out-of-boundary conditions at $\varsigma=0$, 1, take the following forms:

1. simply supported at $\varsigma=0$, 1

$$\tilde{v}_m(\varsigma)=\tilde{w}_m(\varsigma)=\tilde{\varphi}_{xm}(\varsigma)=\sin(m\pi\varsigma);$$

$$\tilde{u}_m(\varsigma)=\tilde{\varphi}_{ym}(\varsigma)=\cos(m\pi\varsigma) \tag{A1}$$

2. clamped at $\varsigma=0$, 1

$$\tilde{v}_m(\varsigma)=\tilde{w}_m(\varsigma)=\tilde{\varphi}_{xm}(\varsigma)=Y_m; \quad \tilde{u}_m(\varsigma)=\tilde{\varphi}_{ym}(\varsigma)=\frac{\partial Y_m}{\partial\varsigma} \tag{A2}$$

where

$$Y_m=\sin\mu_m\varsigma-\sinh\mu_m\varsigma-\theta_m(\cos\mu_m\varsigma-\cosh\mu_m\varsigma)$$

$$\theta_m=(\sin\mu_m-\sinh\mu_m)/(\cos\mu_m-\cosh\mu_m)$$

$$\mu_m=(m+0.5)\pi$$

Appendix B

Details of the ordinary differential operators are given below:

$$L_{11}(\ )=J_0^{(uu)}\frac{d^2(\ )}{d\xi^2}+\frac{\beta^2A_{66}J_2^{(uu)}}{A_{11}}(\ )$$

$$L_{12}(\ )=\frac{\beta(A_{12}+A_{66})J_1^{(uv)}}{A_{11}}\frac{d(\ )}{d\xi}$$

$$L_{13}(\ )=\left(J_{00}^{(uww)}\frac{d^2(\ )}{d\xi^2}+\frac{\beta^2A_{66}J_{02}^{(uww)}}{A_{11}}(\ )\right)\frac{d(\ )}{d\xi}+\frac{\beta^2(A_{12}+A_{66})J_{11}^{(uww)}}{A_{11}}\frac{d(\ )}{d\xi}(\ )$$

$$L_{14}(\ )=\frac{B_{11}J_0^{(ux)}}{A_{11}h}\frac{d^2(\ )}{d\xi^2}+\frac{B_{66}\beta^2J_2^{(ux)}}{A_{11}h}(\ )$$

$$L_{15}(\ )=\frac{\beta(B_{12}+B_{66})J_1^{(uy)}}{A_{11}h}\frac{d(\ )}{d\xi}$$

$$L_{21}(\ )=\frac{\beta(A_{12}+A_{66})J_1^{(vu)}}{A_{11}}\frac{d(\ )}{d\xi}$$

$$L_{22}(\ )=\frac{A_{66}J_0^{(vv)}}{A_{11}}\frac{d^2(\ )}{d\xi^2}+\frac{\beta^2A_{22}J_0^{(vv)}}{A_{11}}(\ )$$

$$L_{23}(\ )=\beta\left(\frac{A_{66}J_{01}^{(vww)}}{A_{11}}\frac{d^2(\ )}{d\xi^2}+\beta^2\frac{A_{22}J_{12}^{(vww)}}{A_{11}}(\ )\right)(\ )+\frac{\beta(A_{12}+A_{66})J_{01}^{(vww)}}{A_{11}}\frac{d(\ )}{d\xi}\frac{d(\ )}{d\xi}$$

$$L_{24}(\ )=\frac{\beta(B_{12}+B_{66})J_1^{(vx)}}{A_{11}h}\frac{d(\ )}{d\xi}$$

$$L_{25}(\ )=\frac{B_{66}J_0^{(vy)}}{A_{11}h}\frac{d^2(\ )}{d\xi^2}+\frac{\beta^2B_{22}J_2^{(vy)}}{A_{11}h}(\ )$$

$$L_{31}(\ )=\left(\frac{\beta^2A_{66}J_{20}^{(wuw)}}{A_{11}}+\frac{2\beta^2A_{66}J_{11}^{(wuw)}}{A_{11}}\right)\frac{d\hat{w}_m}{d\xi}(\ )+\left(J_{00}^{(wuw)}\frac{d^2\hat{w}_m}{d\xi^2}+\frac{\beta^2A_{12}J_{02}^{(wuw)}}{A_{11}}\hat{w}_m+\frac{\beta^2(A_{12}+A_{66})J_{11}^{(wuw)}}{A_{11}}\hat{w}_m\right)\frac{d(\ )}{d\xi}$$
$$+J_{00}^{(wuw)}\frac{d\hat{w}_m}{d\xi}\frac{d^2(\ )}{d\xi^2}$$

$$L_{32}(\ )=\beta\left(\frac{A_{12}J_{10}^{(wvw)}}{A_{11}}\frac{d^2\hat{w}_m}{d\xi^2}+\frac{\beta^2A_{22}(J_{12}^{(wvw)}+J_{21}^{(wvw)})}{A_{11}}\hat{w}_m\right)(\ )+\beta\left(\frac{2A_{66}J_{01}^{(wvw)}}{A_{11}}+\frac{(A_{12}+A_{66})J_{10}^{(wvw)}}{A_{11}}\right)\frac{d\hat{w}_m}{d\xi}\frac{d(\ )}{d\xi}$$
$$+\frac{\beta A_{66}J_{01}^{(wvw)}}{A_{11}}\hat{w}_m\frac{d^2(\ )}{d\xi^2}$$

$$
\begin{aligned}
L_{331}(\;) &= \frac{1}{\gamma^2}\left(-\frac{N_x^* J_0^{(ww)}}{A_{11}}\frac{d^2(\;)}{d\xi^2} - \frac{\beta^2 N_y^* J_2^{(ww)}}{A_{11}}(\;) + \frac{k_5^2 A_{55} J_0^{(ww)}}{A_{11}}\frac{d^2(\;)}{d\xi^2} + \frac{\beta^2 k_4^2 A_{44} J_2^{(ww)}}{A_{11}}(\;)\right)\\
L_{332}(\;) &= \left(\frac{3J_{000}^{(wwww)}}{2}\frac{d^2(\;)}{d\xi^2} + \frac{\beta^2(A_{12}+2A_{66})J_{002}^{(wwww)}}{2A_{11}}(\;)\right)\left(\frac{d(\;)}{d\xi}\right)^2 + \frac{2\beta^2(A_{12}+2A_{66})J_{011}^{(wwww)}}{A_{11}}\frac{d(\;)}{d\xi}\frac{d(\;)}{d\xi}(\;)\\
&\quad + \beta^2\left[\frac{(A_{12}+2A_{66})J_{011}^{(wwww)}}{2A_{11}}\frac{d^2(\;)}{d\xi^2} + \frac{3\beta^2 A_{22} J_{112}^{(wwww)}}{2A_{11}}(\;)\right](\;)^2\\
L_{34}(\;) &= \frac{k_5^2 A_{55} J_0^{(wx)}}{\gamma^2 A_{11}}\frac{d(\;)}{d\xi} + \left(\frac{B_{11}J_{00}^{(wxw)}}{A_{11}h}\frac{d^2\hat{w}_m}{d\xi^2} + \frac{\beta^2 B_{12} J_{02}^{(wxw)}}{A_{11}h}\hat{w}_m\right)\frac{d(\;)}{d\xi} + \frac{2\beta^2 B_{66} J_{11}^{(wxw)}}{A_{11}h}\frac{d\hat{w}_m}{d\xi}(\;) + \frac{B_{11}J_{00}^{(wxw)}}{A_{11}h}\frac{d\hat{w}_m}{d\xi}\frac{d^2(\;)}{d\xi^2}\\
&\quad + \frac{\beta^2(B_{12}+B_{66})J_{11}^{(wxw)}}{A_{11}h}\hat{w}_m\frac{d(\;)}{d\xi} + \frac{\beta^2 B_{66} J_{20}^{(wxw)}}{A_{11}h}\frac{d\hat{w}_m}{d\xi}(\;)\\
L_{35}(\;) &= \frac{\beta k_4^2 A_{44} J_1^{(wy)}}{\gamma^2 A_{11}}(\;) + \frac{2\beta B_{66} J_{01}^{(wyw)}}{A_{11}h}\frac{d\hat{w}_m}{d\xi}\frac{d(\;)}{d\xi} + \beta\left(\frac{B_{12}J_{10}^{(wyw)}}{A_{11}h}\frac{d^2\hat{w}_m}{d\xi^2} + \frac{\beta^2 B_{22} J_{12}^{(wyw)}}{A_{11}h}\hat{w}_m\right)(\;) + \frac{\beta B_{66} J_{01}^{(wyw)}}{A_{11}h}\hat{w}_m\frac{d^2(\;)}{d\xi^2}\\
&\quad + \frac{\beta(B_{12}+B_{66})J_{10}^{(wyw)}}{A_{11}h}\frac{d\hat{w}_m}{d\xi}\frac{d(\;)}{d\xi} + \frac{\beta^3 B_{22} J_{21}^{(wyw)}}{A_{11}h}\hat{w}_m(\;)\\
L_{41}(\;) &= \frac{B_{11}hJ_0^{(xu)}}{D_{11}}\frac{d^2(\;)}{d\xi^2} + \frac{\beta^2 B_{66} h J_2^{(xu)}}{D_{11}}(\;)\\
L_{42}(\;) &= \frac{\beta(B_{12}+B_{66})hJ_1^{(xv)}}{D_{11}}\frac{d(\;)}{d\xi}\\
L_{431}(\;) &= -\frac{k_5^2 A_{55} a^2 J_0^{(xw)}}{D_{11}}\frac{d(\;)}{d\xi}\\
L_{432}(\;) &= \left(\frac{B_{11}hJ_{00}^{(xww)}}{D_{11}}\frac{d^2(\;)}{d\xi^2} + \frac{\beta^2 B_{66} h J_{02}^{(xww)}}{D_{11}}(\;)\right)\frac{d(\;)}{d\xi} + \frac{\beta^2(B_{12}+B_{66})hJ_{11}^{(xww)}}{D_{11}}\frac{d(\;)}{d\xi}(\;)\\
L_{44}(\;) &= J_0^{(xx)}\frac{d^2(\;)}{d\xi^2} + \left(\frac{\beta^2 D_{66} J_2^{(xx)}}{D_{11}}(\;) - \frac{k_5^2 A_{55} a^2 J_0^{(xx)}}{D_{11}}(\;)\right)\\
L_{45}(\;) &= \frac{\beta(D_{12}+D_{66})J_1^{(xy)}}{D_{11}}\frac{d(\;)}{d\xi}\\
L_{51}(\;) &= \frac{\beta(B_{12}+B_{66})hJ_1^{(yu)}}{D_{11}}\frac{d(\;)}{d\xi}\\
L_{52}(\;) &= \frac{B_{66}hJ_0^{(yv)}}{D_{11}}\frac{d^2(\;)}{d\xi^2} + \frac{\beta^2 B_{22} h J_2^{(yv)}}{D_{11}}(\;)\\
L_{531}(\;) &= -\beta\frac{k_4^2 A_{44} a^2 J_1^{(yw)}}{D_{11}}(\;)\\
L_{532}(\;) &= \beta\left(\frac{B_{66}hJ_{01}^{(yww)}}{D_{11}}\frac{d^2(\;)}{d\xi^2} + \frac{\beta^2 B_{22} h J_{12}^{(yww)}}{D_{11}}(\;)\right)(\;) + \frac{\beta(B_{12}+B_{66})hJ_{11}^{(yww)}}{D_{11}}\frac{d(\;)}{d\xi}\frac{d(\;)}{d\xi}\\
L_{54}(\;) &= \frac{\beta(D_{12}+D_{66})J_1^{(yx)}}{D_{11}}\frac{d(\;)}{d\xi}\\
L_{55}(\;) &= \frac{D_{66}J_0^{(yy)}}{D_{11}}\frac{d^2(\;)}{d\xi^2} + \left(\frac{\beta^2 D_{22} J_2^{(yy)}}{D_{11}}(\;) - \frac{k_4^2 A_{44} a^2 J_0^{(yy)}}{D_{11}}(\;)\right)
\end{aligned}
\tag{B1}
$$

in which the laminate stiffness components are calculated by

$$
(A_{ij}, B_{ij}, D_{ij}) = \sum_{k=1}^{N_L}\int_{z_k}^{z_{k+1}} (Q_{ij})^{(k)}(1, z, z^2)\,dz \qquad (i,j=1,2,6)
\tag{B2}
$$

$$A_{ij}=\sum_{k=1}^{N_L} k_i k_j \int_{z_k}^{z_{k+1}} (Q_{ij})^{(k)} dz \quad (i,j=4,5) \qquad (B3)$$

Appendix C

The integral terms in (B-1) are

$$(J_0^{(uu)},J_0^{(uv)},J_0^{(uw)},J_0^{(ux)},J_0^{(uy)}) = \int_0^1 \tilde{u}_m(\tilde{u}_m,\tilde{v}_m,\tilde{w}_m,\tilde{\varphi}_{xm},\tilde{\varphi}_{ym})d\varsigma$$

$$(J_1^{(uu)},J_1^{(uv)},J_1^{(uw)},J_1^{(ux)},J_1^{(uy)}) = \int_0^1 \tilde{u}_m\left(\frac{d\tilde{u}_m}{d\varsigma},\frac{d\tilde{v}_m}{d\varsigma},\frac{d\tilde{w}_m}{d\varsigma},\frac{d\tilde{\varphi}_{xm}}{d\varsigma},\frac{d\tilde{\varphi}_{ym}}{d\varsigma}\right)d\varsigma$$

$$(J_2^{(uu)},J_2^{(uv)},J_2^{(uw)},J_2^{(ux)},J_2^{(uy)}) = \int_0^1 \tilde{u}_m\left(\frac{d^2\tilde{u}_m}{d\varsigma^2},\frac{d^2\tilde{v}_m}{d\varsigma^2},\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d^2\tilde{\varphi}_{xm}}{d\varsigma^2},\frac{d^2\tilde{\varphi}_{ym}}{d\varsigma^2}\right)d\varsigma$$

$$(J_{00}^{(uww)},J_{02}^{(uww)},J_{11}^{(uww)}) = \int_0^1 \tilde{u}_m\left(\tilde{w}_m\tilde{w}_m,\tilde{w}_m\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d\tilde{w}_m}{d\varsigma}\frac{d\tilde{w}_m}{d\varsigma}\right)d\varsigma$$

$$(J_0^{(vu)},J_0^{(vv)},J_0^{(vw)},J_0^{(vx)},J_0^{(vy)}) = \int_0^1 \tilde{v}_m(\tilde{u}_m,\tilde{v}_m,\tilde{w}_m,\tilde{\varphi}_{xm},\tilde{\varphi}_{ym})d\varsigma$$

$$(J_1^{(vu)},J_1^{(vv)},J_1^{(vw)},J_1^{(vx)},J_1^{(vy)}) = \int_0^1 \tilde{v}_m\left(\frac{d\tilde{u}_m}{d\varsigma},\frac{d\tilde{v}_m}{d\varsigma},\frac{d\tilde{w}_m}{d\varsigma},\frac{d\tilde{\varphi}_{xm}}{d\varsigma},\frac{d\tilde{\varphi}_{ym}}{d\varsigma}\right)d\varsigma$$

$$(J_2^{(vu)},J_2^{(vv)},J_2^{(vw)},J_2^{(vx)},J_2^{(vy)}) = \int_0^1 \tilde{v}_m\left(\frac{d^2\tilde{u}_m}{d\varsigma^2},\frac{d^2\tilde{v}_m}{d\varsigma^2},\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d^2\tilde{\varphi}_{xm}}{d\varsigma^2},\frac{d^2\tilde{\varphi}_{ym}}{d\varsigma^2}\right)d\varsigma$$

$$(J_{01}^{(vww)},J_{12}^{(vww)}) = \int_0^1 \tilde{v}_m\left(\tilde{w}_m\frac{d\tilde{w}_m}{d\varsigma},\frac{d\tilde{w}_m}{d\varsigma}\frac{d^2\tilde{w}_m}{d\varsigma^2}\right)d\varsigma$$

$$(J_{00}^{(wuw)},J_{02}^{(wuw)},J_{11}^{(wuw)},J_{20}^{(wuw)}) = \int_0^1 \tilde{w}_m\left(\tilde{u}_m\tilde{w}_m,\tilde{u}_m\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d\tilde{u}_m}{d\varsigma}\frac{d\tilde{w}_m}{d\varsigma},\frac{d^2\tilde{u}_m}{d\varsigma^2}\tilde{w}_m\right)d\varsigma$$

$$(J_{01}^{(wvw)},J_{10}^{(wvw)},J_{12}^{(wvw)},J_{21}^{(wvw)}) = \int_0^1 \tilde{w}_m\left(\tilde{v}_m\frac{d\tilde{w}_m}{d\varsigma},\frac{d\tilde{v}_m}{d\varsigma}\tilde{w}_m,\frac{d\tilde{v}_m}{d\varsigma}\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d^2\tilde{v}_m}{d\varsigma^2}\frac{d\tilde{w}_m}{d\varsigma}\right)d\varsigma$$

$$(J_0^{(ww)},J_2^{(ww)}) = \int_0^1 \tilde{w}_m\left(\tilde{w}_m,\frac{d^2\tilde{w}_m}{d\varsigma^2},(\tilde{w}_m)^3\right)d\varsigma$$

$$(J_{000}^{(wwww)},J_{002}^{(wwww)},J_{011}^{(wwww)},J_{112}^{(wwww)}) = \int_0^1 \tilde{w}_m\left((\tilde{w}_m)^3,(\tilde{w}_m)^2\frac{d^2\tilde{w}_m}{d\varsigma^2},\tilde{w}_m\left(\frac{d\tilde{w}_m}{d\varsigma}\right)^2,\left(\frac{d\tilde{w}_m}{d\varsigma}\right)^2\frac{d^2\tilde{w}_m}{d\varsigma^2}\right) \times d\varsigma$$

$$(J_0^{(wx)},J_{00}^{(wxw)},J_{02}^{(wxw)},J_{11}^{(wxw)},J_{20}^{(wxw)}) = \int_0^1 \tilde{w}_m\left(\tilde{\varphi}_{xm},\tilde{\varphi}_{xm}\tilde{w}_m,\tilde{\varphi}_{xm}\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d\tilde{\varphi}_{xm}}{d\varsigma}\frac{d\tilde{w}_m}{d\varsigma},\frac{d^2\tilde{\varphi}_{xm}}{d\varsigma^2}\tilde{w}_m\right) \times d\varsigma$$

$$(J_1^{(wy)},J_{01}^{(wyw)},J_{10}^{(wyw)},J_{12}^{(wyw)},J_{21}^{(wyw)}) = \int_0^1 \tilde{w}_m\Big(\frac{d\tilde{\varphi}_{ym}}{d\varsigma},\tilde{\varphi}_{ym}\frac{d\tilde{w}_m}{d\varsigma},\frac{d\tilde{\varphi}_{ym}}{d\varsigma}\tilde{w}_m, \frac{d\tilde{\varphi}_{ym}}{d\varsigma}\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d^2\tilde{\varphi}_{ym}}{d\varsigma^2}\frac{d\tilde{w}_m}{d\varsigma}d\varsigma$$

$$(J_0^{(xu)},J_0^{(xv)},J_0^{(xw)},J_0^{(xx)},J_0^{(xy)}) = \int_0^1 \tilde{\varphi}_{xm}(\tilde{u}_m,\tilde{v}_m,\tilde{w}_m,\tilde{\varphi}_{xm},\tilde{\varphi}_{ym})d\varsigma$$

$$(J_1^{(xu)},J_1^{(xv)},J_1^{(xw)},J_1^{(xx)},J_1^{(xy)}) = \int_0^1 \tilde{\varphi}_{xm}\left(\frac{d\tilde{u}_m}{d\varsigma},\frac{d\tilde{v}_m}{d\varsigma},\frac{d\tilde{w}_m}{d\varsigma},\frac{d\tilde{\varphi}_{xm}}{d\varsigma},\frac{d\tilde{\varphi}_{ym}}{d\varsigma}\right)d\varsigma$$

$$(J_2^{(xu)},J_2^{(xv)},J_2^{(xw)},J_2^{(xx)},J_2^{(xy)}) = \int_0^1 \tilde{\varphi}_{xm}\left(\frac{d^2\tilde{u}_m}{d\varsigma^2},\frac{d^2\tilde{v}_m}{d\varsigma^2},\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d^2\tilde{\varphi}_{xm}}{d\varsigma^2},\frac{d^2\tilde{\varphi}_{ym}}{d\varsigma^2}\right)d\varsigma$$

$$(J_{00}^{(xww)},J_{02}^{(xww)},J_{11}^{(xww)}) = \int_0^1 \tilde{\varphi}_{xm}\left(\tilde{w}_m\tilde{w}_m,\tilde{w}_m\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d\tilde{w}_m}{d\varsigma}\frac{d\tilde{w}_m}{d\varsigma}\right)d\varsigma$$

$$(J_0^{(yu)},J_0^{(yv)},J_0^{(yw)},J_0^{(yx)},J_0^{(yy)}) = \int_0^1 \tilde{\varphi}_{ym}(\tilde{u}_m,\tilde{v}_m,\tilde{w}_m,\tilde{\varphi}_{xm},\tilde{\varphi}_{ym})d\varsigma$$

$$(J_1^{(yu)},J_1^{(yv)},J_1^{(yw)},J_1^{(yx)},J_1^{(yy)}) = \int_0^1 \tilde{\varphi}_{ym}\left(\frac{d\tilde{u}_m}{d\varsigma},\frac{d\tilde{v}_m}{d\varsigma},\frac{d\tilde{w}_m}{d\varsigma},\frac{d\tilde{\varphi}_{xm}}{d\varsigma},\frac{d\tilde{\varphi}_{ym}}{d\varsigma}\right)d\varsigma$$

$$(J_2^{(yu)},J_2^{(yv)},J_2^{(yw)},J_2^{(yx)},J_2^{(yy)}) = \int_0^1 \tilde{\varphi}_{ym}\left(\frac{d^2\tilde{u}_m}{d\varsigma^2},\frac{d^2\tilde{v}_m}{d\varsigma^2},\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d^2\tilde{\varphi}_{xm}}{d\varsigma^2},\frac{d^2\tilde{\varphi}_{ym}}{d\varsigma^2}\right)d\varsigma$$

$$(J_{00}^{(yww)},J_{02}^{(yww)},J_{11}^{(yww)}) = \int_0^1 \tilde{\varphi}_{ym}\left(\tilde{w}_m\tilde{w}_m,\tilde{w}_m\frac{d^2\tilde{w}_m}{d\varsigma^2},\frac{d\tilde{w}_m}{d\varsigma}\frac{d\tilde{w}_m}{d\varsigma}\right)d\varsigma \quad (C1)$$

References

[1] Tauchert, T. R., 1991, "Thermally Induced Flexure, Buckling, and Vibration of Plates," Appl. Mech. Rev., **44**(8–9), pp. 347–360.

[2] Noor, A. K., and Burton, W. S., 1992, "Computational Models for High Temperature Multilayered Plates and Shells," Appl. Mech. Rev., **45**(10), pp. 419–446.

[3] Argyris, J., and Tenek, L., 1997, "Recent Advances in Computational Thermostructural Analysis of Composite Plates and Shells With Strong Nonlinearity," Appl. Mech. Rev., **50**(5), pp. 285–307.

[4] Chen, L. W., and Chen, L. Y., 1989, "Thermal Buckling Behaviors of Laminated Composite Plates With Temperature-Dependent Properties," Compos. Struct., **13**(4), pp. 275–287.

[5] Chen, L. W., and Chen, L. Y., 1991, "Thermal Post-Buckling Behaviors of Laminated Composite Plates With Temperature-Dependent Properties," Compos. Struct., **19**(3), pp. 267–283.

[6] Singha, M. K., Ramachandra, L. S., and Bandyopadhyay, J. N., 2001, ''Thermal Post-Buckling Analysis of Laminated Composite Plates,'' Compos. Struct., **54**(4), pp. 453–458.
[7] Argyris, J., and Tenek, L., 1995, ''Post-Buckling of Composite Laminates Under Compressive Loads and Temperature,'' Comput. Methods Appl. Mech. Eng., **128**(1–2), pp. 49–80.
[8] Srikanth, G., and Kumar, A., 2003, ''Post-Buckling Response and Failure of Symmetric Laminates Under Uniform Temperature Rise,'' Compos. Struct., **59**(1), pp. 109–118.
[9] Shen, H. S., 2001, ''Thermal Post-Buckling Behavior of Imperfect Shear Deformable Laminated Plates With Temperature-Dependent Properties,'' Comput. Methods Appl. Mech. Eng., **190**(40–41), pp. 5377–5390.
[10] Shahsiah, R., and Eslami, M. R., 2003, ''Thermal Buckling of Functionally Graded Cylindrical Shells,'' J. Therm. Stresses, **26**(3), pp. 277–294.
[11] Shahsiah, R., and Eslami, M. R., 2003, ''Functionally Graded Cylindrical Shell Thermal Instability Based on Improved Donnell Equations,'' AIAA J., **41**(9), pp. 1819–1826.
[12] Shen, H. S., 2002, ''Postbuckling Analysis of Axially Loaded Functionally Graded Cylindrical Shells in Thermal Environments,'' Compos. Sci. Technol., **62**(7–8), pp. 977–987.
[13] Shen, H. S., 2003, ''Postbuckling Analysis of Pressure-Loaded Functionally Graded Cylindrical Shells in Thermal Environments,'' Eng. Struct., **25**(4), pp. 487–497.
[14] Shen, H. S., 2002, ''Postbuckling Analysis of Axially Loaded Functionally Graded Cylindrical Panels in Thermal Environments,'' Int. J. Solids Struct., **39**(24), pp. 5991–6010.
[15] Shen, H. S., and Leung, A. Y. T., 2003, ''Postbuckling of Pressure-Loaded Functionally Graded Cylindrical Panels in Thermal Environments,'' J. Eng. Mech., **129**(4), pp. 414–425.
[16] Javaheri, R., and Eslami, M. R., 2002, ''Buckling of Functionally Graded Plates Under In-Plane Compressive Loading,'' Z. Angew. Math. Mech., **82**(4), pp. 277–283.
[17] Najafizadeh, M. M., and Eslami, M. R., 2002, ''Buckling Analysis of Circular Plates of Functionally Graded Materials Under Uniform Radial Compression,'' Int. J. Mech. Sci., **44**(12), pp. 2479–2493.
[18] Javaheri, R., and Eslami, M. R., 2002, ''Thermal Buckling of Functionally Graded Plates,'' AIAA J., **40**(1), pp. 162–169.
[19] Javaheri, R., and Eslami, M. R., 2002, ''Thermal Buckling of Functionally Graded Plates Based on Higher Order Theory,'' J. Therm. Stresses, **25**(7), pp. 603–625.
[20] Najafizadeh, M. M., and Eslami, M. R., 2002, ''First-Order-Theory-Based Thermoelastic Stability of Functionally Graded Material Circular Plates,'' AIAA J., **40**(7), pp. 1444–1450.
[21] Qatu, M. S., and Leissa, A. W., 1993, ''Buckling or Transverse Deflections of Unsymmetrically Laminated Plates Subjected to In-Plane Loads,'' AIAA J., **31**(1), pp. 189–194.
[22] Yang, J., and Shen, H. S., 2003, ''Non-Linear Analysis of Functionally Graded Plates Under Transverse and In-Plane Loads,'' Int. J. Non-Linear Mech., **38**(4), pp. 467–482.
[23] Liew, K. M., Yang, J., and Kitipornchai, S., 2003, ''Post-Buckling of Piezoelectric FGM Plates Subjected to Thermo-Electro-Mechanical Loading,'' Int. J. Solids Struct., **40**(15), pp. 881–896.
[24] Jang, S. K., Bert, C. W., and Striz, A. G., 1989, ''Application of Differential Quadrature to Static Analysis of Structural Components,'' Int. J. Numer. Methods Eng., **28**(3), pp. 561–577.
[25] Liew, K. M., Ng, T. Y., and Zhang, J. Z., 2002, ''Differential Quadrature-Layerwise Modeling Technique for Three-Dimensional Analysis of Cross-ply Laminated Plates of Various Edge-Supports,'' Comput. Methods Appl. Mech. Eng., **191**(35), pp. 3811–3832.
[26] Touloukian, Y. S., 1967, *Thermophysical Properties of High Temperature Solid Materials*, McMilan, New York.
[27] Gauthier, M. M. (ed) 1995, *Engineering Materials Handbook-Desk Edition*, ASM International, USA.
[28] Noor, A. K., and Burton, W. S., 1992, ''Three-Dimensional Solutions for Thermal Buckling of Multilayered Anisotropic Plates,'' J. Eng. Mech., **118**(4), pp. 683–701.
[29] Shen, H. S., 1998, ''Thermal Postbuckling Analysis of Imperfect Reissner-Mindlin Plates on Softening Nonlinear Elastic Foundations,'' J. Eng. Math., **33**(3), pp. 259–270.

George G. Adams
Professor,
Fellow ASME,
Department of Mechanical and Industrial Engineering,
Northeastern University,
Boston, MA 02115
e-mail: adams@neu.edu

Adhesion at the Wavy Contact Interface Between Two Elastic Bodies

The plane strain elastic contact of two bodies with a wavy contact interface is investigated. The effect of adhesion is accounted for by using the Maugis model. This periodic mixed boundary value problem is solved using integral transform techniques. Results are obtained for the extent of the contact region as a function of the dimensionless applied pressure and for various values of the dimensionless adhesive stress and peak-to-valley height. These contact length versus applied pressure curves are characterized by discontinuities and hysteresis. A finite contact region exists at zero load, with further loading causing one or more jumps into a complete contact configuration. Unloading is also characterized by one or more jumps before pull-off occurs suddenly with a finite contact length and tensile pressure loading. [DOI: 10.1115/1.1794702]

Introduction

The frictionless two-dimensional elastic contact problem for an elastic half-plane loaded by a periodic system of rigid flat punches was solved by Sadowski [1]. Westergaard [2], using a complex stress function, found a closed-form solution for the two-dimensional contact problem of an elastic half-space with a wavy surface. He obtained an expression for contact stresses as well as for the dependence of the contact area on the applied pressure. Independently, Shtaerman [3] obtained the same result, using the Green's function method in order to formulate the contact problem as an integral equation for the normal contact stress. He found a general form of the mathematical solution for an arbitrary periodic contact profile and a particular solution for a sinusoidal profile. He also obtained an integral equation formulation for the frictional contact problem and a general form of the solution. Dundurs, Tsai, and Keer [4], using Fourier analysis in a stress function approach, produced a series solution to the frictionless problem with a wavy surface, which reduces to the form obtained by Westergaard. Kuznetsov [5] obtained a solution of the same problem, with one rigid body, by using the complex potential method of Muskhelishvili. He also solved the frictional (low-velocity) problem [6]. Nosonovsky and Adams [7] investigated high-speed steady-state frictional sliding of the wavy contact interface.

Johnson, Greenwood, and Higginson [8] obtained a numerical solution, as well as asymptotic solutions for small and large zones of contact, for the frictionless case of two-dimensional waviness (three-dimensional elasticity). Their method was based on Fourier analysis. Manners [9] extended Westergaard's problem to contact surfaces with more complicated periodic profiles.

For sufficiently small size contacts, the adhesion forces between the surfaces affect the contact conditions. Various adhesion models, between an elastic sphere and a flat, have been introduced. The model by Johnson, Kendall, and Roberts (JKR) [10] assumes that the attractive intermolecular surface forces cause elastic deformation beyond that predicted by the Hertz theory, producing a subsequent increase of the contact area. This model also assumes that the attractive forces are confined to the contact area and vanish outside the contact area. The model by Derjaguin, Muller, and Toporov (DMT) [11] assumes that the contact displacement and stress profiles remain the same as in the Hertz theory. However, these quantities are calculated for a higher effective load which includes the applied normal force as well as the attractive adhesive stresses acting outside of the contact area. Due to the assumptions involved, the JKR/DMT models are most suitable when the range of surface forces is small/large compared to the elastic deformations, as pointed out by Tabor [12]. Another model, introduced by Maugis [13], describes a continuous transition between the JKR and DMT models. In order to represent the surface forces, Maugis used a Dugdale approximation in which the attractive stress is constant for surface separations up to a prescribed value h_0. Intimate contact is maintained over the central region; an adhesive stress of constant magnitude acts in an annular ring outside of the contact zone for which the local separation is less than h_0; and the remainder of the separation region is stress free.

Contact problems with a wavy interface, such as [1–9], are relevant to applications such as seals (in which it is desirable to reduce the separation between the surfaces) and in understanding friction (which is related to the real area of contact). As length scales decrease and surfaces become smoother in modern applications such as microelectromechanical (MEMS) and nanoelectromechanical (NEMS), it becomes important to account for the effect of adhesion on the contact area in these problems. In this investigation, the effect of adhesion on the contact problem between two elastic bodies with a wavy contact interface is investigated. Plane strain linear elasticity is used along with the Maugis model of adhesion. Due to the nonlinear nature of adhesion, the contact area during unloading differs from that encountered during loading.

Formulation of the Problem

Consider the plane strain frictionless contact of two semi-infinite elastic bodies, one of which is flat and the other of which has a slightly wavy surface. The bodies are pressed together by a uniform normal traction p applied at infinity (Fig. 1). The lower body is assumed to have a sinusoidal upper surface with a period of 2ℓ and peak-to-valley amplitude g, such that the gap prior to loading $y_0(x)$ is given by

$$y_0(x)=\frac{g}{2}\left(1-\cos\frac{\pi x}{\ell}\right) \tag{1}$$

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, August 8, 2003; final revision, April 3, 2004. 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 of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

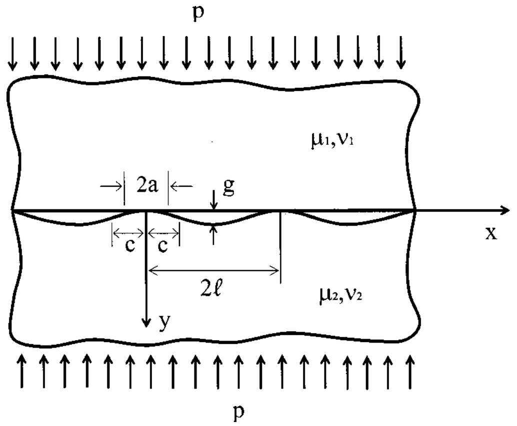


Fig. 1 Two elastic half-spaces with a wavy contact interface subjected to a remotely applied pressure (*p*)

It is assumed that the amplitude of the waviness is small compared to its wavelength, i.e., $g \ll \ell$, so that the lower body may be treated as an elastic half-space with its boundary at $y=0$. From this point forward, the wavelength will be taken as $\ell=\pi$ without loss of generality.

The Navier equations of plane strain elasticity for the displacement components $u(x,y)$, $v(x,y)$ in the x and y directions, respectively, are given by

$$(\lambda_k+2\mu_k)\frac{\partial^2 u^{(k)}}{\partial x^2}+\mu_k\frac{\partial^2 u^{(k)}}{\partial y^2}+(\lambda_k+\mu_k)\frac{\partial^2 v^{(k)}}{\partial x\partial y}=0$$

$$(\lambda_k+2\mu_k)\frac{\partial^2 v^{(k)}}{\partial y^2}+\mu_k\frac{\partial^2 v^{(k)}}{\partial x^2}+(\lambda_k+\mu_k)\frac{\partial^2 u^{(k)}}{\partial x\partial y}=0 \quad (2)$$

in which $k=1$ for the upper body and $k=2$ for the lower body. Due to periodicity, only the strip $-\pi<x<\pi$ need be considered.

The boundary conditions at infinity ($|y|=\infty$) state that stresses approach the uniform state of compression, i.e.,

$$\sigma_{yy}^{(1)}=\sigma_{yy}^{(2)}=-p, \quad \sigma_{xy}^{(1)}=\sigma_{xy}^{(2)}=0, \quad \text{for} \quad -\pi<x<\pi \quad (3)$$

The following boundary conditions pertain to the contact interface ($y=0$):

$$\sigma_{xy}^{(1)}=\sigma_{xy}^{(2)}=0, \quad -\pi<x<\pi \quad (4)$$

$$\sigma_{yy}^{(1)}=\sigma_{yy}^{(2)}, \quad -\pi<x<\pi \quad (5)$$

$$\frac{\partial v^{(1)}}{\partial x}-\frac{\partial v^{(2)}}{\partial x}=\frac{g}{2}\sin x, \quad -a<x<a \quad (6)$$

$$\sigma_{yy}^{(1)}=\sigma_0, \quad a<|x|<c \quad (7)$$

$$\sigma_{yy}^{(1)}=0, \quad c<|x|<\pi \quad (8)$$

The condition of vanishing shear stress (4) is exactly valid only for frictionless contact or for material combinations such that normal/shear stresses do not produce relative tangential/normal displacements. The latter case holds for identical materials; for one material rigid and the other incompressible; for two incompressible materials; or more generally, if the Dundurs bimaterial parameter β, defined in [14], vanishes. However, even if none of these conditions hold, the coupling is generally a small effect [15] and Eq. (4) can be used as an approximation. Condition (6) comes from taking the derivative of $v^{(1)}(x,0)-v^{(2)}(x,0)=y_0(x)$, i.e., the gap vanishes in the contact region.

Condition (7) arises from the effect of adhesion using the Maugis model [13] in which the adhesive tensile stress is a constant (σ_0) in regions in which the local separation is positive but less than h_0. The adhesive stress is taken to be equal to the theoretical strength and by equating the work of adhesion w to the quantity $\sigma_0 h$ (the work of adhesion in the Maugis model). The value of h_0 is approximately equal to the lattice spacing [13]. Thus, the additional condition needed to determine the extra unknown c is obtained from

$$y_0(a)-[v^{(1)}(a,0)-v^{(2)}(a,0)]=0$$

and

$$y_0(c)-[v^{(1)}(c,0)-v^{(2)}(c,0)]=h_0 \quad (9)$$

which states that the interface gap vanishes at $x=a$ and is equal to h_0 at $x=c$. The second of these equations is subtracted from the first, which leads to

$$\int_a^c\left(\frac{\partial v^{(1)}}{\partial x}-\frac{\partial v^{(2)}}{\partial x}\right)dx=y_0(c)-y_0(a)-h_0 \quad (10)$$

in which the normal displacements have been written as the integral of their derivatives.

In addition, the stress and displacement fields are subject to certain inequality constraints on the interface, namely,

$$\sigma_{yy}^{(1)}\leqslant\sigma_0, \quad -a<x<a \quad (11)$$

in the contact region, and

$$v^{(1)}-v^{(2)}\leqslant y_0, \quad a<|x|<\pi \quad (12)$$

in the separation zones.

The solution of the Navier equations (2) can be obtained by using, for example, the Papkovich-Neuber potentials [16]. The solution has been presented by Dundurs et al. [4] and is given as

$$\phi_1=-\frac{py}{2(1-\nu_1)}+\sum_{n=1}^{\infty}a_n n^{-1}e^{ny}\cos nx \quad (13)$$

$$\psi_1=\sum_{n=1}^{\infty}b_n n^{-2}e^{ny}\cos nx \quad (14)$$

$$\phi_2=-\frac{py}{2(1-\nu_2)}+\sum_{n=1}^{\infty}c_n n^{-1}e^{-ny}\cos nx \quad (15)$$

$$\psi_2=\sum_{n=1}^{\infty}d_n n^{-2}e^{-ny}\cos nx \quad (16)$$

where ϕ_k is the y component of the vector potential (the other components are zero), ψ_k is the scalar potential, ν_k are the Poisson ratios, and $k=1,2$ for the upper and lower bodies, respectively. The displacements are related to these potentials by

$$2\mu_k u^{(k)}=-y\frac{\partial\phi_k}{\partial x}-\frac{\partial\psi}{\partial x}$$

$$2\mu_k v^{(k)}=(3-4\nu_k)\phi_k-y\frac{\partial\phi_k}{\partial y}-\frac{\partial\psi}{\partial y} \quad (17)$$

As discussed in [4], the potentials (13)–(16) give the uniform compression field at infinity and satisfy the periodicity conditions

$$u^{(1)}=u^{(2)}=0, \quad \sigma_{xy}^{(1)}=\sigma_{xy}^{(2)}=0 \quad (18)$$

on $x=\pm\pi n$ automatically. Furthermore [4], the boundary conditions (4)–(5) require that

$$b_n(1-2\nu_1)a_n, \quad c_n=-a_n, \quad d_n=(1-2\nu_2)a_n \quad (19)$$

The boundary conditions (6)–(8) lead to

$$\sum_{n=1}^{\infty}a_n\sin nx=-Mg\sin x, \quad -a<x<a \quad (20)$$

$$\sum_{n=1}^{\infty} a_n \cos nx = \sigma_0 + p, \quad a<|x|<c \tag{21}$$

$$\sum_{n=1}^{\infty} a_n \cos nx = p, \quad c<|x|<\pi \tag{22}$$

where

$$M = \frac{(1+\alpha)\mu_1}{4(1-\nu_1)}, \quad \alpha = \frac{\mu_2(1-\nu_1)-\mu_1(1-\nu_2)}{\mu_2(1-\nu_1)+\mu_1(1-\nu_2)} \tag{23}$$

In order to satisfy (22), we introduce the unknown function $\phi(\xi)$ such that

$$a_n = \frac{1}{\pi}\int_{-c}^{c} \phi(\xi)\cos n\xi d\xi \tag{24}$$

Then by substituting (24) into (22), reversing the order of summation and integration, and using the identity [17]

$$\sum_{n=1}^{\infty} \cos nx = -\frac{1}{2} + \pi \sum_{n=0}^{\infty} \delta(x \pm 2\pi n) \tag{25}$$

along with symmetry $(\phi(x)=\phi(-x))$, condition (22) is automatically satisfied provided

$$\int_{-c}^{c} \phi(\xi) d\xi = -2\pi p \tag{26}$$

The interfacial normal stress becomes

$$\sigma_{yy}(x,0) = \phi(x) H(c-|x|) \tag{27}$$

where $H(x)$ is the unit step function. Note that, in view of (27), Eq. (26) may be viewed as an equilibrium condition in the y direction pertaining to a semi-infinite vertical strip of width 2π centered at $x=0$.

Now define another auxiliary function $\psi(x)$ by

$$\psi(x) = \phi(x) - \sigma_0, \quad -c<x<c, \quad \psi(x)=0, \quad a<|x|<c \tag{28}$$

and from (27)–(28), the boundary condition (21) is automatically satisfied.

In order to satisfy the remaining boundary condition (20), we substitute (24) and (28) into (20), reverse the order of summation and integration, and use the identity [17]

$$\sum_{n=1}^{\infty} \sin nx = \frac{1}{2}\cot\frac{x}{2} \tag{29}$$

which leads to the singular integral equation

$$\frac{1}{2\pi}\int_{-a}^{a} [\psi(\xi)+\sigma_0]\cot\frac{x-\xi}{2} d\xi = -Mg \sin x, \quad -a<x<a \tag{30}$$

It is now convenient to introduce the coordinate transformations used by Comninou and Dundurs [18]

$$\tan\frac{\xi}{2} = bu, \quad \tan\frac{x}{2} = bs, \quad b = \tan\frac{a}{2}, \quad d = \tan\frac{b}{2}$$
$$\Psi(u) = \frac{\psi(\xi)}{1+b^2u^2} \tag{31}$$

which, after considerable algebra, reduces the integral Eq. (30) to

$$\frac{1}{\pi}\int_{-1}^{1} \frac{\Psi(u)}{u-s} du = f(s) \tag{32}$$

where

$$f(s) = \left[bs\left(c\sigma_0 + p\pi + \sigma_0 \ln\frac{d+bs}{d-bs}\right)\right] \Big/ \pi(1+b^2s^2) + \frac{2bMgs}{(1+b^2s^2)^2} \tag{33}$$

Furthermore, the resultant condition (26) becomes

$$b\int_{-1}^{1} \Psi(u) du = -(\pi p + \sigma_0 c) \tag{34}$$

Finally, the "adhesion displacement" condition (10) yields, after lengthy algebra

$$-\frac{1}{2M}(I_1+I_2) = \frac{1}{2}g(\cos a - \cos c) - h_0 \tag{35}$$

where

$$I_1 = \frac{2b}{\pi}\int_{-1}^{1} \Psi(u)\ln\left(\frac{d/b-u}{1-u}\right) du + \frac{(\pi p+\sigma_0 c)}{\pi}\ln\left(\frac{1+d^2}{1+b^2}\right)$$

$$I_2 = \frac{\sigma_0}{\pi}\int_a^c \ln\left[\frac{\left(d+\tan\frac{x}{2}\right)(c-x)}{\left(d-\tan\frac{x}{2}\right)2\cos^2\frac{c}{2}}\right] dx + \frac{\sigma_0(c-a)}{\pi}\left[\ln\left(\frac{2\cos^2\frac{c}{2}}{c-a}\right)+1\right] \tag{36}$$

In order to solve the singular integral Eq. (32), which is subject to (34–35), the procedure developed by Erdogan and Gupta [19] is used. The unknown $\Psi(u)$ is taken as

$$\Psi(u) = (1-u^2)^{1/2} g(u) \tag{37}$$

in which $g(u)$ and its derivative are bounded continuous functions in the closed interval $-1 \leqslant u \leqslant 1$. Then (32–35) are approximated by linear algebraic equations. For example, (32) becomes

$$\sum_{i=1}^{n} \frac{1-u_i^2}{n+1}\frac{g(u_i)}{u_i-s_j} = f(s_j), \quad j=1,2,\ldots n+1$$

$$u_i = \cos\left(\frac{i\pi}{n+1}\right), \quad i=1,2,\ldots n \tag{38}$$

$$s_j = \cos\left(\frac{\pi}{2}\frac{2j-1}{n+1}\right), \quad j=1,2,\ldots n+1$$

in which the middle equation is neglected due to symmetry [19].

The analysis thus far has implicitly assumed that $c<\pi$, which need not always be the case. Solutions can also be obtained for $c=\pi$, in which case the adhesion displacement condition (10) is replaced with the inequality constraint given in terms of the maximum gap opening h_M by

$$h_M \equiv y_0(\pi) - y_0(a) - \int_a^{\pi}\left(\frac{\partial v^{(1)}}{\partial x} - \frac{\partial v^{(2)}}{\partial x}\right) dx \leqslant h_0 \tag{39}$$

Using (35), this condition becomes

$$h_M = \frac{1}{2}g(1+\cos a) + \frac{I_1' + I_2'}{2M} \tag{40}$$

where

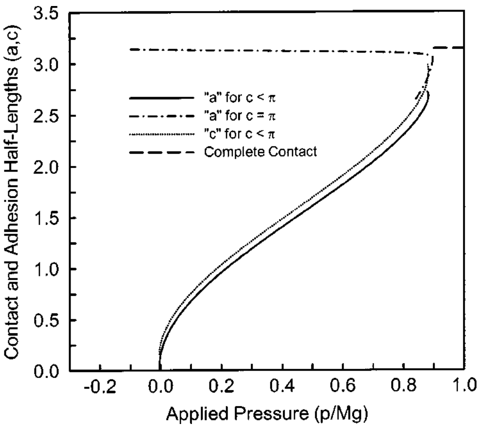


Fig. 2 Contact half-length (*a*) and adhesion half-length (*c*) versus applied pressure (*p*/*Mg*) for $\sigma_0/Mg=0.1$ and $g/h_0=100$

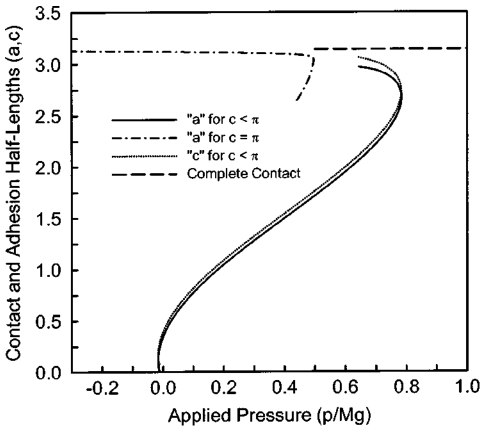


Fig. 3 Contact half-length (*a*) and adhesion half-length (*c*) versus applied pressure (*p*/*Mg*) for $\sigma_0/Mg=0.5$ and $g/h_0=100$

$$I_1'=-\frac{2b}{\pi}\int_{-1}^{1}\Psi(u)\ln[b(1-u)]du-\frac{(\pi p+\sigma_0 c)}{\pi}\ln(1+b^2), \tag{41}$$

$$I_2'=0$$

It is noted that by symmetry, the maximum gap opening defined in (40) occurs at $x=\pi$ and this quantity is useful to know regardless of the value of c. For $c<\pi$, the maximum gap opening h_M is also given by (40) with I_1' given by (41) but with

$$I_2'=\frac{\sigma_0}{\pi}\int_a^c \ln\left[\frac{\left(d+\tan\frac{x}{2}\right)(c-x)}{\left(d-\tan\frac{x}{2}\right)2\cos^2\frac{c}{2}}\right]dx$$
$$+\frac{\sigma_0}{\pi}\left\{(\pi-a)\left[\ln\left(2\cos^2\frac{c}{2}\right)+1\right]\right.$$
$$\left.-(\pi-c)\ln(\pi-c)-(c-a)\ln(c-a)\right\} \tag{42}$$

Another possibility is that the surfaces are in a state of complete contact, in which case $a=c=\pi$. Complete contact requires that the contact pressure cause the inequality (7) to be satisfied on the entire interface. Thus, this state occurs provided that

$$\frac{p}{Mg}\geqslant 1-\frac{\sigma_0}{Mg} \tag{43}$$

In summary, three different types of solutions are possible. In the first case, $c<\pi$ and there exists a region ($c<|x|<\pi$) for which the gap opening exceeds h_0. In the second case, ($c=\pi$) the gap opening is everywhere less than h_0. Finally, the third case corresponds to complete contact of the wavy and flat surfaces.

Numerical Results and Discussion

In Figs. 2–7 are graphs of the dimensionless contact half-length (a) and adhesion half-length (c) as functions of the nondimensional remotely applied pressure p/Mg for various values of dimensionless adhesive stress σ_0/Mg and peak-to-valley height g/h_0. Discontinuities and hysteresis in the loading and unloading curves are observed in all of these graphs.

Figures 2–4 all pertain to $g/h_0=100$ and to various values of the adhesive stress σ_0/Mg. Consider, for example, Fig. 3 for which $\sigma_0/Mg=0.5$. At zero pressure, adhesion causes the surfaces to meet with a finite value of the contact half-length ($a\approx 0.35$) found by the intersection of a vertical line at $p=0$ with the solid curve. The corresponding values of c versus pressure are shown by the dotted line. Note that c is only slightly greater than a. For increasing pressure, a increases until the tangent to its curve becomes vertical. A further increase in pressure causes a jump to complete contact given by the dashed horizontal line. Complete contact persists for increasing contact pressure. During unloading, complete contact continues until reaching the left end of the horizontal line, beyond which point the complete contact solution (43) is no longer valid. A further decrease in pressure causes a jump to the next stable equilibrium position, which is the

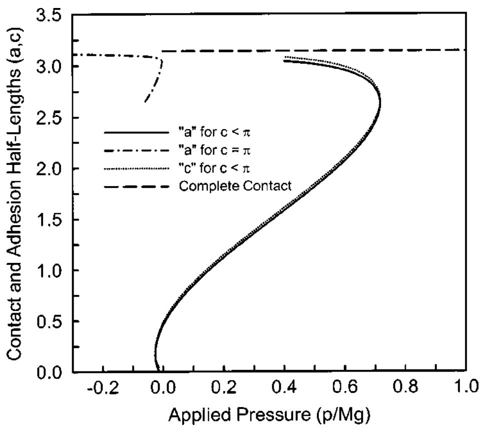


Fig. 4 Contact half-length (*a*) and adhesion half-length (*c*) versus applied pressure (*p*/*Mg*) for $\sigma_0/Mg=1.0$ and $g/h_0=100$

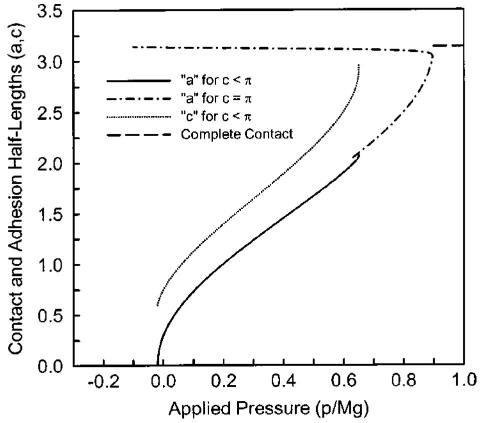


Fig. 5 Contact half-length (*a*) and adhesion half-length (*c*) versus applied pressure (*p*/*Mg*) for $\sigma_0/Mg=0.1$ and $g/h_0=10$

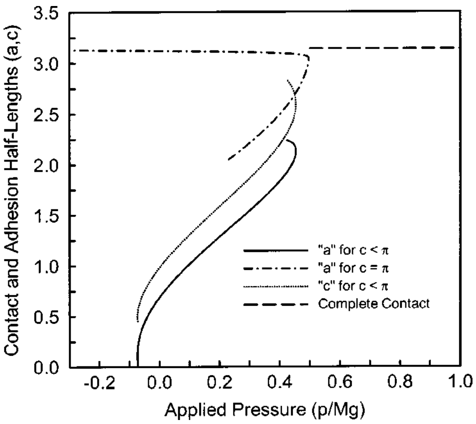


Fig. 6 Contact half-length (a) and adhesion half-length (c) versus applied pressure (p/Mg) for $\sigma_0/Mg=0.5$ and $g/h_0=10$

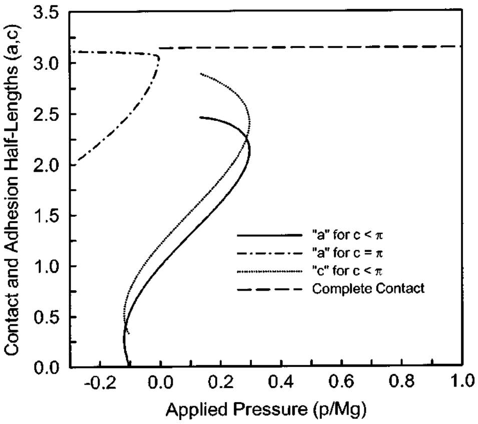


Fig. 7 Contact half-length (a) and adhesion half-length (c) versus applied pressure (p/Mg) for $\sigma_0/Mg=1.0$ and $g/h_0=10$

lower portion of the dot-dashed curve.[1] The entire dot-dashed curve corresponds to $c=\pi$, which requires that the opening gap be everywhere less than the range of the adhesive stress h_0. A further decrease in pressure causes another jump down to the solid line until separation finally occurs at a tensile value of the pressure corresponding to a snap out-of-contact, i.e. a pull-off pressure.

The behavior exhibited in Figs. 2 and 4 are similar to Fig. 3, except for the following differences. For the lower adhesive stress case (i.e., Fig. 2 which corresponds to $\sigma_0/Mg=0.1$), loading jumps from the solid line (a for which $c<\pi$) to the dot-dashed curve (a for which $c=\pi$) before jumping into complete contact. Unloading is qualitatively similar to Fig. 3 but with less hysteresis, as may be expected for this lower adhesion case. For the higher adhesive stress case (i.e., Fig. 4 which corresponds to $\sigma_0/Mg=1.0$), it is the unloading which differs from Fig. 3. After unloading to the left-most point on the lower portion of the dot-dashed line, the bodies jump completely out of contact.

Figures 5–7 all pertain to $g/h_0=10$ (i.e., smoother surfaces than in Figs. 2–4) and to the same values of the adhesive stress σ_0/Mg corresponding to Figs. 2–4, respectively. In all cases, $c-a$ is greater than its corresponding value for the rougher surface. Also the jump into contact occurs with a greater value of a and the stable portion of the dot-dashed curves tend to be more pronounced than in Figs. 2–4.

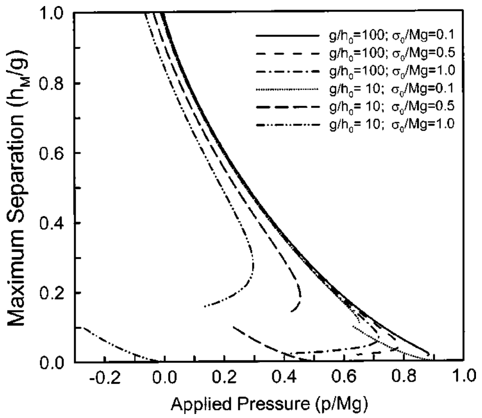


Fig. 8 Maximum separation (h_M/g) versus applied pressure (p/Mg) for various values of σ_0/Mg and g/h_0

Finally, the dimensionless maximum gap opening displacement h_M/g as a function of dimensionless pressure is shown in Fig. 8 for each of the six cases presented. Note that for the three cases corresponding to larger initial gaps (i.e., $g/h_0=100$), the maximum separation is only weakly affected by the adhesive stress. However, for smaller gaps (i.e., $g/h_0=10$), the adhesive stress has a more important effect on the maximum separation. The lower three curves, each of which correspond to $g/h_0=10$ and $c=\pi$, begin at $h_M=h_0$, i.e., $h_M/g=0.1$. The corresponding curves for $g/h_0=100$ are not visible due to the vertical scale of the graph.

Conclusions

The plane strain elastic contact of two bodies with a wavy contact interface has been investigated. The effect of adhesion is accounted for using the Maugis model. Results are obtained for the contact region as a function of the dimensionless applied pressure and for various values of the dimensionless adhesive stress and peak-to-value roughness. The pressure versus contact length curves are characterized by discontinuities and hysteresis. A finite contact region exists at zero load with further loading causing one or more jumps into a complete contact configuration. Unloading is also characterized by one or more jumps before pull-off occurs suddenly with a finite contact length and tensile pressure loading. Adhesion has a greater effect on the maximum surface separation for smoother surfaces.

References

[1] Sadowski, M. A., 1928, “Zwiedimensionale Probleme der Elastizitatshtheorie,” Z. Angew. Math. Mech., **8**, pp. 107–121.
[2] Westergaard, H. M., 1939, “Bearing Pressures and Cracks,” ASME J. Appl. Mech., **6**, pp. A49–A53.
[3] Shtaerman, I. Ya., 1949, *Contact Problem in the Theory of Elasticity* (in Russian), Gostehizdat, Moscow.
[4] Dundurs, J., Tsai, K. C., and Keer, L. M., 1973, “Contact Between Elastic Bodies With Wavy Surfaces,” J. Elast., **3**, pp. 109–115.
[5] Kuznetsov, E. A., 1985, “Effect of Fluid Lubricant on the Contact Characteristics of Rough Elastic Bodies in Compression,” Wear, **157**, pp. 177–194.
[6] Kuznetsov, E. A., 1976, “Periodic Contact Problem for Half-Plane Allowing for Forces of Friction,” Sov. Appl. Mech., **12**, pp. 37–44.
[7] Nosonovsky, M., and Adams, G. G., 2000, “Steady State Frictional Sliding of Two Elastic Bodies With a Wavy Contact Interface,” ASME J. Tribol., **122**, pp. 490–495.
[8] Johnson, K. L., Greenwood, J. A., and Higginson, J. G., 1985, “The Contact of Elastic Regular Wavy Surfaces,” Int. J. Mech. Sci., **27**, pp. 383–396.
[9] Manners, W., 1998, “Partial Contact Between Elastic Surfaces With Periodic Profiles,” Proc. R. Soc. London, Ser. A, **454**, pp. 3203–3221.
[10] 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. 301–313.
[11] Derjaguin, B. V., Muller, V. M., and Toporov, Y. P., 1975, “Effect of Contact

---

[1] Calculations show that in the lower/upper portions of that curve, the opening gap decreases/increases as the pressure increases. Thus, it is only the lower portion of each of these curves that is expected to be stable and therefore physically relevant.

Deformations on the Adhesion of Particles,'' J. Colloid Interface Sci., **53**, pp. 314–326.

[12] Tabor, D., 1976, ''Surface Forces and Surface Interactions,'' J. Colloid Interface Sci., **58**, pp. 2–13.

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

[14] Dundurs, J., 1969, Discussion on ''Edge Bonded Dissimiliar Orthogonal Elastic Wedges Under Normal and Shear Loading,'' ASME J. Appl. Mech., **36**, pp. 650–652.

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

[16] Sokolnikoff, I. S., 1956, *Mathematical Theory of Elasticity*, Second Edition, McGraw-Hill, New York.

[17] Gelfand, I. M., and Shilov, G. E., 1964, *Generalized Functions*, Vol. 1, Academic Press, New York.

[18] Comninou, M., and Dundurs, J., 1977, ''Elastic Interface Waves Involving Separation,'' ASME J. Appl. Mech., **44**, pp. 222–226.

[19] Erdogan, F., and Gupta, G. D., 1972, ''On the Numerical Solution of Singular Integral Equations,'' Q. Appl. Math., **29**, pp. 525–534.

Metin Muradoglu[1]
Department of Mechanical Engineering,
Koc University,
Rumelifeneri Yolu,
Sariyer, Istanbul 34450, Turkey
e-mail: mmuradoglu@ku.edu.tr

Seckin Gokaltun[2]
Computational Science and Engineering
Program,
Informatics Institute,
Istanbul Technical University,
Maslak, Sariyer, Istanbul 34469, Turkey
e-mail: gokaltunse@itu.edu.tr

Implicit Multigrid Computations of Buoyant Drops Through Sinusoidal Constrictions

Two-dimensional computations of dispersed multiphase flows involving complex geometries are presented. The numerical algorithm is based on the front-tracking method in which one set of governing equations is written for the whole computational domain and different phases are treated as a single fluid with variable material properties. The front-tracking methodology is combined with a newly developed finite volume solver based on dual time-stepping, diagonalized alternating direction implicit multigrid method. The method is first validated for a freely rising drop in a straight channel, and it is then used to compute a freely rising drop in various constricted channels. Interaction of two buoyancy-driven drops in a continuously constricted channel is also presented.

[DOI: 10.1115/1.1795222]

1 Introduction

Dynamics of dispersed bubbles or drops in capillary flows involving complex geometries has attracted considerable interest due to its applications in enhanced oil recovery, hazardous waste management, microfluidic devices, and biological systems [1–3].

The presence of deforming phases makes the multiphase flow computations a challenging task, and strong interactions between the phases and complex geometries add further complexity to the problem. Therefore, the progress was rather slow and the computations of multiphase flows have been usually restricted to simple geometries [4] or to moderately complex geometries in the limiting case of creeping flow regime [5,6]. Since nearly all-multiphase flows of practical importance involve complex geometries, it is of obvious interest to extend the modeling and computational techniques to treat multiphase flows in arbitrarily complex geometries.

The motion of a drop in a constricted capillary tube has been studied experimentally by Olbricht and Leal [1], Olbricht and Kung [7], and Hemmat and Borhan [2], and computationally in the creeping flow regime by Tsai and Miksis [5] and Magna [6]. Udaykumar et al. [3] performed computations of the motion of droplets in a constricted channel at finite Reynolds numbers by using a mixed Eulerian-Lagrangian method.

In the present work, a finite-volume/front-tracking (FV/FT) method is used to simulate dynamics of two-dimensional drops rising due to buoyancy in various constricted channels. The front-tracking (FT) method developed by Unverdi and Tryggvason [8] is incorporated into a newly developed finite-volume (FV) algorithm in order to facilitate efficient and accurate simulations of dispersed multiphase flows in arbitrarily complex geometries. The front-tracking method is based on writing one set of governing equations for the whole computational domain and treating different phases as a single fluid with variable material properties. In this method, the fronts are explicitly tracked in a Lagrangian frame and the effects of surface tension are accounted for by treating them as body forces. The front tracking method has been successfully applied to a variety of dispersed multiphase flow problems, but all in relatively simple geometries. A detailed description of the front-tracking method can be found in the review paper by Tryggvason et al. [4]. The FV method used in the present work is based on the concept of the dual (or pseudo) time-stepping method and is developed for unsteady computations of incompressible laminar flows. The dual time-stepping method uses subiterations in pseudotime and has a number of advantages, such as direct coupling of the continuity and momentum equations in incompressible flow equations, the elimination of factorization error in factored implicit schemes, the elimination of errors due to approximations made in the implicit operator to improve numerical efficiency, the elimination of errors due to lagged boundary conditions at both solid and internal fluid boundaries, and the ability to use nonphysical, preconditioned iterative methods for more efficient convergence of the subiterations as discussed by Caughey [9]. In order to combine the front-tracking methodology with the FV method, an algorithm is developed for tracking the front in curvilinear grids and is found to be very efficient and robust. The details of the present FV/FT method can be found in Muradoglu and Kayaalp [10].

The paper begins with a brief description of the governing equations and the numerical solution algorithm. The results are then presented and discussed in Section 4. The present FV/FT method is first validated for a freely rising drop in a straight channel, and the results are compared with the results of the finite-difference/front-tracking (FD/FT) method implemented in the FTC2D code of Unverdi and Tryggvason [8]. It is then applied to a single drop rising in various constricted channels. Interactions of two identical drops are also studied in the continuously constricted channel. Finally, some conclusions are drawn in Section 5.

2 Mathematical Formulation

Following Unverdi and Tryggvason [8], the Navier-Stokes equations are written for the whole flow field, and different phases are treated with variable material properties. The effects of surface tension are modeled as body forces and are included in the momentum equations as δ functions at the phase boundaries. In the

---

[1] To whom correspondence should be addressed.

[2] Present address: Department of Mechanical Engineering, Florida International University, EAS 2417, 10855 W. Flagler Street, Miami, FL 33174.

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 25, 2003; final revision; June 17, 2004. 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 of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

Cartesian coordinates, two-dimensional time-dependent Navier-Stokes equations for incompressible flow can be written in conservative form as

$$\frac{\partial \mathbf{q}}{\partial t}+\frac{\partial \mathbf{f}}{\partial x}+\frac{\partial \mathbf{g}}{\partial y}=\frac{\partial \mathbf{f_v}}{\partial x}+\frac{\partial \mathbf{g_v}}{\partial y}-\Delta\rho\mathbf{G}+\int \delta(\mathbf{x}-\mathbf{x}_\mathrm{f})\sigma\kappa\mathbf{n}ds \quad (1)$$

where

$$\mathbf{q}=\begin{pmatrix}0\\ \rho u\\ \rho v\end{pmatrix},\quad \mathbf{f}=\begin{pmatrix}u\\ p+\rho u^2\\ \rho uv\end{pmatrix},\quad \mathbf{g}=\begin{pmatrix}v\\ \rho uv\\ p+\rho v^2\end{pmatrix},$$

$$\mathbf{f}_\mathrm{v}=\begin{pmatrix}0\\ \tau_{xx}\\ \tau_{xy}\end{pmatrix},\quad \mathbf{g}_\mathrm{v}=\begin{pmatrix}0\\ \tau_{xy}\\ \tau_{yy}\end{pmatrix} \quad (2)$$

and the viscous stresses are given for a Newtonian fluid as

$$\tau_{xx}=2\mu\frac{\partial u}{\partial x},\quad \tau_{xy}=\mu\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right),\quad \tau_{yy}=2\mu\frac{\partial v}{\partial y} \quad (3)$$

In Eqs. (1)–(3), u, v, p, ρ, and μ denote the velocity components in x and y directions, the pressure, the density, and the dynamic viscosity, respectively. The third term on the right-hand side of Eq. (1) represents the body force due to buoyancy with $\mathbf{G}$ being the gravitational acceleration and $\Delta\rho=\rho_o-\rho$, where ρ_o is the density of the ambient fluid. The last term represents the effects of the surface tension and δ, $\mathbf{x_f}$, σ, κ, $\mathbf{n}$ and ds denote the Dirac delta function, the location of the front, the surface tension coefficient, the curvature, the outward unit normal vector on the interface, and the arc length along the interface, respectively.

The fluids in and out of the drop are assumed to be incompressible, and the effects of heat transfer are neglected. Therefore, the viscosity and the density remain constant in each fluid particle, i.e.,

$$\frac{D\rho}{Dt}=0,\quad \frac{D\mu}{Dt}=0 \quad (4)$$

The flow regime of bubbly flows is characterized by four nondimensional parameters as discussed by Clift et al. [11]. These are the Morton number $M=\mu_o^4(\rho_o-\rho_b)G/\rho_o^2\sigma^3$, the Eötvös number $Eo=(\rho_o-\rho_b)d_e^2G/\sigma$, the density ratio $\gamma=\rho_b/\rho_o$, and the viscosity ratio $\zeta=\mu_b/\mu_o$, where d_e is the equivalent drop diameter and the subscripts o and b refer to the ambient and the drop fluids, respectively. The Reynolds number is defined as $\mathrm{Re}=\rho_o V d_e/\mu_o$, where V is the rise velocity.

3 Numerical Procedure

As can be seen in Eq. (1), the continuity equation is decoupled from the momentum equations because it does not have any time derivative term in incompressible flows. To circumvent this difficulty and to be able to use time-marching algorithms, pseudotime derivative terms augmented with a preconditioning matrix are added to Eq. (1) yielding

$$\Gamma^{-1}\frac{\partial \mathbf{w}}{\partial \tau}+\frac{\partial \mathbf{q}}{\partial t}+\frac{\partial \mathbf{f}}{\partial x}+\frac{\partial \mathbf{g}}{\partial y}=\frac{\partial \mathbf{f_v}}{\partial x}+\frac{\partial \mathbf{g_v}}{\partial y}-(\rho_o-\rho)\mathbf{G}+\int \delta(\mathbf{x}-\mathbf{x_f})\sigma\kappa\mathbf{n}ds \quad (5)$$

with

$$\mathbf{w}=\begin{pmatrix}p\\ u\\ v\end{pmatrix},\quad \Gamma^{-1}=\begin{pmatrix}\frac{1}{\rho\beta^2} & 0 & 0\\ \frac{2\alpha u}{\beta^2} & \rho & 0\\ \frac{2\alpha v}{\beta^2} & 0 & \rho\end{pmatrix} \quad (6)$$

where β is a preconditioning parameter with dimensions of velocity and α is a dimensionless parameter to be determined. In Eq. (5), τ denotes the pseudotime and the dual time-stepping method is based on marching in pseudotime until a convergence is reached for each physical time step. Since the transient solution in pseudotime is not of interest, we are free to use any nonphysical convergence acceleration technique, such as preconditioning, local time-stepping, and multigrid methods. To facilitate treatment of complex geometries, Eq. (5) is transformed into a curvilinear coordinates defined by

$$\xi=\xi(x,y),\quad \eta=\eta(x,y) \quad (7)$$

Using the relation $\mathbf{q}=\mathbf{I^1}\rho\mathbf{w}$ where the incomplete identity matrix $\mathbf{I}^1$ is defined as

$$\mathbf{I}^1=\begin{bmatrix}0 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} \quad (8)$$

and the transformation given by Eq. (7), the transformed equations in the curvilinear coordinates can be written as

$$\Gamma^{-1}\frac{\partial h\mathbf{w}}{\partial \tau}+\mathbf{I^1}\frac{\partial h\rho\mathbf{w}}{\partial t}+\frac{\partial h\mathbf{F}}{\partial \xi}+\frac{\partial h\mathbf{G}}{\partial \eta}=\frac{\partial h\mathbf{F_v}}{\partial \xi}+\frac{\partial h\mathbf{G}_\mathrm{v}}{\partial \eta}+h\mathbf{f}_b \quad (9)$$

where $h=x_\xi y_\eta-x_\eta y_\xi$ is the determinant of the Jacobian of the transformation and $h\mathbf{F}$, $h\mathbf{G}$, $h\mathbf{F_v}$, and $h\mathbf{G_v}$ are the transformed convective and viscous fluxes given by

$$\begin{aligned}h\mathbf{F}&=y_\eta\mathbf{f}-x_\eta\mathbf{g},\quad h\mathbf{F_v}=y_\eta\mathbf{f_v}-x_\eta\mathbf{g_v},\\ h\mathbf{G}&=-y_\xi\mathbf{f}+x_\xi\mathbf{g},\quad h\mathbf{G_v}=-y_\xi\mathbf{f_v}+x_\xi\mathbf{g_v}\end{aligned} \quad (10)$$

The vector $\mathbf{f}_b$ represents the last two terms on the right-hand side of Eq. (1), namely, the sum of the buoyancy forces and the surface tension. Following Caughey [9], subiterated implicit scheme to solve Eq. (9) can be written as

$$\begin{aligned}\Gamma^{-1}h\frac{\mathbf{w}^{p+1}-\mathbf{w}^p}{\Delta\tau}+\mathbf{I}^1\frac{3(\rho h\mathbf{w})^{p+1}-4(\rho h\mathbf{w})^n+(\rho h\mathbf{w})^{n-1}}{2\Delta t}\\ =\theta\left[\frac{\partial h\mathbf{F_v}}{\partial \xi}+\frac{\partial h\mathbf{G_v}}{\partial \eta}+h\mathbf{f_b}\right]^p-\theta\left[\frac{\partial h\mathbf{F}}{\partial \xi}+\frac{\partial h\mathbf{G}}{\partial \eta}\right]^{p+1}-(1-\theta)\\ \times\left[\frac{\partial h(\mathbf{F}-\mathbf{F_v})}{\partial \xi}+\frac{\partial h(\mathbf{G}-\mathbf{G_v})}{\partial \eta}-h\mathbf{f_b}\right]^n\end{aligned} \quad (11)$$

where $(\,)^p$ denotes the pth level of the subiteration and $(\,)^n$ denotes the nth level of the physical time step. The iterations in the physical and pseudotimes are called the outer and inner iterations, respectively. The parameter θ is the implicitness factor with $\theta=1$, corresponding to a fully implicit method in pseudotime. As can be seen in Eq. (11), the viscous and source terms are treated implicitly in the physical time and explicitly in the pseudotime. The correction $\Delta\mathbf{w}=\mathbf{w}^{p+1}-\mathbf{w}^p$ is computed in each subiteration according to

$$h\left(\frac{\Gamma^{-1}}{\Delta\tau}+\frac{3\rho}{2\Delta t}I^{1}\right)\Delta\mathbf{w}^{p}=-\mathbf{I^{1}}\left[\frac{3(\rho h\mathbf{w})^{p}-4(\rho h\mathbf{w})^{n}+(\rho h\mathbf{w})^{n-1}}{2\Delta t}\right]$$
$$+\theta\left[\frac{\partial h\mathbf{F_{v}}}{\partial\xi}+\frac{\partial h\mathbf{G_{v}}}{\partial\eta}+h\mathbf{f_{b}}\right]^{p}-\theta\left[\frac{\partial h\mathbf{F}}{\partial\xi}\right.$$
$$\left.+\frac{\partial h\mathbf{G}}{\partial\eta}\right]^{p+1}-(1-\theta)\left[\frac{\partial h(\mathbf{F}-\mathbf{F_{v}})}{\partial\xi}\right.$$
$$\left.+\frac{\partial h(\mathbf{G}-\mathbf{G_{v}})}{\partial\eta}-h\mathbf{f_{b}}\right]^{n}\tag{12}$$

As can be seen in Eq. (12), when a steady state is reached in the pseudotime, i.e., $\Delta\mathbf{w}^{p}=0$, we have $(\ )^{p}\rightarrow(\ )^{n+1}$. Therefore the method is equivalent to the second-order implicit backward Euler method in the physical time. To solve Eq. (12), the convective fluxes are linearized in pseudotime according to

$$\begin{aligned}\left(\frac{\partial h\mathbf{F}}{\partial\xi}\right)^{p+1}&=\left(\frac{\partial h\mathbf{F}}{\partial\xi}\right)^{p}+A\,\frac{\partial\Delta\mathbf{w}^{p}}{\partial\xi}+O(\Delta\tau^{2})\\ \left(\frac{\partial h\mathbf{G}}{\partial\eta}\right)^{p+1}&=\left(\frac{\partial h\mathbf{G}}{\partial\eta}\right)^{p}+B\,\frac{\partial\Delta\mathbf{w}^{p}}{\partial\eta}+O(\Delta\tau^{2})\end{aligned}\tag{13}$$

where the Jacobian matrices are defined as

$$\mathbf{A}=\left(\frac{\partial\mathbf{F}}{\partial\mathbf{w}}\right)^{p},\quad \mathbf{B}=\left(\frac{\partial\mathbf{G}}{\partial\mathbf{w}}\right)^{p}\tag{14}$$

From Eqs. (12)–(14), the linearized equations can be written as

$$\left[\mathbf{I}+\frac{\theta\Delta\tau}{h}\left(\tilde{\mathbf{A}}\frac{\partial}{\partial\xi}+\tilde{\mathbf{B}}\frac{\partial}{\eta}\right)\right]\Delta\mathbf{w}^{p}=\mathbf{R}\tag{15}$$

where the residual vector is defined as

$$\mathbf{R}=-\mathbf{D^{-1}I^{1}}\frac{\Delta\tau}{\rho h}\left[\frac{3(\rho h\mathbf{w})^{p}-4(\rho h\mathbf{w})^{n}+(\rho h\mathbf{w})^{n-1}}{2\Delta t}\right]$$
$$-\frac{\Delta\tau\mathbf{D^{-1}}\Gamma}{h}\left\{\theta\left[\frac{\partial h(\mathbf{F}-\mathbf{F_{v}})}{\partial\xi}+\frac{\partial h(\mathbf{G}-\mathbf{G_{v}})}{\partial\eta}-h\mathbf{f_{b}}\right]^{p}+(1-\theta)\right.$$
$$\left.\times\left[\frac{\partial h(\mathbf{F}-\mathbf{F}_{\mathrm{v}})}{\partial\xi}+\frac{\partial h(\mathbf{G}-\mathbf{G_{v}})}{\partial\eta}-h\mathbf{f_{b}}\right]^{n}\right\}\tag{16}$$

and

$$\tilde{\mathbf{A}}=\mathbf{D^{-1}}\Gamma\mathbf{A};\quad \tilde{\mathbf{B}}=\mathbf{D^{-1}\Gamma B};\quad \mathbf{D}=\mathbf{I}+\frac{3\Delta\tau}{2\mathbf{\Delta}\mathrm{t}}\mathbf{I^{1}}\tag{17}$$

Following Caughey [9], Eq. (15) is factorized as

$$\left[I+\frac{\theta\Delta\tau}{h}\tilde{\mathbf{A}}\frac{\partial}{\partial\xi}\right]\left[I+\frac{\theta\Delta\tau}{h}\tilde{\mathbf{B}}\frac{\partial}{\eta}\right]\Delta\mathbf{w}^{p}=\mathbf{R}\tag{18}$$

which can be solved efficiently in two steps by using a block tridiagonal solver. However, Eq. (18) can be solved more efficiently using the diagonalization procedure. The diagonalization is possible because the inviscid part of the preconditioned equations are hyperbolic, so there exist modal matrices $\mathbf{Q}_{\tilde{\mathbf{A}}}$ and $\mathbf{Q}_{\tilde{\mathbf{B}}}$ such that

$$\mathbf{\Lambda}_{\tilde{\mathbf{A}}}=\mathbf{Q}_{\tilde{\mathbf{A}}}^{-1}\tilde{\mathbf{A}}\mathbf{Q}_{\tilde{\mathbf{A}}};\quad \mathbf{\Lambda}_{\tilde{\mathbf{B}}}=\mathbf{Q}_{\tilde{\mathbf{B}}}^{-1}\tilde{\mathbf{B}}\mathbf{Q}_{\tilde{\mathbf{B}}}\tag{19}$$

and the diagonal matrices having real eigenvalues. The diogonalized algorithm is then given by

$$\left[I+\frac{\theta\Delta\tau}{h}\mathbf{\Lambda}_{\tilde{\mathbf{A}}}\frac{\partial}{\partial\xi}\right]\mathbf{Q}_{\tilde{\mathbf{A}}}^{-1}\mathbf{Q}_{\tilde{\mathbf{B}}}\left[I+\frac{\theta\Delta\tau}{h}\mathbf{\Lambda}_{\tilde{\mathbf{B}}}\frac{\partial}{\eta}\right]\mathbf{\Delta V}^{p}=\mathbf{Q}_{\tilde{\mathbf{A}}}^{-1}\mathbf{R}\tag{20}$$

where $\mathbf{\Delta V}^{p}=\mathbf{Q}_{\tilde{\mathbf{B}}}^{-1}\Delta\mathbf{w}^{p}$. Equation (20) is solved in two steps using a scalar tridiagonal solver in each step. Note that the spatial derivatives are approximated by a cell-centered finite volume method, which is equivalent to second-order central differences on a regular Cartesian grid and fourth-order numerical dissipation terms similar to that of Caughey [9] are added explicitly to the right-hand side of Eq. (18) to prevent the odd-even decoupling.

A front-tracking method similar to that of Unverdi and Tryggvason [8] is developed for treatment of the different phases and the surface tension. In this method, the interface is divided into small line segments called front elements, and the end points of each element are tracked explicitly in a Lagrangian frame. The details of the numerical method can be found in Muradoglu and Kayaalp [10]. The complete solution procedure can be summarized as follows:

In advancing solutions from physical time level n $(t_{n}=n\cdot\Delta t)$ to level $n+1$, the locations of front points at the new time level $n+1$ are first predicted using an explicit Euler method, i.e.,

$$\mathbf{X_{f}}^{n+1}=\mathbf{X_{f}}^{n}+\Delta t\mathbf{V_{f}}^{n}\tag{21}$$

where $\mathbf{X_{f}}$ and $\mathbf{V_{f}}$ denote the position of front points and the flow velocity interpolated from the neighboring fixed grid points onto front point $\mathbf{X_{f}}$, respectively. Then the material properties and surface tension are evaluated using the predicted front position $\mathbf{X_{f}}$, i.e.,

$$\rho^{n+1}=\rho(\mathbf{X_{f}}^{n+1});\quad \mu^{n+1}=\mu(\mathbf{X_{f}}^{n+1});\quad \mathbf{f_{b}}^{n+1}=\mathbf{f_{b}}(\mathbf{X_{f}}^{n+1})\tag{22}$$

The velocity and pressure fields at new physical time level $(n+1)$ are then computed by solving the flow equations by the FV method for a single physical time step, and finally the positions of the front points are corrected as

$$\mathbf{X_{f}}^{n+1}=\mathbf{X_{f}}^{n}+\frac{\Delta t}{2}(\mathbf{V_{f}}^{n}+\mathbf{V_{f}}^{n+1})\tag{23}$$

After this step, the material properties and the body forces are also reevaluated using the corrected front position. Cubic B-splines are used for all the interpolations from the fixed curvilinear grid onto front points and from the front points onto fixed curvilinear grid, and for distributing surface tension onto fixed curvilinear grid. The overall method is second-order accurate both in time and space. It is emphasized that the method is implicit in physical time and the physical time step Δt is solely determined by accuracy considerations.

An auxiliary regular Cartesian grid is utilized for tracking the positions of the front points in the curvilinear grid and is found to be very robust and efficient. Details of the tracking algorithm can also be found in Muradoglu and Kayaalp [10]. The auxiliary regular Cartesian grid is also used to determine the material properties using the procedure developed by Unverdi and Tryggvason [8], which involves solution of a Poisson equation. Bilinear interpolations are used to interpolate the material properties from the regular grid onto the curvilinear grid. The front marker points are reflected back into the computational domain in the case that the points cross the solid boundary due to numerical errors. The surface tension is distributed only onto the neighboring grid points when the front is close to the solid boundary and care is taken to make sure that the distributed forces are equivalent to the surface tension. However, no special treatment is done when two fronts come close to each other in the computational domain.

We note that, in addition to the preconditioning method, a multigrid method similar to that of Caughey [9] and a local time-stepping method are used to further accelerate convergence rate in pseudotime stepping. The details of the FV method can be found in Muradoglu and Kayaalp [10].

4 Results and Discussion

The method is first validated for the test case of a buoyancy-driven drop rising in a straight channel, and the results obtained with the present method and with the well-tested FD/FT method of Unverdi and Tryggvason [8] implemented in the FTC2D code are compared. Then the method is applied to more challenging test cases of the buoyancy-driven drops in various constricted channels. Although the method is general and can handle many drop

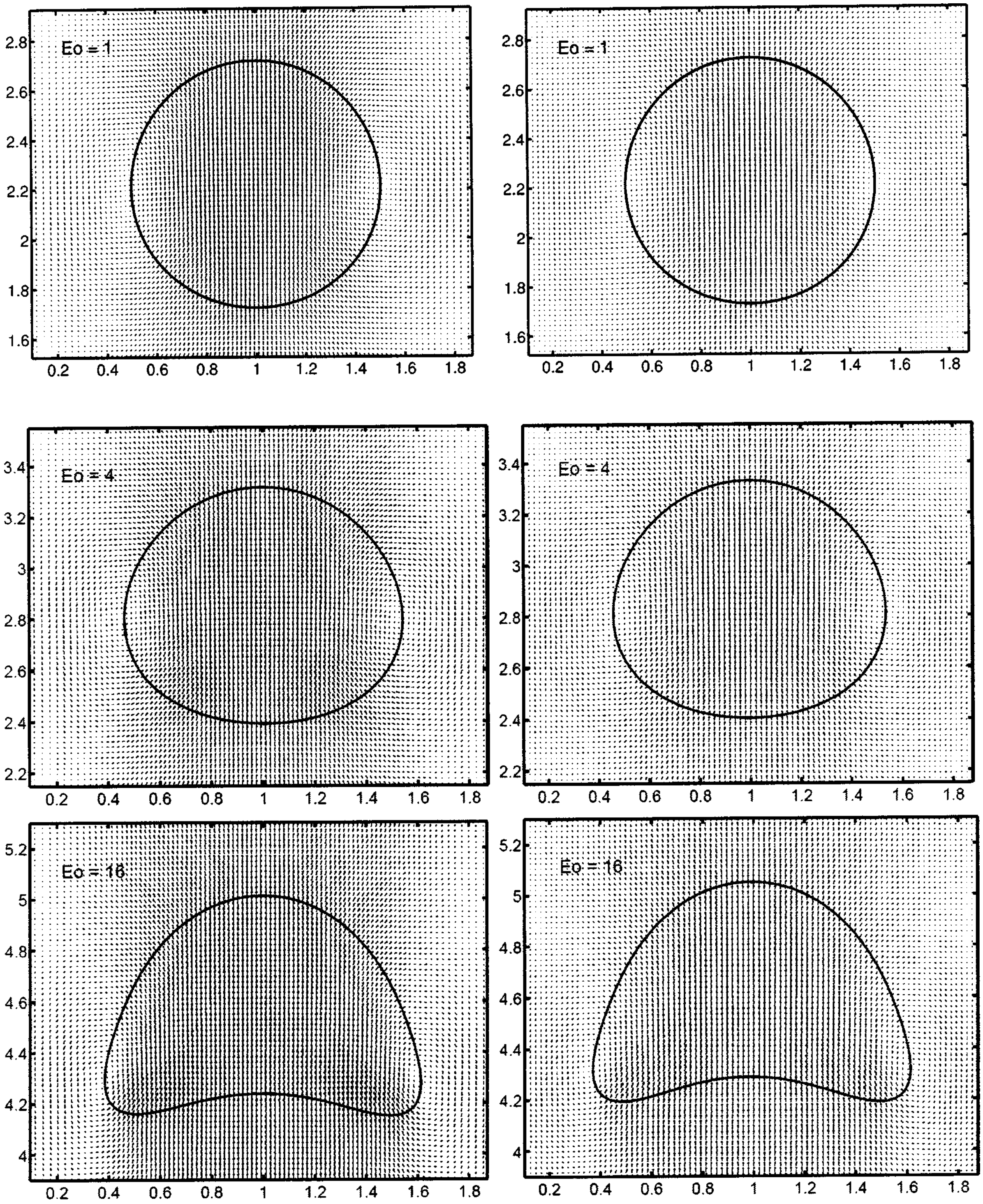


Fig. 1 Velocity vectors around a light drop rising in a straight channel for Eötvös and Morton numbers $Eo=1$, $M=10^{-4}$ (top plots), $Eo=4$, $M=4\times10^{-4}$ (middle plots) and $Eo=16$, $M=16\times10^{-4}$ (bottom plots) at $t^*=9.487$. Present results (left plots) are compared with the FTC2D results (right plot). Grid: 96×384, $dt^*=0.0316$.

interactions, here only a single drop and two drop cases are considered. The computational results are expressed in terms of nondimensional quantities. For this purpose, the length, time, and velocity scales are defined as $L=d_e$, $T=\sqrt{d_e/G}$ and $V_r=\mu_o/\rho_o d_e$, respectively, and the nondimensional quantities are denoted by*. For example, the *x* and *y* coordinates are nondimensionalized as $x^*=x/L$ and $y^*=y/L$, respectively. Although a three orders-of-magnitude reduction in the *rms* residuals of the subiterations is found to produce essentially the same results; the *rms* residuals are reduced by four orders of magnitude in each inner iteration in pseudotime for all the simulations presented in this paper.

4.1 Freely Rising Drop in a Straight Channel. The method is first applied to a two-dimensional freely rising drop in a straight channel. The purpose of this test case is to validate the

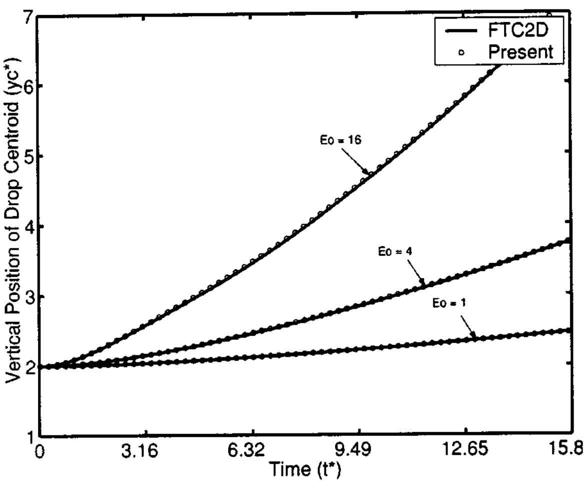


Fig. 2 The vertical positions (left plot) and the rise velocities (right plot) of the drop centroid taken from the simulations of the light drop rising in a straight channel. Computations are performed for Eötvös numbers 1, 4, and 16. The solid lines denote the FTC2D results and the symbols are the present calculations. Grid: 96×384, $dt^*=0.0316$.

method against the FD/FT method implemented in FTC2D code of Unverdi and Tryggvason [8] that can only use regular Cartesian grids. The computational domain is $2d_e\times 8d_e$, where d_e is the initial drop diameter and is resolved by a 96×384 uniform regular Cartesian grid. No-slip boundary conditions are applied on the side walls (i.e., $x=0$ and $x=2d_e$), while periodic boundary conditions are used in the vertical direction. The drop is initially an infinitely long circular cylinder centered at $(x_c^*,y_c^*)=(1.0,2.0)$ and starts rising from the rest due to buoyancy forces at time $t^*=0$. Freely rising drops take various shapes depending essentially on the Eötvös number.

To show this effect, computations are performed for three different Eötvös and Morton numbers (i.e., $Eo=1$, $M=10^{-4}$; $Eo=4$, $M=4\times 10^{-4}$; and $Eo=16$, $M=16\times 10^{-4}$), while the viscosity ratio is kept constant at $\zeta=1$. The corresponding density ratios are $\gamma=0.975$, 0.9, and 0.6, respectively. The physical time step is fixed at $dt^*=0.0316$, and the residuals are reduced by four orders of magnitude in each inner iteration in pseudotime. The drop shapes and the velocity vector field in the vicinity of the drop are plotted in Fig. 1 at time $t^*=9.487$ for the cases of $Eo=1$ (top plots), $Eo=4$ (middle plots), and $Eo=16$ (bottom plots). The results obtained with the FTC2D code are also shown in the right plots of Fig. 1. It can be seen in this figure that the present results are in good agreement with the FTC2D results demonstrating the accuracy of the present method. It is also observed that drop deformation increases as the Eötvös number increases as expected. To better quantify the accuracy of the present method the vertical position of the drop centroid and the drop rise velocity computed with the present method as well as with the FTC2D code are plotted as a function of time in Fig. 2. As can be seen in this figure, the present results are overall in very good agreement with the results of the FTC2D code except for the small discrepancies observed between the two results for $Eo=16$, which is partly attributed to the time-stepping error in the present results. Note that the time step used in the present method is about 20 times larger than that used in the FTC2D code for the case of $Eo=16$. Although the flow is incompressible, drop volume (area) changes due to numerical errors and the percentage change in the drop volume is a good indicator for the accuracy of the method. Figure 3 shows the percentage change of the drop area obtained with the present method and the FTC2D code for all three sets of dimensionless numbers. It can be observed in this figure that, in contrast with the FTC2D results, the drop volume reduces in time as the drop rises in the present method for this test case, but the overall percentage change in the drop volume is comparable in magnitude in both methods. In summary the present results compare well with the results of the well-tested FD/FT method of Unverdi and Tryggvason [8] demonstrating the accuracy of the present algorithm.

4.2 Freely Rising Drops in Various Constricted Channels.

After validating the method for the case of a freely rising drop in a straight channel, we now consider buoyancy-driven drops rising in various constricted channels to show the ability of the method for treating dispersed multiphase flows in complex geometries where the phases strongly interact with the solid boundaries. The first test case concerns a single initially cylindrical drop rising due to buoyancy in a constricted channel. The channel is $2d_e$ wide, extends to $12d_e$ in the y direction, and is constricted sinusoidally at the middle by 75% as shown in Fig. 4(b). In Fig. 4(a), a portion of a coarse grid containing 32×192 grid cells is plotted in the vicinity of the constriction to show the overall structure of the body-fitted curvilinear grid used in the simulations. The governing nondimensional numbers are set to $Eo=8$, $M=8\times 10^{-4}$, $\gamma=0.8$, and $\zeta=1$. No-slip boundary conditions are applied on the solid

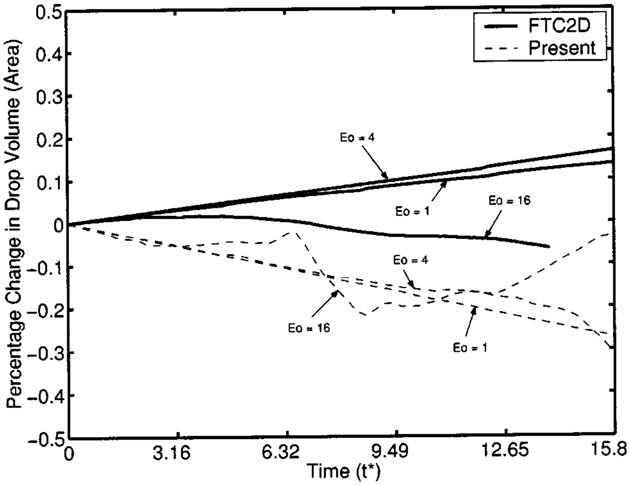


Fig. 3 Percentage change in the drop area for Eötvös numbers 1, 4, and 16 in the computations of the freely rising drop in the straight channel. Dashed curves denote the present results, and the solid curves are the FTC2D results. Grid: 96×384, $dt^*=0.0316$.

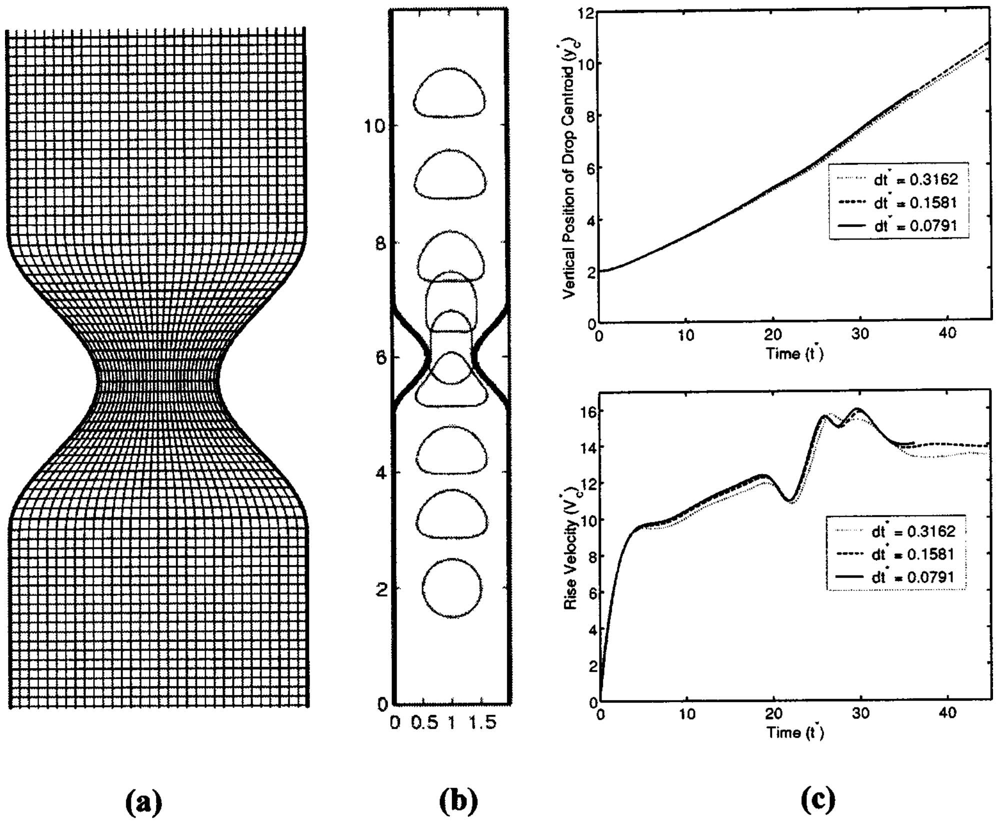


Fig. 4 Freely rising drop in a sinusoidally constricted channel. (*a*) A portion of the body-fitted curvilinear coarse grid containing 32×192 grid cells. (*b*) Snapshots taken at time frames $t^*=0$, 9.49, 15.81, 22.14, 28.46, 31.62, 37.95, and 44.27. Time progresses from bottom to top. (*c*) The vertical position (top plot) and the rise velocity (bottom plot) of the drop centroid computed with the physical time steps $dt^*=0.3162$ (dotted line), $dt^*=0.1581$ (dashed line), and $dt^*=0.0791$ (solid line). $Eo=2$, $M=8\times10^{-4}$, $\gamma=0.8$, $\zeta=1$. Grid: 128×768.

walls, and periodic boundary conditions are employed in the vertical direction. The drop is initially centered at $(x_c^*, y_c^*) = (1.0, 2.0)$ and starts rising from the rest due to buoyancy forces. The snapshots taken at the time frames $t^*=0$, 9.49, 15.81, 22.14, 28.46, 31.62, 37.95, and 44.27 are plotted in Fig. 4(*b*) to show the overall behavior of the drop. The computations are performed on a 128×768 grid. In order to better quantify the drop motion, the vertical position and the rise velocity of the drop centroid are plotted in Fig. 4(*c*). Computations are repeated for three different physical time steps (i.e., $dt^*=0.3162$, 0.1581, and 0.0791) on the same 128×768 grid to demonstrate the time-stepping error convergence. The small differences between the results computed

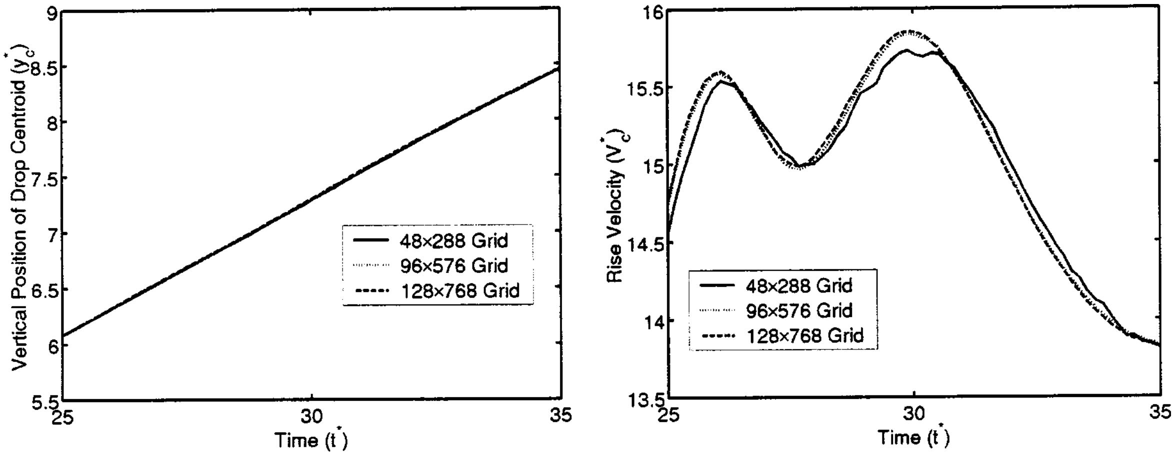


Fig. 5 Grid convergence analysis for the freely rising drop in the straight channel. The vertical position (left plot) and the rise velocity (right plot) of the drop centroid as a function of time computed on the body-fitted curvilinear grids containing 48×288 (solid line), 96×576 (dotted line) and 128×768 (dashed line) grid cells in the time interval $t^*=25$ and $t^*=35$. $dt^*=0.1581$, $Eo=4$.

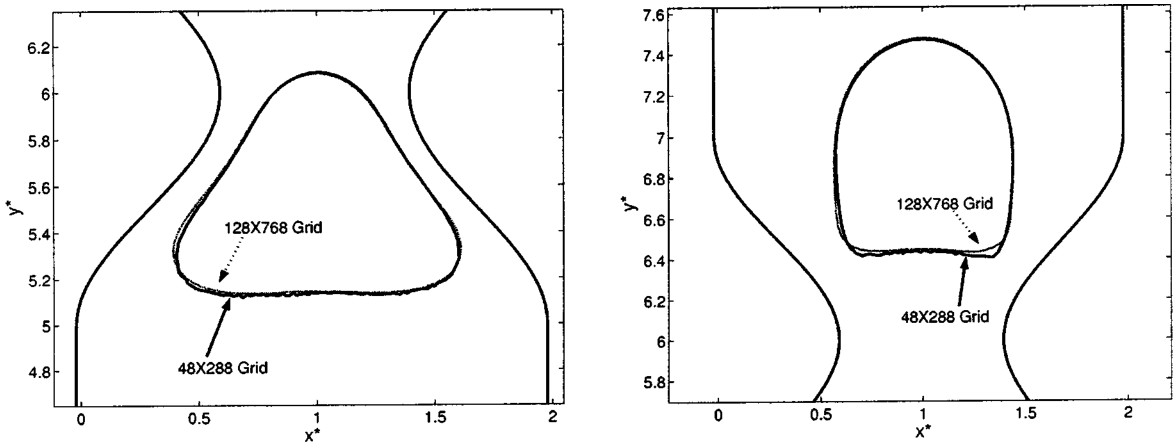
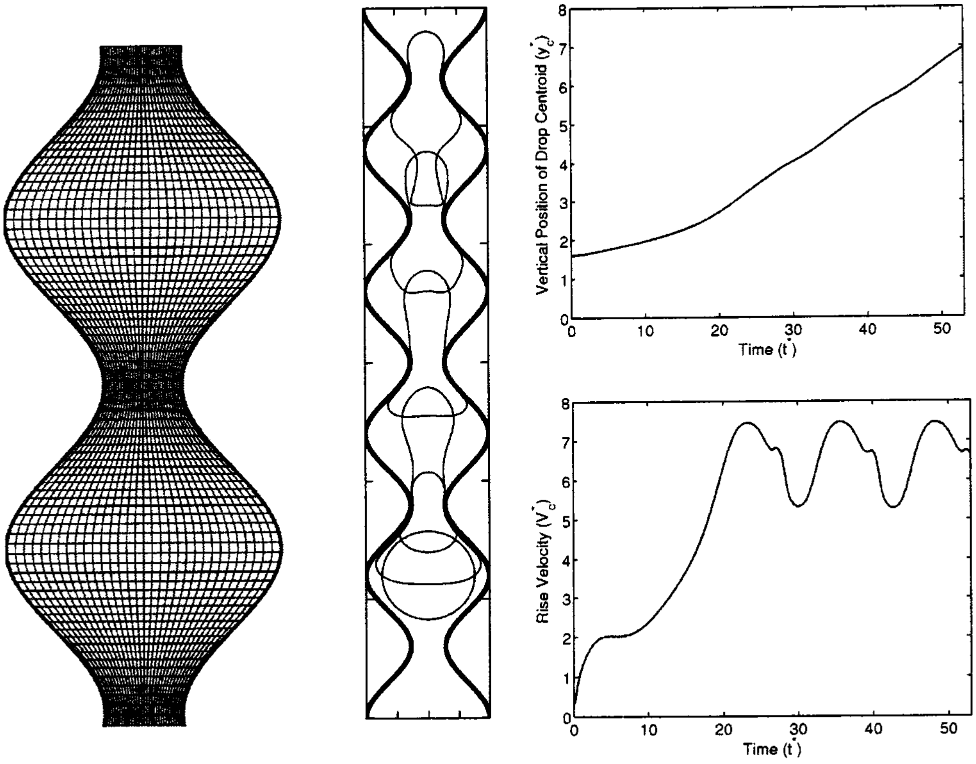


Fig. 6 Effects of grid refinement of the front structure. Before the drop enters (left plot, time $t^*=12.65$) and after it passes (right plot, time $t^*=22.14$) the constriction computed on 48×288 (solid line) and 128×768 (dashed line) grids. $dt^*=0.1581$, $Eo=4$. The coarse grid results in wiggles on the front while the front remains smooth in the case of the fine grid.

with the two smallest physical time steps indicate that the time-stepping error convergence is achieved and $dt^*=0.1581$ is sufficient for this test problem. Figures 4(*b*) and 4(*c*) together show that the drop motion initially resembles that of the straight-channel case before the drop starts feeling the effects of the constriction (i.e., before about $t^*=15$), but then it is strongly affected by the presence of the constriction as the drop passes through the constriction. In Fig. 5, the vertical position and the rise velocity of the drop centroid computed on 48×288, 96×576, and 128×768 grids are plotted to show the grid convergence. The physical time



(a) **(b)** **(c)**

Fig. 7 Freely rising drop in a continuously constricted channel. (*a*) A portion of the body-fitted curvilinear coarse grid containing 32×192 grid cells. (*b*) Snapshots taken at time frames $t^*=0$, 10.33, 20.66, 30.98, 41.31, and 51.64. Time progresses from bottom to top. (*c*) The vertical position (top plot) and the rise velocity (bottom plot) of the drop centroid computed with the physical time step $dt^*=0.1291$ on a 128×768 grid. $Eo=18$, $M=8\times10^{-4}$, $\gamma=0.8$, $\zeta=1$.

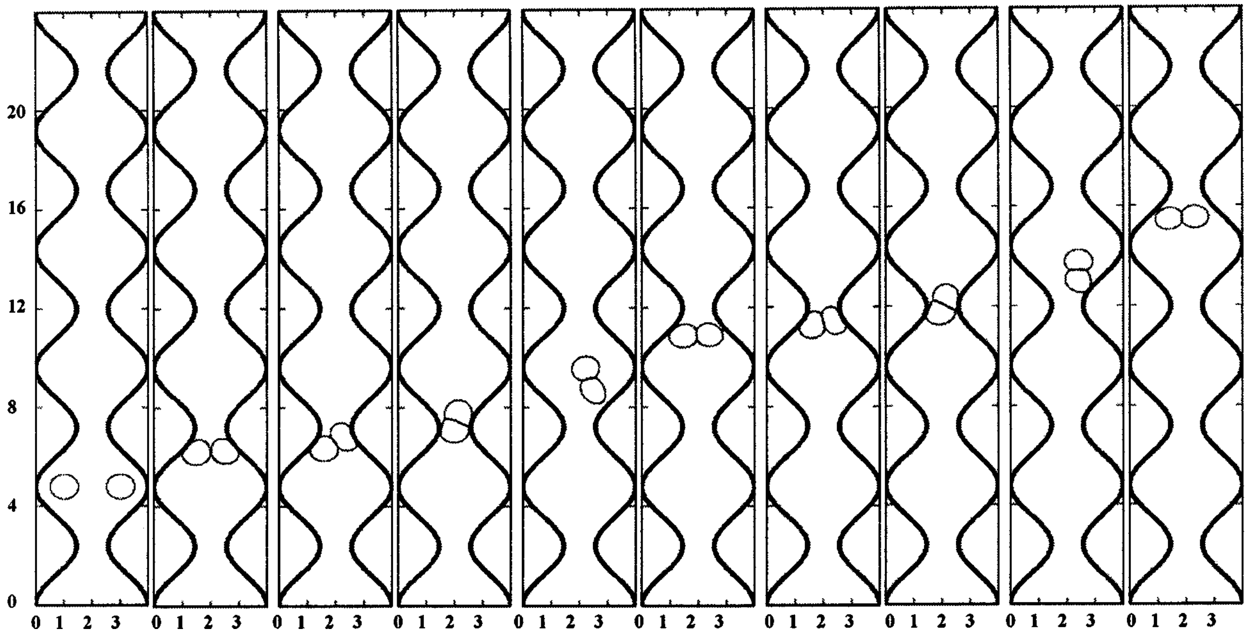


Fig. 8 Buoyancy-driven two-drop interaction in the continuously constricted channel. Snapshots taken at $t^*=0$, 35.78, 53.67, 71.55, 89.44, 107.33, 125.22, 143.11, 161.00, and 178.89. $Eo=2$, $M=8\times10^{-4}$, $\gamma=0.8$ and $\zeta=1$. Grid: 96×576, $dt^*=0.2236$.

step is kept constant at $dt^*=0.1581$, and the results are shown only in the time interval $t^*=25$ and $t^*=35$, during which the drop passes through the constriction. The decreasing differences between results obtained on the successively finer grids indicate that the grid convergence is achieved and the 96×576 grid is sufficient for this test case. Figure 6 shows the fronts computed on the 48×288 and 128×768 grids just before the drop enters and after it passes the constriction in order to demonstrate the effects of the grid refinement on the front structure. As can be seen in this figure, the coarse grid results in wiggles on the front while the front remains very smooth in the case of the fine grid. Note that results obtained on the 96×576 grid is not plotted because it is almost indistinguishable from the 128×768 grid case.

The next test case concerns a buoyancy-driven drop freely rising in a continuously constricted channel depicted in Fig. 7(*b*). The channel is $4d_e/3$ wide, and extends to $8d_e$ in the y direction, and is constricted along the y-axis by sinusoidal wavy walls. The constriction ratio is 75% as in the singly constricted channel case. The initial and the boundary conditions are the same as the singly constricted channel case, and the drop is initially centered at $(x_c^*,y_c^*)=(0.6667,1.6)$. The governing nondimensional numbers are $Eo=18$, $M=8\times10^{-4}$, $\gamma=0.8$, and $\zeta=1$. Computations are performed on a 96×576 grid with the constant physical time step $dt^*=0.1291$. A portion of a coarser grid containing 32×192 grid cells is plotted in Fig. 7(*a*). Note that a similar geometry was used by Hemmat and Borhan [2] in their experimental study of the buoyancy-driven drops and bubbles. The snapshots taken at the time frames $t^*=0.0$, 10.33, 20.66, 30.98, 41.31, and 51.64 are plotted in Fig. 7(*b*) to show the overall evolution of the drop motion. The strong interactions between the drop and the solid

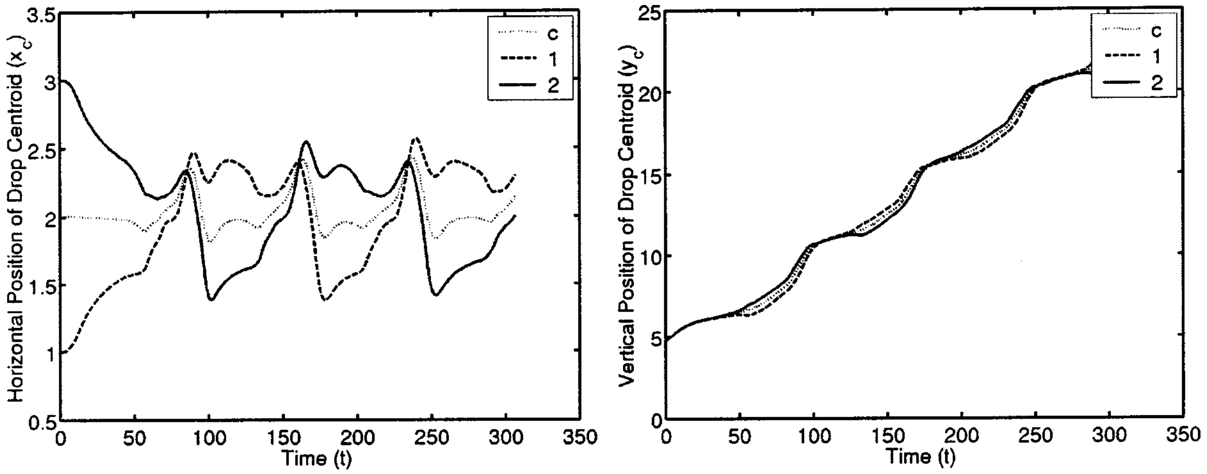


Fig. 9 Buoyancy-driven two-drop interaction in the continuously constricted channel. The horizontal (left plot) and the vertical (right plot) positions of the (initially) left drop centroid (solid line), the (initially) right drop (dashed line), and the center of the mass of the drop system (dotted line). $Eo=2$, $M=8\times10^{-4}$, $\gamma=0.8$, $\zeta=1$. Grid: 96×576, $dt^*=0.2236$.

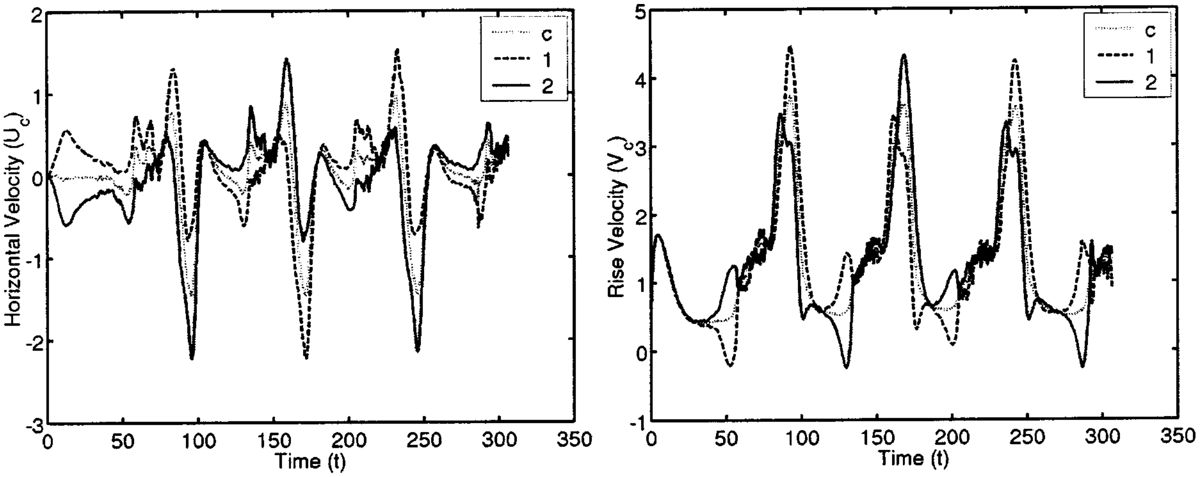


Fig. 10 Buoyancy-driven two-drop interaction in the continuously constricted channel. The horizontal (left plot) and the vertical (right plot) velocities of the (initially) left drop centroid (solid line), the (initially) right drop (dashed line), and the center of the mass of the drop system (dotted line). $Eo=2$, $M=8\times10^{-4}$, $\gamma=0.8$, $\zeta=1$. Grid: 96×576, $dt^*=0.2236$.

walls can be clearly seen in this figure. It is emphasized that, in spite of large deformations, the front remains smooth, which might be considered as a good indication for the accuracy of the simulation. The vertical position and the rise velocity of the drop centroid are plotted in Fig. 7(*c*). It can be seen in this figure that the rise velocity becomes periodic after a transient period (i.e., after about $t^*=20$).

Finally, the method is used to compute two identical drops freely rising in the continuously constricted channel. In this case, the drop diameters are relatively small compared to the channel width, and the ratio of the initial drop diameter to the maximum channel width is $d_e/d_{\max}=0.25$. The corresponding nondimensional numbers are $Eo=2$, $M=8\times10^{-4}$, $\gamma=0.8$, and $\zeta=1$. The same 96×576 grid is employed as used in the single-drop case and the physical time step is taken as $dt^*=0.2236$. The drops are initially located at $(x_c^*,y_c^*)=(1.0,4.8)$ and $(x_c^*,y_c^*)=(3.0,4.8)$. The snapshots taken at the time frames $t^*=0$, 35.78, 53.67, 71.55, 89.44, 107.33, 125.22, 143.11, 161.00, and 178.89 are plotted in Fig. 8 to show the overall behavior of the drops. The drops strongly interact with the solid walls as well as with themselves and deform considerably as they pass through the constrictions. It is interesting to observe that the drops initially rise side by side but the right drop passes the first constriction little earlier than the left drop and then the left drop catches up and passes the right drop. After that the drops switches their positions and this behavior is repeated periodically after each constriction. This behavior can also be seen in Figs. 9 and 10, where the horizontal and vertical locations, the horizontal and vertical (rise) velocities of the individual drop centroids, and the center of the mass of the drop system are plotted. Note that neither breakup nor coalescence is allowed in the present simulations. It can be seen from these figures that the motion of a two-drop system becomes periodic after a transient period. Due to relatively poor grid resolution and large physical time step, drops lose about 20% volume when they move about 14 drop diameters (which corresponds to $t^*=215$). In spite of very strong interactions between the drops and the solid walls as well as between drops themselves and a relatively coarse grid, the periodic behavior of the two-drop system observed in Figs. 9 and 10 indicates that the main features of the two-drop interactions are well captured by the present method.

5 Conclusions

The computations of buoyant drops in constricted channels have been reported in this paper. The FV/FT method is first validated for the case of a buoyancy-driven drop in a straight channel by comparing the results with the computations obtained by the well-tested FD/FT method implemented in FTC2D code [8]. The method has been successfully applied to buoyant drops in various constricted channels. Some error convergence studies have also been performed. It is found that the present method is a viable tool to model dispersed multiphase flows in arbitrarily complex geometries.

References

[1] Olbricht, W. L., and Leal, L. G., 1983, "The Creeping Motion of Immiscible Drops Through a Diverging/Converging Tube," J. Fluid Mech., **134**, 329–355.

[2] Hemmat, M., and Borhan, A., 1996, "Buoyancy-Driven Motion and Deformation of Drops and Bubbles in a Periodically Constricted Capillary," Chem. Eng. Commun., **148–150**, 363–384.

[3] Udaykumar, H. S., Kan, H.-C., Shyy, W., and Tran-Son-Tay, R., 1997, "Multiphase Dynamics in Arbitrary Geometries on Fixed Cartesian Grids," J. Comput. Phys., **137**, 366–405.

[4] Tryggvason, G., Bunner, B., Esmaeeli, A., Juric, D., Al-Rawahi, N., Tauber, W., Han, J., Nas, S., and Jan, Y.-J., 2001, "A Front Tracking Method for the Computations of Multiphase Flows," J. Comput. Phys., **169**, 708–759.

[5] Tsai, T. M., and Miksis, J. M., 1994, "Dynamics of a Drop in a Constricted Capillary Tube," J. Fluid Mech., **274**, 197–217.

[6] Manga, M., 1996, "Dynamics of Drops in Branched Tubes," J. Fluid Mech., **315**, 105–117.

[7] Olbricht, W. L., and Kung, D. M., 1992, "The Deformation and Breakup of Liquid Drops in Low Reynolds Number Flows Through Capillaries," Phys. Fluids, **4**, 1347–1354.

[8] Unverdi, S. O., and Tryggvason, G., 1992, "A Front-Tracking Method for Viscous, Incompressible Flows," J. Comput. Phys., **100**, 25–37.

[9] Caughey, D. A., 2001, "Unsteady Flows Past Cylinders of Square Cross-Section," Comput. Fluids, **30**, 939–960.

[10] Muradoglu, M., and Kayaalp, A. D., 2004, "A Finite Volume/Front-Tracking Method for Computations of Dispersed Multiphase Flows in Complex Geometries" J. Comput. Phys. (submitted for publication).

[11] Clift, R., Grace, J. R., and Weber, M. E., 1978, *Bubbles, Drops, and Particles*, Academic Press, San Diego, CA.

Xian-Fang Li
Institute of Mechanics and Sensor Technology,
School of Civil Engineering and Architecture,
Central South University, Changsha,
Hunan 410083, China
e-mail: xfli25@yahoo.com.cn

Kang Yong Lee[1]
School of Mechanical Engineering,
Yonsei University,
Seoul 120-749, South Korea
e-mail: KLY2813@yonsei.ac.kr

Three-Dimensional Electroelastic Analysis of a Piezoelectric Material With a Penny-Shaped Dielectric Crack

Previous studies assumed that a crack is either impermeable or permeable, which are actually two limiting cases of a dielectric crack. This paper considers the electroelastic problem of a three-dimensional transversely isotropic piezoelectric material with a penny-shaped dielectric crack perpendicular to the poling axis. Using electric boundary conditions controlled by the boundaries of an opening crack, the electric displacements at the crack surfaces are determined. The Hankel transform technique is employed to reduce the considered problem to dual integral equations. By solving resulting equations, the results are presented for the case of remote uniform loading, and explicit expressions for the electroelastic field at any point in the entire piezoelectric body are given in terms of elementary functions. Moreover, the distribution of asymptotic field around the crack front and field intensity factors are determined. Numerical results for a cracked PZT-5H ceramic are evaluated to examine the influence of the dielectric permittivity of the crack interior on the field intensity factors, indicating that the electric boundary conditions at the crack surfaces play an important role in determining electroelastic field induced by a crack, and that the results are overestimated for an impermeable crack, and underestimated for a permeable crack. [DOI: 10.1115/1.1795219]

1 Introduction

Piezoelectric materials have been used widely in techniques, such as actuators, sensors, transducers, etc., due to the intrinsic coupling feature between elastic and electric behaviors [1]. However, a main disadvantage is that they are very susceptible to fracture because of their brittleness. Owing to various causes, cracks or flaws are inevitably present in piezoelectric materials, which gives rise to electroelastic field concentration under applied electromechanical loading, rising high enough to cause the crack advance, and finally leads to serious degradation of the performance of piezoelectric materials. To understand the failure mechanism of piezoelectric materials and maintain the stability of cracked piezoelectric structures operating in an environment of combined electromechanical loading, the analysis of elastic and electric behaviors is prerequisite. So far, many efforts in theory have been made in this field (e.g., [1–13], among others).

In determining electroelastic field induced by cracks embedded in piezoelectric materials, cracks of two types prevail, which usually are referred to as impermeable and permeable cracks. In other words, for electric boundary conditions at the crack surfaces, it is assumed that the electric displacements at the crack surfaces and inside the crack vanishes for the former case [3,4], and that electric displacement and electric potential are continuous across the crack for latter case [8]. However, a real crack is a dielectric with permittivity $\varepsilon^{(c)}$, $\varepsilon^{(c)}=\varepsilon_0$, $\varepsilon_0=8.85\times10^{-12}$ F/m corresponding to a crack full of air or vacuum. An impermeable crack simply imposes the requirement of $\varepsilon^{(c)}=0$, and further exact analysis of a slender elliptical hole by Dunn [5], Sosa and Khutoryansky [7], and Zhang et al. [9] indicates that the assumption of the impermeable crack may give rise to significant errors in determining the electroelastic field. In contrast, for permeable crack having no thickness, since the electric boundary conditions are assumed to be fulfilled at the boundaries of an *undeformed* crack in solving electroelastic fields, it ignores the contribution of the dielectric of the opening crack interior under applied electromechanical loading on the electroelastic field. Therefore, in order to characterize the dependence of the crack opening displacement (COD) and the electric displacements at the crack surfaces and inside the opening crack, it is natural to assume that electric field inside the opening crack obeys

$$E_z^{(c)}=-\frac{\Delta\phi}{\Delta u_z} \tag{1}$$

where a quantity with the superscript c designates the one inside the opening crack, and $\Delta\phi$ and Δu_z are the voltage and COD across the crack, respectively. Consequently, the electric displacement vector $D^{(c)}$ inside the opening crack obeys

$$D_z^{(c)}=\varepsilon^{(c)}E_z^{(c)} \tag{2}$$

or

$$D_z^{(c)}=-\varepsilon^{(c)}\frac{\Delta\phi}{\Delta u_z} \tag{3}$$

which was first proposed by Hao and Shen [14], who used the above electric boundary condition together with solving a Riemann-Hilbert boundary value problem, and gave an electroelastic analysis of a crack in an infinite piezoelectric plane.

It is emphasized that (3) is fulfilled for a crack posterior to deformation, not prior to deformation, or Δu_z denotes the final displacement jump across the crack under combined electroelastic loading. In fact, finite element analysis given by McMeeking [15] demonstrated that it is more suitable and reasonable to consider the boundaries of a deformed crack, rather than an undeformed

---

[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, September 3, 2003; final revision, March 17, 2004. Associate Editor: H. Gao. 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.

crack. Moreover, the validity of (3) is observed from recent experiments by Schneider et al. [16], who found that there is a distinct drop of electric potential between the two surfaces of an opening crack, implying that the existence of electric field in the interior of an opening crack is full of dielectric such as air or vacuum. As a result, the condition (3) is adopted as an electric boundary condition, reflecting a geometrically nonlinear relation between electric displacement and COD, i.e., the electric displacement is controlled by unknown COD. Strictly speaking, the above electric boundary condition is not an exact electric boundary condition, since exact electric boundary conditions indicate that electric displacements and electric potentials (or fields) are continuous along the normal and tangential directions of the interface between a piezoelectric matrix and an inhomogeneity, respectively [17]. However, for an opening crack, due to the COD small compared to the crack length or radius. As a simple modelling, in what follows the electric boundary condition (3) is adopted.

For the two-dimensional case, such a treatment has been used in dealing with certain crack problems, such as Hao and Shen [14], Hao [18], Liu et al. [19], Xu and Rajapakse [20], Wang and Jiang [21], and Wang and Mai [22]. However, for three-dimensional crack problems in piezoelectric materials, similar treatment is lacking, although some field intensity factors have been determined for a penny-shaped crack. Such works can be found in [23–33], which are mainly concentrated on determining field intensity factors for either permeable or impermeable cracks via various approaches. For example, using the results of electroelastic behavior in an infinite piezoelectric medium containing a spheroidal piezoelectric inclusion and letting a spheroidal piezoelectric inclusion reduce to a penny-shaped crack, Kogan et al. [25] derived the results of field intensity factors for a permeable penny-shaped crack. A similar problem has also been studied under the electrically impermeable condition by Jiang and Sun [29], who found that the electric displacement intensity factor is independent of applied mechanical loading, which is completely opposite to that for a permeable crack [25]. This means that the dielectric of the crack interior (or equivalently the electric boundary conditions at the crack surfaces) plays a significant role in determining the electroelastic field in a piezoelectric material containing cracks. Nevertheless, there are few papers dealing with a dielectric crack in a three-dimensional piezoelectric space. Therefore, it is very necessary to investigate the effects of the dielectric of the crack interior on failure behavior of three-dimensional piezoelectric materials.

This paper is concerned with the electroelastic behavior analysis of a transversely isotropic piezoelectric material with a penny-shaped dielectric crack. The emphasis is focused on the influence of the dielectric permittivity of the crack interior on the electroelastic field, or the influence of the electric boundary condition (3) on the electroelastic field. The electric displacements at the crack surfaces are determined by applied electromechanical loading as well as the dielectric permittivity of the crack interior. By means of the Hankel transform technique, a complete solution for electroelastic behavior in the entire piezoelectric space is given in terms of elementary functions, and the corresponding field intensity factors are determined. The effects of the dielectric inside the crack on the fiend intensity factors are presented graphically for a cracked PZT-5H ceramic.

2 Statement of the Problem

Consider a class of axisymmetric problems of an infinite transversely isotropic piezoelectric body with the poling axis as the z axis and the isotropic plane as the xy plane. The constitutive equations with the framework of the theory of linear piezoelectricity take the form

$$\sigma_{rr}=c_{11}u_{r,r}+c_{12}\frac{u_r}{r}+c_{13}u_{z,z}+e_{31}\phi_{,z} \tag{4a}$$

$$\sigma_{\theta\theta}=c_{12}u_{r,r}+c_{11}\frac{u_r}{r}+c_{13}u_{z,z}+e_{31}\phi_{,z} \tag{4b}$$

$$\sigma_{zz}=c_{13}u_{r,r}+c_{13}\frac{u_r}{r}+c_{33}u_{z,z}+e_{33}\phi_{,z} \tag{4c}$$

$$\sigma_{rz}=c_{44}(u_{z,r}+u_{r,z})+e_{15}\phi_{,r} \tag{4d}$$

$$D_r=e_{15}(u_{z,r}+u_{r,z})-\varepsilon_{11}\phi_{,r} \tag{4e}$$

$$D_z=e_{31}\left(u_{r,r}+\frac{u_r}{r}\right)+e_{33}u_{z,z}-\varepsilon_{33}\phi_{,z} \tag{4f}$$

where c_{ij}, ε_{ij}, and e_{ij} are the elastic stiffnesses, the dielectric permittivities, the piezoelectric constants, respectively.

In the absence of body forces and free charges, the equilibrium equations of stresses and electric displacements are, respectively,

$$\sigma_{rr,r}+\sigma_{rz,z}+\frac{\sigma_{rr}-\sigma_{\theta\theta}}{r}=0 \tag{5a}$$

$$\sigma_{rz,r}+\sigma_{zz,z}+\frac{\sigma_{rz}}{r}=0 \tag{5b}$$

$$D_{r,r}+D_{z,z}+\frac{D_r}{r}=0 \tag{5c}$$

Substituting the constitutive equations into the above equilibrium equations yields the basic governing equations for elastic displacements u_r and u_z, and electric potential ϕ as follows:

$$c_{11}\left(u_{r,rr}+\frac{u_{r,r}}{r}-\frac{u_r}{r^2}\right)+c_{44}u_{r,zz}+(c_{13}+c_{44})u_{z,rz}+(e_{31}+e_{15})\phi_{,rz}=0 \tag{6a}$$

$$c_{44}\left(u_{z,rr}+\frac{u_{z,r}}{r}\right)+c_{33}u_{z,zz}+(c_{13}+c_{44})\left(u_{r,rz}+\frac{u_{r,z}}{r}\right)+e_{15}\left(\phi_{,rr}+\frac{\phi_{,r}}{r}\right)+e_{33}\phi_{,zz}=0 \tag{6b}$$

$$e_{15}\left(u_{z,rr}+\frac{u_{z,r}}{r}\right)+e_{33}u_{z,zz}+(e_{31}+e_{15})\left(u_{r,rz}+\frac{u_{r,z}}{r}\right)-\varepsilon_{11}\left(\phi_{,rr}+\frac{\phi_{,r}}{r}\right)-\varepsilon_{33}\phi_{,zz}=0 \tag{6c}$$

In what follows special attention is focussed on a penny-shaped crack of radius a perpendicular to the poling axis, as shown in Fig. 1. In order to obtain a desired electroelastic field of a piezoelectric body weakened by a penny-shaped crack, appropriate boundary conditions must be furnished. First of all, from the above analysis the electric boundary condition at the surfaces of a dielectric crack is

$$D_z^{(c)}\Delta u_z(r,0)=-\varepsilon^{(c)}\Delta\phi(r,0),\quad r<a \tag{7}$$

It is worth noting that the above electric boundary condition at the crack surface is equivalent to (3) for an opening crack $[\Delta u_z(r,0)>0]$. However, in case of $\Delta u_z(r,0)=0$, (3) becomes meaningless. As a matter of fact, in this case two crack surfaces keep contact with each other, and the electric potential jump

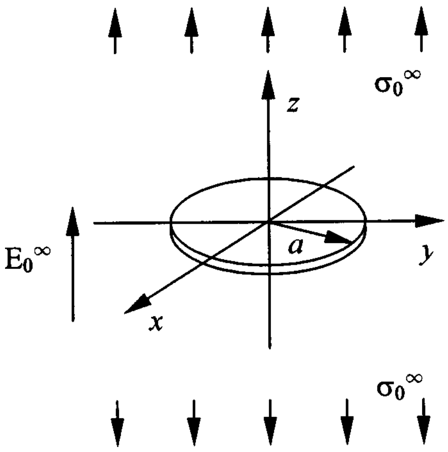


Fig. 1 Geometry of a piezoelectric material with a penny-shaped crack along with the corresponding coordinates

across the crack should vanish. Therefore, it is more convenient to rewrite (3) in the above form, which pertains to an opening crack as well as a closed crack. Previous assumptions on an electrically permeable crack are that electric potentials and electric displacements are continuous across an undeformed crack, rather than a deformed crack. In fact, once a crack opens, the electric potential jumps across the crack should exist. Therefore, electric boundary condition (7) is more suitable than the electrically impermeable and permeable assumptions, and moreover, the latter two cases are two limiting cases of the assumption (7). Very recently, Ou and Chen [34] also gave a discussion of validity of the condition (7).

For a penny-shaped crack embedded in an infinite piezoelectric body subjected to a uniform mechanical tension σ_0^∞ and constant electric field E_0^∞ at infinity along the poling axis, remote mechanical and electric boundary conditions can be written, respectively, below

$$\sigma_{zz}(r,z)\to\sigma_0^\infty, \quad z\to\infty \tag{8a}$$

$$\sigma_{rr}(r,z)\to 0, \quad r\to\infty \tag{8b}$$

$$\sigma_{rz}(r,z)\to 0, \quad z\to\infty \tag{8c}$$

$$E_z(r,z)\to E_0^\infty, \quad z\to\infty \tag{8d}$$

In general, it is easy to measure and control the electric field strength, rather than the electric displacement, in experiment. Hence, for electric boundary conditions, E_0^∞ is supposed to be prescribed. Of course, the case when electric displacement is given at infinity can be solved in a similar manner, which is omitted here.

3 Solution of the Problem

In this section, we consider a penny-shaped crack lying at a plane perpendicular to the poling direction. Obviously, due to the symmetry it is sufficient to consider the upper half piezoelectric body. The electroelastic field in the lower part can be directly given by symmetry from the counterpart in the upper part. Hence in what follows we confine our attention to the upper half-space.

In order to obtain a desired solution, it is expedient to employ the Hankel transform technique. We introduce three generalized harmonic functions by Hankel transform of the zeroth order

$$F_j(r,z) = -\int_0^\infty \frac{1}{\xi} A_j(\xi) e^{-\gamma_j \xi z} J_0(\xi r)\, \mathrm{d}\xi \tag{9}$$

for $z \geqslant 0$, where $A_j(\xi)$'s are unknown functions to be determined through appropriate electric and elastic boundary conditions, γ_j^2 $(j=1,2,3)$ are three distinct roots of the following characteristic equation:

$$a_0\gamma^6 + b_0\gamma^4 + c_0\gamma^2 + d_0 = 0 \tag{10}$$

where a_0, b_0, c_0, and d_0 are given in the Appendix, and γ_j is chosen such that $\mathrm{Re}(\gamma_j)$ is larger than zero to guarantee that the integral in (9) vanishes as z approaches infinity. Then a formal solution can be obtained in terms of $A_j(\xi)$ as follows:

$$u_r(r,z) = \sum_{j=1}^{3} \int_0^\infty A_j(\xi) e^{-\gamma_j \xi z} J_1(\xi r)\, \mathrm{d}\xi + B_1 r \tag{11a}$$

$$u_z(r,z) = \sum_{j=1}^{3} \eta_{3j}\gamma_j \int_0^\infty A_j(\xi) e^{-\gamma_j \xi z} J_0(\xi r)\, \mathrm{d}\xi + B_3 z \tag{11b}$$

$$\phi(r,z) = \sum_{j=1}^{3} \eta_{4j}\gamma_j \int_0^\infty A_j(\xi) e^{-\gamma_j \xi z} J_0(\xi r)\, \mathrm{d}\xi + B_4 z \tag{11c}$$

where B_k $(k=1,3,4)$ are unknown constants, and η_{3j} and η_{4j} are constants that can be determined by the following relations:

$$\frac{c_{11}}{c_{44}+(c_{13}+c_{44})\eta_{3j}+(e_{31}+e_{15})\eta_{4j}} = \frac{c_{13}+c_{44}+c_{44}\eta_{3j}+e_{15}\eta_{4j}}{c_{33}\eta_{3j}+e_{33}\eta_{4j}} = \frac{e_{31}+e_{15}+e_{15}\eta_{3j}-\varepsilon_{11}\eta_{4j}}{e_{33}\eta_{3j}-\varepsilon_{33}\eta_{4j}} = \gamma_j^2 \tag{12}$$

Moreover, from the constitutive equations, expressions for the stresses and electric displacements in terms of $A_j(\xi)$ are also obtainable. For example, we have

$$\sigma_{rr}(r,z) = -\sum_{j=1}^{3} \beta_{0j} \int_0^\infty \xi A_j(\xi) e^{-\gamma_j \xi z} J_0(\xi r)\, \mathrm{d}\xi - (c_{11}-c_{12}) r^{-1} \times \sum_{j=1}^{3} \int_0^\infty A_j(\xi) e^{-\gamma_j \xi z} J_1(\xi r)\, \mathrm{d}\xi + (c_{11}+c_{12})B_1 + c_{13}B_3 + e_{31}B_4 \tag{13a}$$

$$\sigma_{zz}(r,z) = -\sum_{j=1}^{3} \beta_{1j} \int_0^\infty \xi A_j(\xi) e^{-\gamma_j \xi z} J_0(\xi r)\, \mathrm{d}\xi + 2c_{13}B_1 + c_{33}B_3 + e_{33}B_4 \tag{13b}$$

$$\sigma_{rz}(r,z) = -\sum_{j=1}^{3} \beta_{2j} \int_0^\infty \xi A_j(\xi) e^{-\gamma_j \xi z} J_1(\xi r)\, \mathrm{d}\xi \tag{13c}$$

$$D_r(r,z) = -\sum_{j=1}^{3} \beta_{3j} \int_0^\infty \xi A_j(\xi) e^{-\gamma_j \xi z} J_1(\xi r)\, \mathrm{d}\xi \tag{13d}$$

$$D_z(r,z) = -\sum_{j=1}^{3} \beta_{4j} \int_0^\infty \xi A_j(\xi) e^{-\gamma_j \xi z} J_0(\xi r)\, \mathrm{d}\xi + 2e_{31}B_1 + e_{33}B_3 - \varepsilon_{33}B_4 \tag{13e}$$

where β_{kj} $(k=0,1,\ldots,4, j=1,2,3)$ are given in the Appendix.

As a straightforward check, substitution of (13) into the equilibrium equations reveals that these equations are satisfied identically. The remaining task is how to get unknown B_j and $A_j(\xi)$

through appropriate electric and elastic boundary conditions and further determine electroelastic field for the crack problem posed by associated mixed boundary conditions.

Consideration of symmetry of the problem allows us to conclude that the shear stress at the crack plane vanishes, i.e.,

$$\sigma_{rz}(r,0)=0, \quad 0\leqslant r<\infty \tag{14a}$$

Since attention is restricted to the upper half-space, the following condition:

$$u_z(r,0)=0, \quad \phi(r,0)=0, \quad r\geqslant a \tag{14b}$$

must be supplemented owing to symmetry of the problem. Besides, at the crack surfaces, electromechanical boundary conditions

$$\sigma_{zz}(r,0)=0, \quad r<a \tag{14c}$$

$$D_z(r,0)=D^{(c)}, \quad r<a \tag{14d}$$

should be satisfied, where $D^{(c)}$ is a constant to be determined through (7).

First, from the remote electromechanical loading expressed by (8), we can get a system of linear equations that can be used to determine uniquely $B_k\,(k=1,3,4)$. The final result is

$$B_1=\frac{(c_{33}e_{31}-c_{13}e_{33})E_0^\infty-c_{13}\sigma_0^\infty}{(c_{11}+c_{12})c_{33}-2c_{13}^2} \tag{15a}$$

$$B_3=\frac{[(c_{11}+c_{12})e_{33}-2c_{13}e_{31}]E_0^\infty+(c_{11}+c_{12})\sigma_0^\infty}{(c_{11}+c_{12})c_{33}-2c_{13}^2} \tag{15b}$$

$$B_4=-E_0^\infty \tag{15c}$$

Knowledge of $B_k\,(k=1,3,4)$ permits us to further seek the disturbed electroelastic field in a piezoelectric body weakened by a penny-shaped crack. To this end, by substituting the above results into (11) and (13), application of (13*b*) and (13*e*) to the conditions (14*c*) and (14*d*), respectively, leads to

$$-\sum_{j=1}^{3}\beta_{1j}\int_0^\infty \xi A_j(\xi)J_0(\xi r)\mathrm{d}\xi+\sigma_0^\infty=0, \quad r<a \tag{16a}$$

$$-\sum_{j=1}^{3}\beta_{4j}\int_0^\infty \xi A_j(\xi)J_0(\xi r)\mathrm{d}\xi+D_0^\infty=D^{(c)}, \quad r<a \tag{16b}$$

with

$$D^{(c)}\sum_{j=1}^{3}\eta_{3j}\gamma_j\int_0^\infty A_j(\xi)J_0(\xi r)\mathrm{d}\xi =-\varepsilon^{(c)}\sum_{j=1}^{3}\eta_{4j}\gamma_j\int_0^\infty A_j(\xi)J_0(\xi r)\mathrm{d}\xi, \quad r<a \tag{17a}$$

$$D_0^\infty=\frac{(c_{11}+c_{12})e_{33}-2c_{13}e_{31}}{(c_{11}+c_{12})c_{33}-2c_{13}^2}\sigma_0^\infty +\left[\frac{(c_{11}+c_{12})e_{33}^2+2c_{33}e_{31}^2-4c_{13}e_{33}e_{31}}{(c_{11}+c_{12})c_{33}-2c_{13}^2}+\varepsilon_{33}\right]E_0^\infty \tag{17b}$$

Additionally, from (11*b*) and (11*c*) in conjunction with the conditions in (14*b*), we have

$$\sum_{j=1}^{3}\eta_{3j}\gamma_j\int_0^\infty A_j(\xi)J_0(\xi r)\mathrm{d}\xi=0, \quad r\geqslant a \tag{18a}$$

$$\sum_{j=1}^{3}\eta_{4j}\gamma_j\int_0^\infty A_j(\xi)J_0(\xi r)\mathrm{d}\xi=0, \quad r\geqslant a \tag{18b}$$

Because $D^{(c)}$ is assumed to be an unknown constant,

$$\sum_{j=1}^{3}[D^{(c)}\eta_{3j}+\varepsilon^{(c)}\eta_{4j}]\gamma_jA_j(\xi)=0 \tag{19}$$

follows from (17*a*) together with (18). On the other hand, utilizing the boundary conditions (14*a*) yields

$$\sum_{j=1}^{3}\beta_{2j}A_j(\xi)=0 \tag{20}$$

Now we arrive at two equations (20) and (19) for $A_j(\xi)\,(j=1,2,3)$, which are solvable up to two unknowns. In other words, two unknowns may be represented by the remaining one. To achieve this, we may choose a new intermediate auxiliary function $A(\xi)$ such that

$$A_j(\xi)=a_jA(\xi) \tag{21}$$

where a_j's are constants. Putting the above into (20) and (19) yields, respectively,

$$\sum_{j=1}^{3}\beta_{2j}a_j=0 \tag{22a}$$

$$\sum_{j=1}^{3}[D^{(c)}\eta_{3j}+\varepsilon^{(c)}\eta_{4j}]\gamma_ja_j=0 \tag{22b}$$

Furthermore, substituting (21) into (16*a*) and (16*b*), by comparison we find

$$\sigma_0^\infty\sum_{j=1}^{3}\beta_{4j}a_j+[D^{(c)}-D_0^\infty]\sum_{j=1}^{3}\beta_{1j}a_j=0 \tag{22c}$$

Accordingly, Eqs. (22*a*)–(22*c*) form a system of linear algebraic equations for a_j, which can be rewritten in a compact form

$$\Lambda\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix} \tag{23}$$

with

$$\Lambda=\begin{bmatrix}\beta_{11}[D^{(c)}-D_0^\infty]+\beta_{41}\sigma_0^\infty & \beta_{12}[D^{(c)}-D_0^\infty]+\beta_{42}\sigma_0^\infty & \beta_{13}[D^{(c)}-D_0^\infty]+\beta_{43}\sigma_0^\infty\\ \beta_{21} & \beta_{22} & \beta_{23}\\ [\eta_{31}D^{(c)}+\eta_{41}\varepsilon^{(c)}]\gamma_1 & [\eta_{32}D^{(c)}+\eta_{42}\varepsilon^{(c)}]\gamma_2 & [\eta_{33}D^{(c)}+\eta_{43}\varepsilon^{(c)}]\gamma_3\end{bmatrix} \tag{24}$$

In order to obtain a nontrivial solution of this equation, the determinant of the matrix Λ must take zero, from which $D^{(c)}$ can be determined. In particular, for several special situations, $D^{(c)}$ may be given via simple expressions.

i. In the absence of applied mechanical loading at infinity, this situation gives

$$D^{(c)}=D_0^\infty, \quad \text{or} \quad D^{(c)}=-\frac{\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2]}{\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_1]}\varepsilon^{(c)} \tag{25}$$

Hereafter β_k denotes the vector composed of $(\beta_{k1},\beta_{k2},\beta_{k3})^T$ $(k=0,1,\ldots,4)$, and η_{k-2} denotes the vector composed of $(\eta_{k1}\gamma_1,\eta_{k2}\gamma_2,\eta_{k3}\gamma_3)^T$ $(k=3,4)$, T being the transposition. Obviously, the first solution pertains to the case where a piezoelectric body without crack or two crack surfaces contact each other, and the second solution is reliant on the material properties and not on applied loading suitable for an opening crack. From this, one further finds the electric field inside the opening crack to be a constant $-\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2]/\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_1]$, independent of applied electric field and the dielectric permittivity. This phenomenon has been found in studying a dielectric crack in the two-dimensional case [19].

ii. In the case of an impermeable crack, $\varepsilon^{(c)}$ is assumed to be zero, and we then find

$$D^{(c)}=0, \quad \text{or} \quad D^{(c)}=D_0^\infty-\sigma_0^\infty\frac{\det[\boldsymbol{\beta}_4,\boldsymbol{\beta}_2,\boldsymbol{\eta}_1]}{\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_1]} \tag{26}$$

Clearly, the first solution $D^{(c)}=0$ corresponds to an opening impermeable crack, and the second corresponds to a closed crack, since two crack surfaces for a closed crack contact each other and the electric displacement at the crack surfaces is certainly no longer vanishing.

iii. For a permeable crack, $\Delta\phi(r,0)=0$ follows the assumption of the electrically permeable crack. Owing to the fact that $D_z^{(c)}(r,0)$ and $\Delta u_z(r,0)$ are finite, in this situation, it is easy to derive

$$D^{(c)}=D_0^\infty-\sigma_0^\infty\frac{\det[\boldsymbol{\beta}_4,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2]}{\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2]} \tag{27}$$

which indicates that $D^{(c)}$ is a linear function of applied electric loading as well as applied mechanical loading. In addition, if letting $\varepsilon^{(c)}\to\infty$, the derived electric displacements at the crack surface are the same as that given by (27). This is to say that the electric displacements given by (27) are, in effect, equal to those for a conducting crack where applied electric fields are parallel or antiparallel to the poling direction. Due to the fact that a real permeable crack is not a conducting crack, even for the present study, i.e., applied electric fields parallel or antiparallel to the poling direction, the above result indicates a drawback of the electrically permeable assumption at the boundaries of an undeformed crack.

For a general case, expanding the determinant yields a quadric equation for $D^{(c)}$, and it admits two solutions

$$D^{(c)}=\frac{-m_1\pm\sqrt{m_1^2-4m_0m_2}}{2m_2} \tag{28}$$

where

$$m_0=\varepsilon^{(c)}\sigma_0^\infty\det[\boldsymbol{\beta}_4,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2]-\varepsilon^{(c)}D_0^\infty\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2] \tag{29a}$$

$$m_1=\varepsilon^{(c)}\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2]+\sigma_0^\infty\det[\boldsymbol{\beta}_4,\boldsymbol{\beta}_2,\boldsymbol{\eta}_1]-D_0^\infty\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_1] \tag{29b}$$

$$m_2=\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_1] \tag{29c}$$

Clearly, apart from the material properties of the piezoelectric matrix, the electric displacement $D^{(c)}$ in the opening crack interior is also dependent on the dielectric permittivity of the interior of the opening crack. Moreover, not only applied electric loading but also applied mechanical loading at infinity have a pronounced influence on $D^{(c)}$. The above result coincides with those for the analysis of a dielectric crack in the two-dimensional case [14,18,20–22].

Generally speaking, in the two roots given by (28) for a dielectric crack, only one is reasonable and the other is superfluous, which should be neglected. Since $\varepsilon^{(c)}$ is finite and lies in a range of 0 to ∞, corresponding to the impermeable and conducting cracks, respectively, the corresponding electric displacement $D^{(c)}$ inside the opening crack should be located at the range between two limiting values corresponding to $\varepsilon^{(c)}=0$ and $\varepsilon^{(c)}=\infty$, respectively. Thus, an acceptable $D^{(c)}$ may be selected. An alternative approach for determining an acceptable $D^{(c)}$ is to look for the one such that $\Delta u_z(x,0)\geqslant 0$, the physical interpretation of which is obvious, avoiding penetration of two crack surfaces. By computing for many practical examples, we find that $D^{(c)}$ selected from the abovementioned two methods are identical. The variation of $D^{(c)}$ with E_0^∞ for a dielectric crack embedded in a PZT-5H ceramic is plotted in Fig. 2, from which it is seen for a vacuum crack, the electric displacements at the crack surfaces lie between those for an impermeable crack and a conducting crack. In addition, the curve corresponding to $\varepsilon^{(c)}=0$ has an apparent turning point at about $E_0^\infty=-3$ kV/cm, i.e., the crack is closed as $E_0^\infty<-3$ kV/cm, and the crack starts to open as E_0^∞ exceeds -3 kV/cm in the presence of $\sigma_0^\infty=5$ MPa.

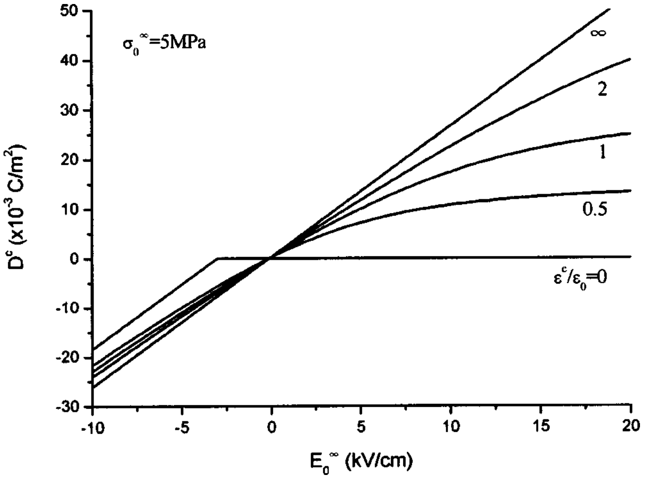


Fig. 2 Electric displacement versus applied electric field with $\sigma_0^\infty=5$ MPa for various values of $\varepsilon^{(c)}$ of the crack interior for a cracked PZT-5H ceramic

Due to the complexity of dependence (28) of $D^{(c)}$, we here present an approximate linear relation between $D^{(c)}$ and applied electric fields E_0^∞ for computing $D^{(c)}$ by the following form:

$$D_{\rm ap}^{(c)} = \lambda_E E_0^\infty \tag{30}$$

with

$$\lambda_E = \frac{\varepsilon^{(c)} g_2 \det[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\eta}_2]}{\varepsilon^{(c)} \det[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\eta}_2] + \sigma_0^\infty \det[\boldsymbol{\beta}_4, \boldsymbol{\beta}_2, \boldsymbol{\eta}_1] - g_1 \sigma_0^\infty \det[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\eta}_1]} \tag{31a}$$

where

$$g_1 = \frac{(c_{11}+c_{12})e_{33} - 2c_{13}e_{31}}{(c_{11}+c_{12})c_{33} - 2c_{13}^2},$$
$$g_2 = \frac{(c_{11}+c_{12})e_{33}^2 + 2c_{33}e_{31}^2 - 4c_{13}e_{33}e_{31}}{(c_{11}+c_{12})c_{33} - 2c_{13}^2} \tag{31b}$$

which is very accurate for E_0^∞ taking in a range from −10 kV/cm to 10 kV/cm.

Once $D^{(c)}$ is determined, nontrivial solution a_j can be expressed in terms of $D^{(c)}$ via solving arbitrary two equations among (22*a*)–(22*c*). For example, solving (22*a*) and (22*c*) we obtain

$$\begin{bmatrix} a_2/a_1 \\ a_3/a_1 \end{bmatrix} = -\begin{bmatrix} \beta_{22} & \beta_{23} \\ \sigma_0^\infty \beta_{42} + (D^{(c)} - D_0^\infty)\beta_{12} & \sigma_0^\infty \beta_{43} + (D^{(c)} - D_0^\infty)\beta_{13} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \beta_{21} \\ \sigma_0^\infty \beta_{41} + (D^{(c)} - D_0^\infty)\beta_{11} \end{bmatrix} \tag{32}$$

Next, applying the boundary conditions (16*a*) and (18*a*) we get a pair of simultaneous dual integral equations for $A(\xi)$

$$\kappa_m \int_0^\infty \xi A(\xi) J_0(\xi r) \mathrm{d}\xi = \sigma_0^\infty, \quad r < a \tag{33a}$$

$$\int_0^\infty A(\xi) J_0(\xi r) \mathrm{d}\xi = 0, \quad r \geqslant a \tag{33b}$$

where

$$\kappa_m = \sum_{j=1}^{3} \beta_{1j} a_j \tag{34}$$

This is a system of simultaneous dual integral equations involving Bessel functions, and a solution can be obtained according to a standard approach [35] is found to be

$$A(\xi) = \sqrt{\frac{2}{\pi}} \frac{a^{3/2}\sigma_0^\infty}{\kappa_m} \frac{J_{3/2}(\xi a)}{\sqrt{\xi}} \tag{35}$$

where $J_{3/2}(\cdot)$ is the Bessel function of the first kind.

It is pointed out that the above result is suitable only for the case where applied mechanical loading at infinity is nonzero $\sigma_0^\infty \neq 0$. In case of $\sigma_0^\infty = 0$, from (16*a*) we get $\Sigma_{j=1}^{3} \beta_{1j} a_j = 0$, and in this case Eq. (33*a*) becomes an identity. However, in the absence of mechanical loading, in view of the coupling characteristic, applied electric loading also causes deformation of a piezoelectric body and further may give rise to crack opening. Taking into account that the opening of crack is attributed to application of electric loading for this case, we therefore use Eq. (16*b*) instead of Eq. (16*a*) to obtain dual integral equations similar to (33). An analogous treatment gives a solution as

$$A(\xi) = \sqrt{\frac{2}{\pi}} \frac{a^{3/2}[D_0^\infty - D^{(c)}]}{\kappa_e} \frac{J_{3/2}(\xi a)}{\sqrt{\xi}} \tag{36}$$

where

$$\kappa_e = \sum_{j=1}^{3} \beta_{4j} a_j \tag{37}$$

In view of (22*c*), it is readily seen that in the presence of mechanical loading, the solution (35) coincides with (36). Similarly, (36) is valid only for $D_0^\infty \neq D^{(c)}$, since Eq. (16*b*) also becomes an identity when $D_0^\infty = D^{(c)}$. For convenience, we rewrite these two results in a uniform form, i.e.,

$$A(\xi) = \sqrt{\frac{2}{\pi}} \frac{a^{3/2}P}{\kappa} \frac{J_{3/2}(\xi a)}{\sqrt{\xi}} \tag{38}$$

where

$$P = \sigma_0^\infty, \quad \kappa = \kappa_m, \quad \text{as } \sigma_0^\infty \neq 0 \tag{39a}$$

$$P = D_0^\infty - D^{(c)}, \quad \kappa = \kappa_e, \quad \text{as } \sigma_0^\infty = 0 \tag{39b}$$

D_0^∞ and $D^{(c)}$ being defined by (17*b*) and (28), respectively.

With the above obtained results, the entire electroelastic field of a cracked piezoelectric body can be determined. This can be achieved by substituting (38) into (11) for elastic displacements and potential. Making use of some known results involving infinite integrals of Bessel functions, which are listed in the Appendix, explicit expressions for the elastic displacements and potential are obtained as follows:

$$u_r(r,z) = \frac{P}{\pi\kappa} \sum_{j=1}^{3} a_j \left[r \sin^{-1}\left(\frac{l_{1j}}{r}\right) - \frac{l_{1j}}{r} \sqrt{r^2 - l_{1j}^2} \right] + B_1 r \tag{40a}$$

$$u_z(r,z) = \frac{P}{\pi\kappa} \sum_{j=1}^{3} \eta_{3j} \gamma_j a_j \left[\sqrt{a^2 - l_{1j}^2} - \gamma_j z \sin^{-1}\left(\frac{l_{1j}}{r}\right) \right] + B_3 z \tag{40b}$$

$$\phi(r,z) = \frac{P}{\pi\kappa} \sum_{j=1}^{3} \eta_{4j} \gamma_j a_j \left[\sqrt{a^2 - l_{1j}^2} - \gamma_j z \sin^{-1}\left(\frac{l_{1j}}{r}\right) \right] - E_0^\infty z \tag{40c}$$

where B_1 and B_3 are given by (15*a*) and (15*b*), respectively, and

$$l_{1j} = \frac{1}{2}\left[\sqrt{(r+a)^2 + (\gamma_j z)^2} - \sqrt{(r-a)^2 + (\gamma_j z)^2} \right] \tag{41a}$$

$$l_{2j} = \frac{1}{2}\left[\sqrt{(r+a)^2 + (\gamma_j z)^2} + \sqrt{(r-a)^2 + (\gamma_j z)^2} \right] \tag{41b}$$

In a similar fashion, from (13) we can further give a complete solution of elastic stresses, electric displacements, and electric fields in the entire space. Or rather, utilizing the relevant formulas listed in the Appendix, the distribution of electroelastic field in the entire space is

$$\sigma_{rr}(r,z)=\frac{P}{\pi\kappa}\sum_{j=1}^{3}\left[2\beta_{0j}a_j h_{2j}(r,z)+(c_{11}-c_{12})a_j h_{0j}(r,z)\right] \tag{42a}$$

$$\sigma_{\theta\theta}(r,z)=\frac{P}{\pi\kappa}\sum_{j=1}^{3}\left[2\beta_{5j}a_j h_{2j}(r,z)-(c_{11}-c_{12})a_j h_{0j}(r,z)\right] \tag{42b}$$

$$\sigma_{zz}(r,z)=\frac{2P}{\pi\kappa}\sum_{j=1}^{3}\beta_{1j}a_j h_{2j}(r,z)+\sigma_0^{\infty} \tag{42c}$$

$$\sigma_{rz}(r,z)=-\frac{2P}{\pi\kappa}\sum_{j=1}^{3}\beta_{2j}a_j h_{1j}(r,z) \tag{42d}$$

$$D_r(r,z)=-\frac{2P}{\pi\kappa}\sum_{j=1}^{3}\beta_{3j}a_j h_{1j}(r,z),\quad D_z(r,z)=\frac{2P}{\pi\kappa}\sum_{j=1}^{3}\beta_{4j}a_j h_{2j}(r,z)+D_0^{\infty} \tag{43a}$$

$$E_r(r,z)=\frac{2P}{\pi\kappa}\sum_{j=1}^{3}\eta_{4j}\gamma_j a_j h_{1j}(r,z),\quad E_z(r,z)=-\frac{2P}{\pi\kappa}\sum_{j=1}^{3}\eta_{4j}\gamma_j^2 a_j h_{2j}(r,z)+E_0^{\infty} \tag{43b}$$

where β_{ij} $(j=1,2,3,i=0,\ldots,4)$ are defined as before,

$$\beta_{5j}=(c_{13}\eta_{3j}+e_{31}\eta_{4j})\gamma_j^2-c_{12} \tag{44}$$

and

$$h_{1j}(r,z)=\frac{l_{1j}^2\sqrt{a^2-l_{1j}^2}}{r(l_{2j}^2-l_{1j}^2)},\quad h_{2j}(r,z)=\left[\frac{a\sqrt{l_{2j}^2-a^2}}{l_{2j}^2-l_{1j}^2}-\sin^{-1}\left(\frac{l_{1j}}{r}\right)\right] \tag{45a}$$

$$h_{0j}(r,z)=\frac{l_{1j}}{r^2}\sqrt{r^2-l_{1j}^2}-\sin^{-1}\left(\frac{l_{1j}}{r}\right) \tag{45b}$$

Therefore, explicit analytic expressions for the complete electroelastic field are given in terms of elementary functions. From the above results, explicit expressions for electroelastic field for an impermeable crack and for a conducting crack can be directly written out only if setting $\varepsilon^{(c)}=0$ and $\varepsilon^{(c)}=\infty$, respectively. In addition, if imposing the piezoelectric coefficients vanish, the electric and elastic behaviors are uncoupled, and the corresponding elastic field reduces to the results relating to a penny-shaped crack embedded in an infinite transversely isotropic medium subjected to uniform tension at infinity. It is interesting to note that for the latter, explicit expressions in terms of elementary functions are not given by Sneddon and Lowengrub [36], who gave their results involving complicated infinite integrals.

In what follows we confine our attention to the crack plane, $z=0$. In this case, using the properties of l_{1j} and l_{2j} [37]

$$\lim_{z\to 0} l_{1j}(r,z)=\begin{cases} r, & \text{as } r<a,\\ a, & \text{as } r>a\end{cases},\quad \lim_{z\to 0} l_{2j}(r,z)=\begin{cases} a, & \text{as } r<a\\ r, & \text{as } r>a\end{cases} \tag{46}$$

we get immediately elastic displacements and potential in the crack plane, from the results (40)

$$u_r(r,0)=\frac{PH(r-a)}{\pi\kappa}\sum_{j=1}^{3}a_j\left[r\sin^{-1}\left(\frac{a}{r}\right)-\frac{a}{r}\sqrt{r^2-a^2}\right]+\frac{rPH(a-r)}{2\kappa}\sum_{j=1}^{3}a_j+\frac{(c_{33}e_{31}-c_{13}e_{33})E_0^{\infty}-c_{13}\sigma_0^{\infty}}{(c_{11}+c_{12})c_{33}-2c_{13}^2}r \tag{47a}$$

$$u_z(r,0)=\frac{PH(a-r)}{\pi\kappa}\sum_{j=1}^{3}\eta_{3j}\gamma_j a_j\sqrt{a^2-r^2} \tag{47b}$$

$$\phi(r,0)=\frac{PH(a-r)}{\pi\kappa}\sum_{j=1}^{3}\eta_{4j}\gamma_j a_j\sqrt{a^2-r^2} \tag{47c}$$

where $H(t)$ denotes the Heaviside unit step function, i.e., $H(t)=1$ as $t>0$ and $H(t)=0$ as $t<0$. Moreover, from the results (42) in conjunction with (39), it follows that aside from the apparent conclusion $\sigma_{rz}(r,0)=0$, other nonvanishing elastic stresses, electric displacement, and electric field in the crack plane are

$$\sigma_{zz}(r,0)=\begin{cases}\dfrac{2\sigma_0^{\infty}}{\pi}\left[\dfrac{a}{\sqrt{r^2-a^2}}-\sin^{-1}\left(\dfrac{a}{r}\right)\right]+\sigma_0^{\infty}, & \text{as } r>a\\ 0, & \text{as } r<a,\end{cases} \tag{48a}$$

$$D_z(r,0)=\begin{cases}\dfrac{2[D_0^{\infty}-D^{(c)}]}{\pi}\left[\dfrac{a}{\sqrt{r^2-a^2}}-\sin^{-1}\left(\dfrac{a}{r}\right)\right]+D_0^{\infty}, & \text{as } r>a\\ D^{(c)}, & \text{as } r<a,\end{cases} \tag{48b}$$

$$E_z(x,0)=\begin{cases}-\dfrac{2P}{\pi\kappa}\displaystyle\sum_{j=1}^{3}\eta_{4j}\gamma_j^2 a_j\left[\dfrac{a}{\sqrt{r^2-a^2}}-\sin^{-1}\left(\dfrac{a}{r}\right)\right]+E_0^{\infty}, & \text{as } r>a\\ \dfrac{1}{\varepsilon^{(c)}}D^{(c)}, & \text{as } r<a.\end{cases} \tag{48c}$$

Note that from (48a), $\sigma_{zz}(r,z)$ for $z=0$ is independent of applied electric loading and material properties. Nevertheless, the distribution of $\sigma_{zz}(r,z)$ for $z\neq 0$ around the crack front is reliant on the material properties, which is apparently seen from (42c) since a_j together with κ are determined by combined loading, including mechanical loading as well as electric loading. In contrast, $D_z(r,z)$ around the crack front is dependent on applied mechanical loading no matter whether $z=0$ or not, which is due to the fact that $D^{(c)}$ is controlled by applied mechanical loading, apart from electric loading. Furthermore, it is not difficult to find that the behavior of $E_z(r,z)$ is the same as $D_z(r,z)$. Obviously, it is seen that all the electroelastic quantities exhibit a usual square-root singularity near the crack front $r=a$, in accordance with the counterpart for a two-dimensional case.

By comparing these expressions for electroelastic field in the crack plane obtained above with those given by Zhang et al. [17], it is found that two results are the same in form except for the electric displacements $D^{(c)}$ at the crack surfaces. Zhang et al. [17] started with an elliptical hole before deformation and derived the electric displacements $D^{(c)}$ at the hole boundary by use of the exact electric boundary conditions. The results indicate that when considering a very slender elliptical hole (i.e., reducing a crack), $D^{(c)}$ are strongly dependent on a competition factor α/β^*, α and β^* being defined by Zhang et al. ([17], p. 227). In the present study, a *flat* penny-shaped crack *before deformation* and an opening elliptical hole after deformation are assumed, and further $D^{(c)}$ are related to the eventual COD of the *deformed* elliptical hole

Table 1 Relevant material properties

	Elastic Stiffnessess ($\times 10^{10\ \mathrm{N/m^2}}$)					Piezoelectric constants (C/m^2)			Dielectric permittivities ($\times 10^{10}$ F/m)	
	c_{11}	c_{33}	c_{44}	c_{12}	c_{13}	e_{31}	e_{33}	e_{15}	ε_{11}	ε_{33}
PZT-5H	12.6	11.7	3.53	5.5	5.3	−6.5	23.3	17	151	130

(opening crack). Under these circumstances, values of $D^{(c)}$ evaluated by (28) are also strongly dependent on both materials properties, including the permittivity and applied electromechanical loadings, which is in exact agreement with those for the two-dimensional analysis (see e.g., [14,18,20–22]).

4 Asymptotic Field and Intensity Factors

In studying the stability of a cracked piezoelectric structure, of importance from the viewpoint of fracture mechanics is the distribution of the asymptotic electroelastic field around the crack front, which can be characterized as field intensity factors.

In order to obtain the required asymptotic field, we introduce a local polar coordinate system (ρ,θ) with the origin at the periphery of the crack, which satisfies

$$\rho=\sqrt{(r-a)^2+z^2}, \quad \theta=\tan^{-1}[z/(r-a)] \tag{49}$$

In the close vicinity of the crack front, i.e., $\rho \ll a$, we have

$$l_{1j} \simeq a+\frac{\rho}{2}\left[\cos(\theta)-\sqrt{\cos^2(\theta)+\gamma_j^2 \sin^2(\theta)}\right] \tag{50a}$$

$$l_{2j} \simeq a+\frac{\rho}{2}\left[\cos(\theta)+\sqrt{\cos^2(\theta)+\gamma_j^2 \sin^2(\theta)}\right] \tag{50b}$$

Upon substitution of these into (42), by neglecting some higher-order infinitesimal terms, the asymptotic expressions for electroelastic field in the vicinity of the crack front are derived below:

$$\begin{bmatrix} \sigma_{\rho\rho}(\rho,\theta) \\ \sigma_{zz}(\rho,\theta) \\ \sigma_{\rho z}(\rho,\theta) \end{bmatrix} \simeq \frac{2P}{\pi\kappa}\sqrt{\frac{a}{2\rho}}\sum_{j=1}^{3}\begin{bmatrix} \beta_{0j}a_j f_{2j}(\theta) \\ \beta_{1j}a_j f_{2j}(\theta) \\ -\beta_{2j}a_j f_{1j}(\theta) \end{bmatrix} \tag{51a}$$

$$\begin{bmatrix} D_{\rho}(\rho,\theta) \\ D_{z}(\rho,\theta) \\ E_{\rho}(\rho,\theta) \\ E_{z}(\rho,\theta) \end{bmatrix} \simeq \frac{2P}{\pi\kappa}\sqrt{\frac{a}{2\rho}}\sum_{j=1}^{3}\begin{bmatrix} -\beta_{3j}a_j f_{1j}(\theta) \\ \beta_{4j}a_j f_{2j}(\theta) \\ \eta_{4j}\gamma_j a_j f_{1j}(\theta) \\ -\eta_{4j}\gamma_j^2 a_j f_{2j}(\theta) \end{bmatrix} \tag{51b}$$

where $f_{1j}(\theta)$ and $f_{2j}(\theta)$ denote the functions of angle distribution, defined by

$$f_{1j}(\theta)=\frac{1}{\sqrt[4]{\cos^2(\theta)+\gamma_j^2 \sin^2(\theta)}} \times \sqrt{\frac{1}{2}\left[1-\frac{\cos(\theta)}{\sqrt{\cos^2(\theta)+\gamma_j^2 \sin^2(\theta)}}\right]} \tag{52a}$$

$$f_{2j}(\theta)=\frac{1}{\sqrt[4]{\cos^2(\theta)+\gamma_j^2 \sin^2(\theta)}} \times \sqrt{\frac{1}{2}\left[1+\frac{\cos(\theta)}{\sqrt{\cos^2(\theta)+\gamma_j^2 \sin^2(\theta)}}\right]} \tag{52b}$$

From the above, the intensity factors of stress, electric-displacement, and electric-field near the crack front, according to their definitions

$$K^q=\lim_{r\to a^+}\sqrt{2\pi(r-a)}\,q(r,0) \tag{53}$$

where q stands for σ_{zz}, D_z, and E_z, respectively, can be evaluated as

$$K^{\sigma}=\frac{2\sigma_0^{\infty}}{\pi}\sqrt{\pi a}, \quad K^{D}=\frac{2[D_0^{\infty}-D^{(c)}]}{\pi}\sqrt{\pi a},$$
$$K^{E}=-\frac{2P}{\pi\kappa}\sum_{j=1}^{3}\eta_{4j}\gamma_j^2 a_j\sqrt{\pi a} \tag{54}$$

As seen from the above, no matter how applied electric loading varies, the stress intensity factor maintains unchanged, implying that stress intensity factors near the crack front are inapplicable to predicting crack growth of piezoelectric materials. On the contrary, the intensity factors of electric displacement and electric field depend on the material properties and applied mechanical loading. In particular, if setting $\varepsilon^{(c)}=0$, we find $K^D=2D_0^{\infty}\sqrt{a/\pi}$, independent of applied mechanical stress, which is in agreement with existing results, such as [27–29], whereas if setting $\Delta\phi(r,0)=0$ or $\varepsilon^{(c)}=\infty$, we find $K^D=2\sigma_0^{\infty}\det[\boldsymbol{\beta}_4,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2]\sqrt{a}/(\det[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2]\sqrt{\pi})$, dependent solely on applied mechanical stress, in agreement with those obtained in [25,30,32,33]. Therefore, the above obtained conclusions under impermeable and permeable conditions are completely opposite. However, a real crack is neither electrically impermeable nor electrically permeable (at the boundaries of an undeformed crack). Consideration of a dielectric crack results in an important conclusion. That is, applied mechanical loading strongly affects the singularity of the electric displacement near the crack front, and also its intensity factor varies with the dielectric permittivity of the crack interior. Moreover, the impermeable and permeable cracks can be taken as two limiting cases of a dielectric crack.

Additionally, we introduce a factor to characterize the behavior of the COD $u_z(r,0)$ near the crack front, referred as to the COD intensity factor, which is defined by

$$K_{\mathrm{COD}}=\lim_{r\to a^-}\sqrt{\frac{2\pi}{a-r}}\,u_z(r,0) \tag{55}$$

With the help of (47*b*), one can immediately obtain

$$K_{\mathrm{COD}}=\frac{2P}{\pi\kappa}\sqrt{\pi a}\sum_{j=1}^{3}\eta_{3j}\gamma_j a_j \tag{56}$$

identical to the crack center opening displacement $2u_z(0,0)$, aside from a factor of $\sqrt{a/\pi}$.

5 Numerical Results and Discussions

In this section numerical computations are carried out for commercially available PZT-5H ceramic with the relevant materials properties, listed in Table 1 [3].

Figure 3 shows the variation of the COD intensity factor K_{COD} with applied electric field E_0^{∞} for three different crack models, an impermeable crack, a vacuum dielectric crack, and a permeable crack, in the case of mechanical loadings $\sigma_0^{\infty}=5$, 20, and 0 MPa, respectively. Here the radius of the penny-shaped crack is chosen as $a=1$ mm. It is noted that the solution corresponding to a permeable crack is identical to that corresponding to a conducting crack for applied electric fields parallel or antiparallel to the pol-

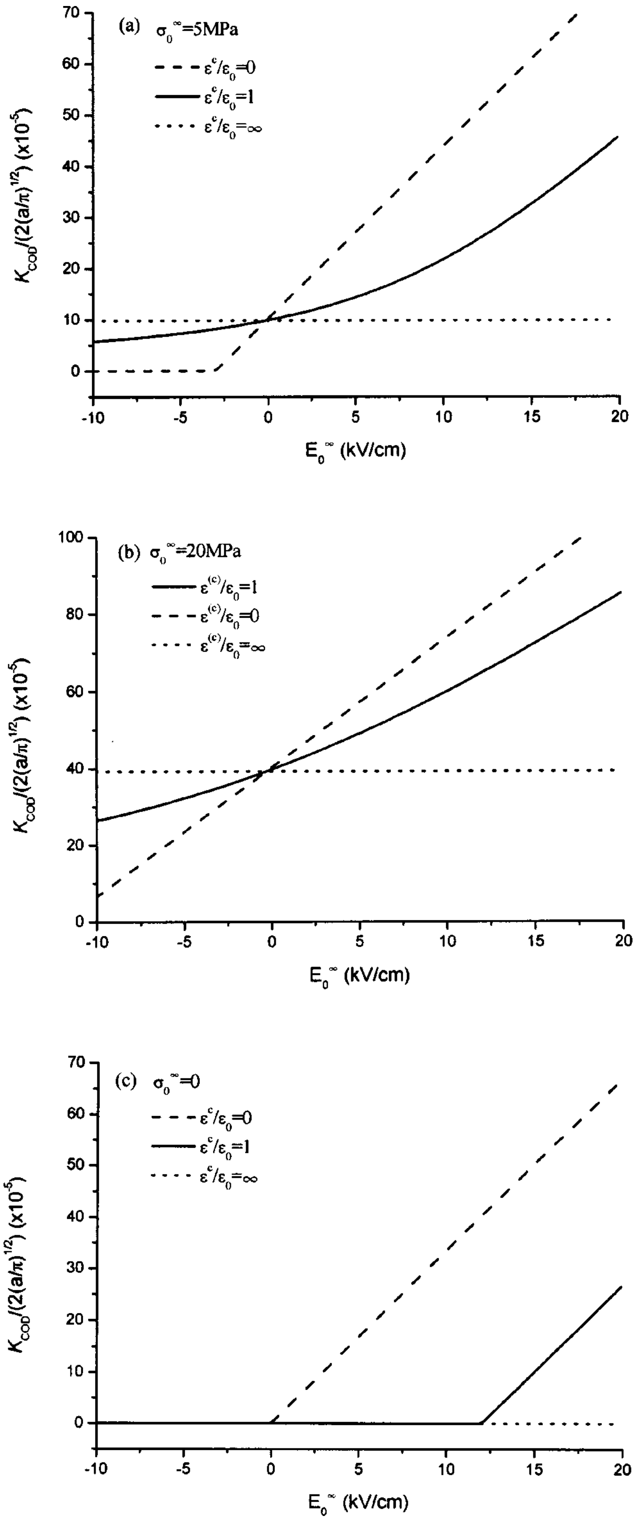


Fig. 3 Variation of the COD intensity factor K_{COD} with applied electric field E_0^∞ for $\varepsilon^{(c)}/\varepsilon_0=0$, 1, ∞, (*a*) $\sigma_0^\infty=5$ MPa, (*b*) $\sigma_0^\infty=20$ MPa, and (*c*) $\sigma_0^\infty=0$ MPa

ing axis because of (27). It is found from Fig. 3(*a*) for $\sigma_0^\infty=5$ MPa that for a vacuum dielectric crack, a positive electric field increases K_{COD}, while a negative electric field decrease K_{COD}, which implies that applied electric fields can aid or impede crack growth depending on positive or negative directions of electric fields, in agreement with the experimental observations by Park and Sun [38]. Therefore, it is more reasonable to adopt the COD intensity factor as a fracture criterion to assess the fracture behavior of piezoelectric materials. Especially, in the case of $\varepsilon^{(c)}/\varepsilon_0=0$ corresponding to an impermeable crack, the crack keeps closed ($K_{\mathrm{COD}}=0$) when $E_0^\infty<-3$ kV/cm, and starts to open when $E_0^\infty>-3$ kV/cm. In contrast, for a permeable crack or a conducting crack, the opening of a crack does not vary with applied electric fields, as expected. Also it is found that three curves corresponding to $\varepsilon^{(c)}/\varepsilon_0=0$, 1, and ∞ intersect a point at $E_0^\infty=0$, implying that values of K_{COD} for purely mechanical loading are the same for a crack, irrespective of the dielectric permittivity of the crack interior. Similar trends are seen in Fig. 3(*b*) for $\sigma_0^\infty=20$ MPa. The only difference lies in the fact that the impermeable crack does not close as $E_0^\infty>-10$ kV/cm for $\sigma_0^\infty=20$ MPa, while it is closed as $E_0^\infty<-3$ kV/cm for $\sigma_0^\infty=5$ MPa, which is easily understood since applied mechanical tension $\sigma_0^\infty=20$ MPa is much larger than $\sigma_0^\infty=5$ MPa. In the absence of mechanical loading [Fig. 3(*c*)], a purely electric field cannot drive crack growth for a permeable or conducting crack. This is easily understood because the electric boundary conditions for a permeable crack are adopted at the boundaries of an undeformed crack. However, this is not true for a dielectric crack and an impermeable crack. For a vacuum crack, the crack surfaces remain in contact for a negative electric field because the piezoelectric material shrinks in the poling direction, in this case, and they start to open until the electric field exceeds a certain positive electric field E_c because it expands in the poling directions. Here E_c denotes a turning point from a closed crack to an opening crack, which is related to the dielectric permittivity of the crack interior. Obviously, E_c is shifted to the origin when $\varepsilon^{(c)}/\varepsilon_0$ drops down to 0. In other words, for an impermeable crack, the crack surfaces keep contact for negative electric fields and open immediately only when positive electric fields are applied.

Similarly, the variation of K_{COD} with applied mechanical loading σ_0^∞ is displayed for three different values $\varepsilon^{(c)}/\varepsilon_0=0$, 1, ∞, for $E_0^\infty=0$, ± 5 kV/cm in Fig. 4. Clearly, in the absence of electric fields, values of the COD caused by purely mechanical loading are almost the same for various values of $\varepsilon^{(c)}/\varepsilon_0$. Nevertheless, the results manifest an important difference in the presence of electric fields. For example, from Figs. 4(b) and 4(c) it is observed that values of K_{COD} for an opening dielectric crack gradually become close to those of K_{COD} for an impermeable crack with an increase of σ_0^∞. Conversely, they gradually become close to those for a conducting crack with the decrease of σ_0^∞, which implies the influence of $\varepsilon^{(c)}$ inside the opening crack. This coincides with the theoretical prediction for a two-dimensional cracked piezoelectric medium by Wang and Jiang [21]. Especially, it is observed from Figs. 4(b) and 4(c) that for an impermeable crack, application of positive electric fields drives crack opening, even in the absence of mechanical loading, while application of negative electric fields cannot drive crack opening for lower mechanical loading, which is attributed to negative electric fields that cause the piezoelectric material to shrink in the poling direction. When applied mechanical loading is large enough, the crack begins to open, agreeing with that in Fig. 3.

Figures 5 and 6 are devoted to the effects of applied mechanical loading on the intensity factors of electric displacement and electric field, respectively. As pointed out above, the stress intensity factor is independent of applied electric field. However, the electric displacement intensity factor K^D depends strongly on applied mechanical loading, which is seen in Fig. 5. When an applied electric field is absent, the effect of σ_0^∞ is small, as compared to that when an electric field is present, and exhibits almost a linear relation. For the latter, applied mechanical loading has a strong influence on K^D. For a positive electric field, K^D for a dielectric crack increases and gradually approaches that for an impermeable crack as σ_0^∞ is raised from 0, and decreases and gradually ap-

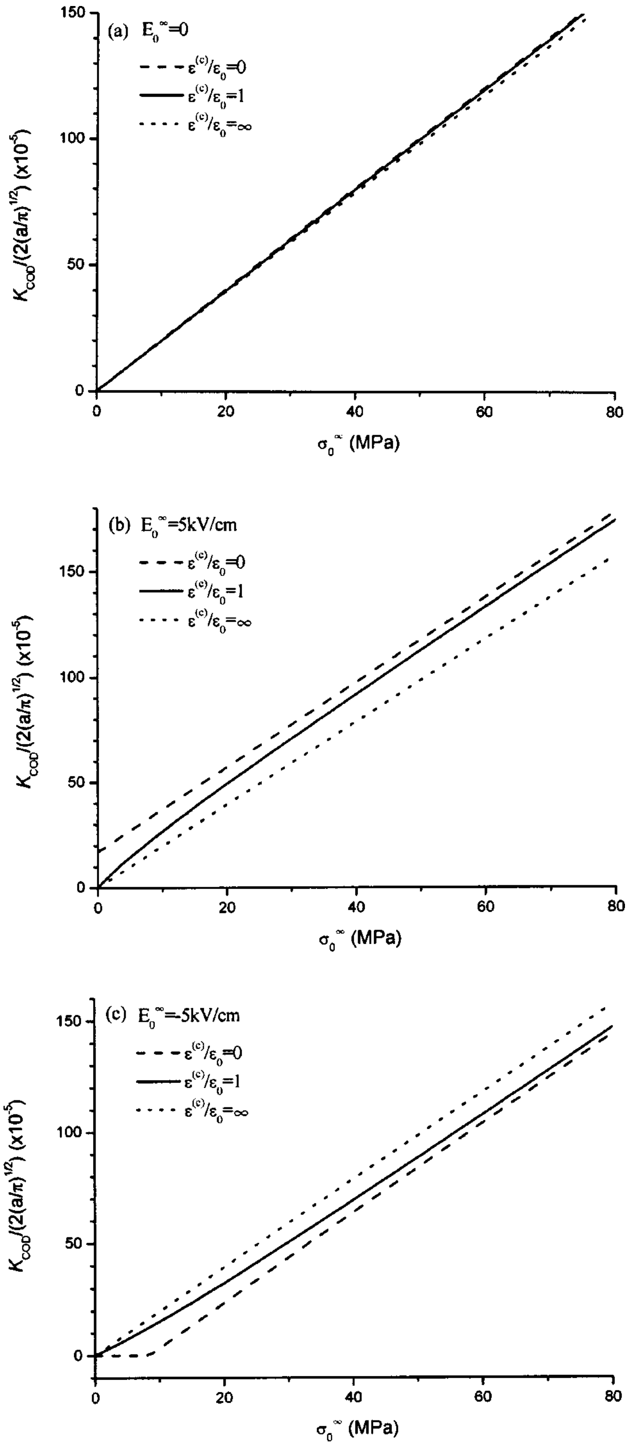


Fig. 4 Variation of the COD intensity factor K_{COD} with applied mechanical stress σ_0^∞ for $\varepsilon^{(c)}/\varepsilon_0=0,\ 1,\ \infty$, (*a*) $E_0^\infty=0$ kV/cm, (*b*) $E_0^\infty=5$ kV/cm, and (*c*) $E_0^\infty=-5$ kV/cm

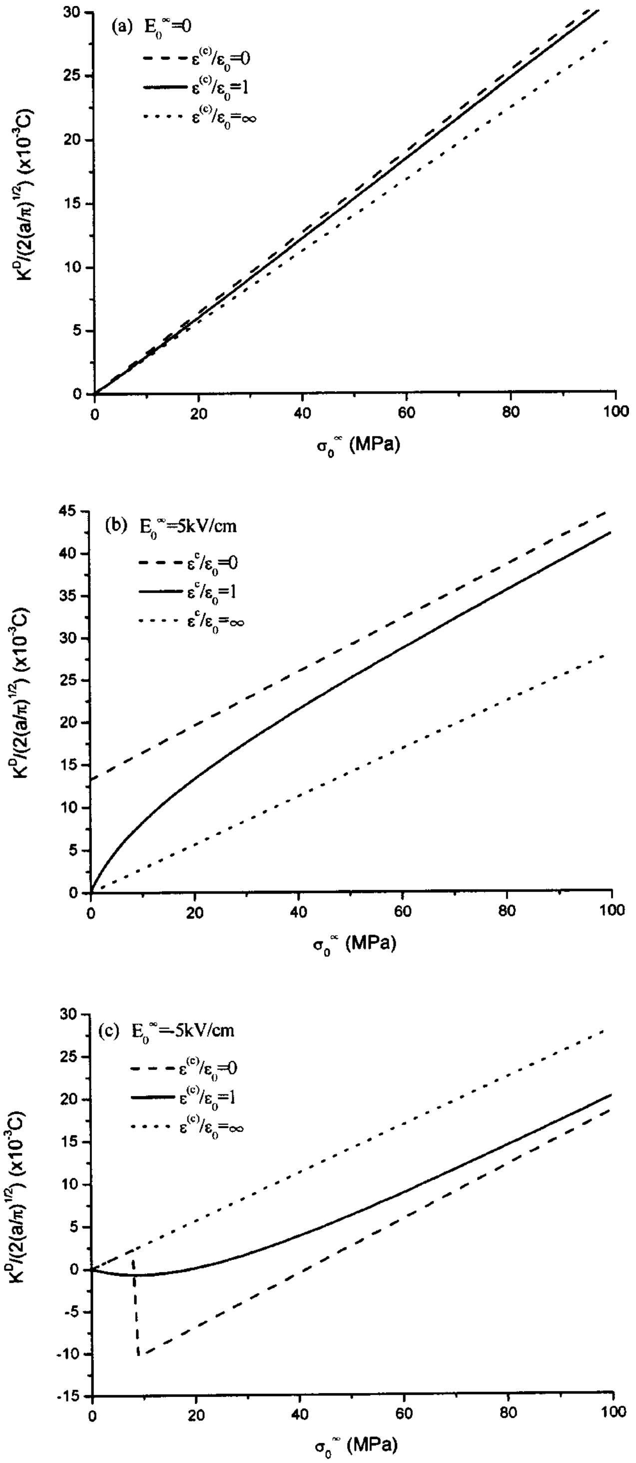


Fig. 5 Electric displacement intensity factor K^{D} as a function of applied mechanical stress σ_0^∞ for $\varepsilon^{(c)}/\varepsilon_0=0,\ 1,\ \infty$, (*a*) $E_0^\infty=0$ kV/cm, (*b*) $E_0^\infty=5$ kV/cm, and (*c*) $E_0^\infty=-5$ kV/cm

proaches that for a conducting crack as σ_0^∞ drops to 0. While for a negative electric field, the tread is similar to the above, aside from that the curve corresponding to $\varepsilon^{(c)}=0$ is below that corresponding to $\varepsilon^{(c)}=\infty$ for $E_0^\infty<0$, which is reversed for $E_0^\infty>0$. This is attributed to the fact that positive and negative electric fields cause piezoelectric material to expand and shrink, respectively. Also with the increase of σ_0^∞, K^D for a dielectric crack still gradually approaches that for an impermeable crack. It is also seen that the curve corresponding to $\varepsilon^{(c)}=0$ coincides with that corresponding to $\varepsilon^{(c)}=\infty$ for lower σ_0^∞, which is still due to a closed crack, whose two surfaces keep contact with each other in this stage. Similar behaviors can be found for the electric field intensity factor K^E near the crack front in Fig. 6.

6 Conclusions

The electroelastic analysis of a transversely isotropic piezoelectric body with a penny-shaped dielectric crack perpendicular to

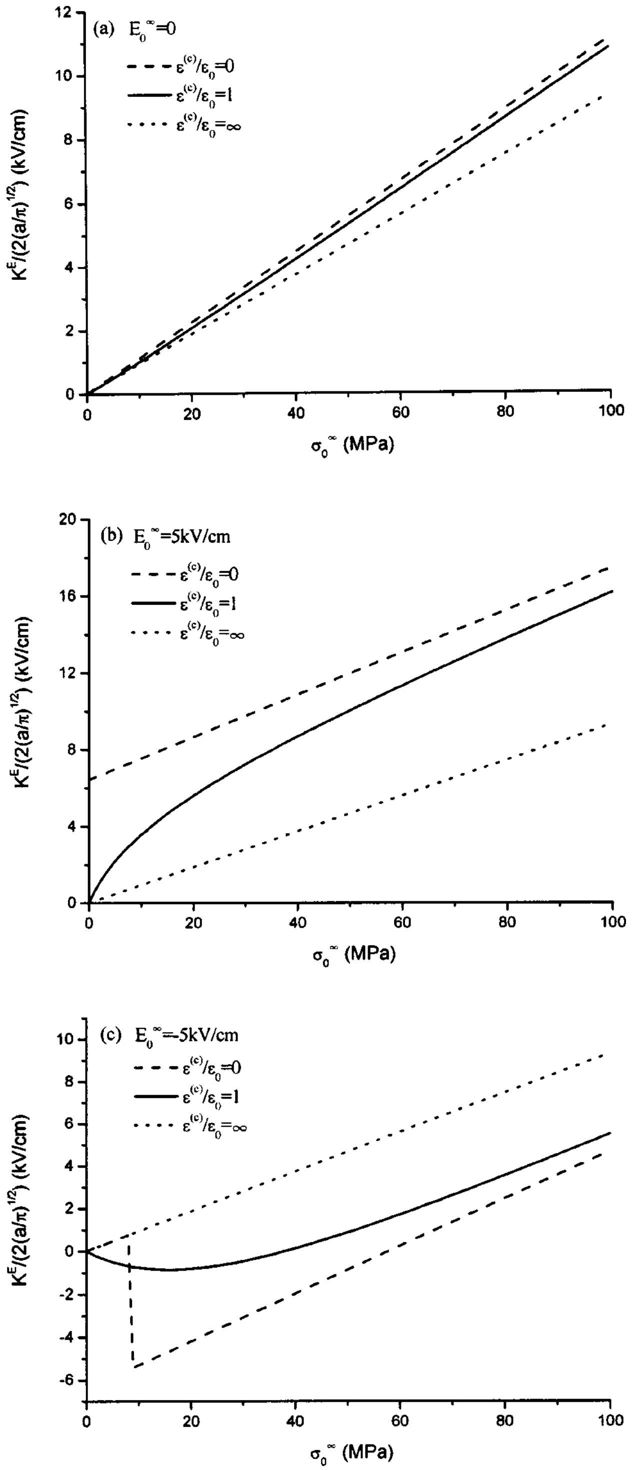


Fig. 6 Electric field intensity factor K^E as a function of applied mechanical stress σ_0^∞ for $\varepsilon^{(c)}/\varepsilon_0=0$, 1, ∞, (*a*) $E_0^\infty=0$ kV/cm, (*b*) $E_0^\infty=5$ kV/cm, and (*c*) $E_0^\infty=-5$ kV/cm

the poling direction is made within the framework of the theory of linear piezoelectricity. Using the electric boundary conditions at the crack surfaces dependent on the COD, the electric displacements at the crack surfaces are determined in terms of applied electromechanical loading as well as the dielectric permittivity of the crack interior. The Hankel transform technique is employed to reduce the mixed boundary value problem to dual integral equations. The results are presented for the case of remote uniform electromechanical loading. Explicit expressions for the complete electroelastic field at any point in the entire piezoelectric space are obtained in terms of elementary functions. In particular, the asymptotic electroelastic field around the crack front is also given in terms of angle distribution functions, and the field intensity factors are determined. The previous results for a penny-shaped impermeable or permeable crack are two special cases of the present results.

Acknowledgments

The authors are very grateful to the anonymous reviewers for their helpful comments for the improvement of this article and for bringing the authors' attention to additional important literature. This work was supported by the National Natural Science Foundation of China under Grant No. 10272043 and the Korea Institute of Science and Technology Evaluation and Planning.

Appendix

In deriving explicit expressions for the complete electroelastic field, the following integral equalities have been utilized:

$$\int_0^\infty e^{-c\xi}J_1(\xi r)J_{3/2}(\xi a)\sqrt{\xi}\,d\xi=\frac{2l_1^2\sqrt{a^2-l_1^2}}{\sqrt{2\pi}a^{3/2}r(l_2^2-l_1^2)} \qquad (A1)$$

$$\int_0^\infty e^{-c\xi}J_1(\xi r)J_{3/2}(\xi a)\frac{d\xi}{\sqrt{\xi}}=\frac{1}{\sqrt{2\pi}a^{3/2}r}\left[-l_1\sqrt{r^2-l_1^2}+r^2\sin^{-1}\left(\frac{l_1}{r}\right)\right] \qquad (A2)$$

$$\int_0^\infty e^{-c\xi}J_0(\xi r)J_{3/2}(\xi a)\sqrt{\xi}\,d\xi=\frac{2}{\sqrt{2\pi}a^{3/2}}\left[\sin^{-1}\left(\frac{l_1}{r}\right)-\frac{a\sqrt{l_2^2-a^2}}{l_2^2-l_1^2}\right] \qquad (A3)$$

$$\int_0^\infty e^{-c\xi}J_0(\xi r)J_{3/2}(\xi a)\frac{d\xi}{\sqrt{\xi}}=\frac{2}{\sqrt{2\pi}a^{3/2}}\left[\sqrt{a^2-l_1^2}-c\sin^{-1}\left(\frac{l_1}{r}\right)\right] \qquad (A4)$$

for $\mathrm{Re}(c)>|\mathrm{Im}(a\pm b)|$, where

$$l_1=\tfrac{1}{2}\left[\sqrt{(r+a)^2+c^2}-\sqrt{(r-a)^2+c^2}\right] \qquad (A5)$$

$$l_2=\tfrac{1}{2}\left[\sqrt{(r+a)^2+c^2}+\sqrt{(r-a)^2+c^2}\right] \qquad (A6)$$

It is noted that the first two have been given in Fabrikant [39], while the last two are derived as follows.

Multiplying (*A*2) by r, then differentiating with respect to r, and considering the known result

$$\frac{\partial}{\partial r}[rJ_1(\xi r)]=\xi rJ_0(\xi r) \qquad (A7)$$

we get

$$\int_0^\infty e^{-c\xi}J_0(\xi r)J_{3/2}(\xi a)\sqrt{\xi}\,d\xi=\frac{2}{\sqrt{2\pi}a^{3/2}}\left[\sin^{-1}\left(\frac{l_1}{r}\right)-\frac{l_1}{r\sqrt{r^2-l_1^2}}\left(r-l_1\frac{\partial l_1}{\partial r}\right)\right] \qquad (A8)$$

Furthermore, upon substitution of the following results [37]:

$$\frac{\partial l_1}{\partial r}=\frac{a(l_2^2-r^2)}{l_2(l_2^2-l_1^2)}, \quad l_1l_2=ar \qquad (A9)$$

$$\sqrt{r^2-l_1^2}\sqrt{l_2^2-r^2}=cr, \quad \sqrt{l_2^2-a^2}\sqrt{l_2^2-r^2}=cl_2 \qquad (A10)$$

into (*A*8), after some manipulations one can obtain (*A*3).

In order to derive (*A*-4), putting the known identity

$$J_{3/2}(\xi a)=-\frac{\sqrt{\xi}}{a}\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{J_{1/2}(\xi a)}{\sqrt{\xi}}\right) \tag{A11}$$

into the left-hand side of (*A*4) and performing integration by parts, we find

$$\begin{aligned}\int_0^\infty & e^{-c\xi}J_0(\xi r)J_{3/2}(\xi a)\,\frac{\mathrm{d}\xi}{\sqrt{\xi}}\\ &=\frac{1}{a}\left[\sqrt{\frac{2a}{\pi}}-c\int_0^\infty e^{-c\xi}J_0(\xi r)J_{1/2}(\xi a)\,\frac{\mathrm{d}\xi}{\sqrt{\xi}}\right.\\ &\quad\left.-r\int_0^\infty e^{-c\xi}J_1(\xi r)J_{1/2}(\xi a)\,\frac{\mathrm{d}\xi}{\sqrt{\xi}}\right]\end{aligned} \tag{A12}$$

With the aid of the known results [39]

$$\begin{aligned}\int_0^\infty e^{-c\xi}J_1(\xi r)J_{1/2}(\xi a)\,\frac{\mathrm{d}\xi}{\xi\sqrt{\xi}}&=\frac{\sqrt{2}}{\sqrt{\pi a r}}\left[\frac{l_1}{2}\sqrt{r^2-l_1^2}+\frac{r^2}{2}\sin^{-1}\left(\frac{l_1}{r}\right)\right.\\ &\quad\left.+c(\sqrt{a^2-l_1^2}-a)\right]\end{aligned} \tag{A13}$$

differentiation of the above equality with respect to *c*, considering (*A*10), yields

$$\int_0^\infty e^{-c\xi}J_1(\xi r)J_{1/2}(\xi a)\,\frac{\mathrm{d}\xi}{\sqrt{\xi}}=\frac{\sqrt{2}}{\sqrt{\pi a r}}(a-\sqrt{a^2-l_1^2}) \tag{A14}$$

On the other hand, we multiply (*A*13) by *r* and then differentiate with respect to *r*. Employing the relations in (*A*9) and (*A*10), together with

$$\sqrt{a^2-l_1^2}\sqrt{r^2-l_1^2}=cl_1 \tag{A15}$$

we derive

$$\int_0^\infty e^{-c\xi}J_0(\xi r)J_{1/2}(\xi a)\,\frac{\mathrm{d}\xi}{\sqrt{\xi}}=\frac{\sqrt{2}}{\sqrt{\pi a}}\sin^{-1}\left(\frac{l_1}{r}\right) \tag{A16}$$

As a consequence, substituting (*A*14) and (*A*16) into (*A*12) yields the desired result (*A*-4).

The constants appearing in (10) are respectively

$$\begin{aligned}a_0&=c_{44}(c_{33}\varepsilon_{33}+e_{33}^2)\\ b_0&=-c_{33}c_{44}\varepsilon_{11}+c_{13}^2\varepsilon_{33}-c_{11}c_{33}\varepsilon_{33}+2c_{13}c_{44}\varepsilon_{33}-c_{33}e_{15}^2-c_{11}e_{33}^2\\ &\quad-2c_{33}e_{31}e_{15}-c_{33}e_{31}^2+2c_{13}e_{33}e_{15}+2c_{13}e_{33}e_{31}+2c_{44}e_{33}e_{31}\\ c_0&=-c_{13}^2\varepsilon_{11}+c_{11}c_{33}\varepsilon_{11}-2c_{13}c_{44}\varepsilon_{11}+c_{11}c_{44}\varepsilon_{33}-2c_{13}e_{15}^2\\ &\quad-2c_{13}e_{15}e_{31}+c_{44}e_{31}^2+2c_{11}e_{15}e_{33}\\ d_0&=-c_{11}(c_{44}\varepsilon_{11}+e_{15}^2)\end{aligned} \tag{A17}$$

The constants appearing in (13) are respectively

$$\begin{aligned}\beta_{0j}&=(c_{13}\eta_{3j}+e_{31}\eta_{4j})\gamma_j^2-c_{11}\\ \beta_{1j}&=(c_{33}\eta_{3j}+e_{33}\eta_{4j})\gamma_j^2-c_{13}\\ \beta_{2j}&=[c_{44}(1+\eta_{3j})+e_{15}\eta_{4j}]\gamma_j\\ \beta_{3j}&=[e_{15}(1+\eta_{3j})-\varepsilon_{11}\eta_{4j}]\gamma_j\\ \beta_{4j}&=(e_{33}\eta_{3j}-\varepsilon_{33}\eta_{4j})\gamma_j^2-e_{31}\end{aligned} \tag{A18}$$

References

[1] Rao, S. S., and Sunar, M., 1994, "Piezoelectricity and its Use in Disturbance Sensing and Control of Flexible Structures: A Survey," Appl. Mech. Rev., **47**, pp. 113–123.

[2] Suo, Z., Kuo, C.-M., Barnett, D. M., and Willis, J. R., 1992, "Fracture Mechanics for Piezo-Electric Ceramics," J. Mech. Phys. Solids, **40**, pp. 739–765.

[3] Pak, Y. E., 1990, "Crack Extension Force in a Piezoelectric Material," ASME J. Appl. Mech., **57**, pp. 647–653.

[4] Pak, Y. E., 1992, "Linear Electroelastic Fracture Mechanics of Piezoelectric Materials," Int. J. Fract., **54**, pp. 79–100.

[5] Dunn, M. L., 1994, "The Effects of Crack Face Boundary Conditions on the Fracture Mechanics of Piezoelectric Solids," Eng. Fract. Mech., **48**, pp. 25–39.

[6] Gao, H., Zhang, T.-Y., and Tong, P., 1997, "Local and Global Energy Release Rates for an Electrically Yielded Crack in a Piezoelectric Ceramic," J. Mech. Phys. Solids, **45**, pp. 491–510.

[7] Sosa, H., and Khutoryansky, N., 1996, "New Developments Concerning Piezoelectric Materials With Defects," Int. J. Solids Struct., **33**, pp. 3399–3414.

[8] Shindo, Y., Tanaka, K., and Narita, F., 1997, "Singular Stress and Electric Fields of a Piezoelectric Ceramic Strip With a Finite Crack Under Longitudinal Shear," Acta Mech., **120**, pp. 31–45.

[9] Zhang, T.-Y., Qian, C.-F., and Tong, P., 1998, "Linear Electroelastic Analysis of a Cavity or a Crack in a Piezoelectric Material," Int. J. Solids Struct., **35**, pp. 2121–2149.

[10] Ru, C. Q., 1999, "Electric-Field Induced Crack Closure in Linear Piezoelectric Media," Acta Mater., **47**, pp. 4683–4693.

[11] McMeeking, R. M., 2001, "Towards a Fracture Mechanics for Brittle Piezoelectric and Dielectric Materials," Int. J. Fract., **108**, pp. 25–41.

[12] Yang, F., 2001, "Fracture Mechanics for a Mode I Crack in Piezoelectric Materials," Int. J. Solids Struct., **38**, pp. 3813–3830.

[13] Liu, M., and Hsia, K. J., 2003, "Interfacial Cracks Between Piezoelectric and Elastic Materials Under In-Plane Electric Loading," J. Mech. Phys. Solids, **51**, pp. 921–944.

[14] Hao, T. H., and Shen, Z. Y., 1994, "A New Electric Boundary Condition of Electric Fracture Mechanics and Its Applications," Eng. Fract. Mech., **47**, pp. 793–802.

[15] McMeeking, R. M., 1999, "Crack Tip Energy Release Rate For a Piezoelectric Compact Tension Specimen," Eng. Fract. Mech., **64**, pp. 217–244.

[16] Schneider, G. A., Felten, F., and McMeeking, R. M., 2003, "The Electrical Potential Difference Across Cracks in PZT Measured by Kelvin Probe Microscopy and the Implications for Fracture," Acta Mater., **51**, pp. 2235–2241.

[17] Zhang, T.-Y., Zhao, M., and Tong, P., 2002, "Fracture of Piezoelectric Ceramics," Adv. Appl. Mech., **38**, pp. 147–289.

[18] Hao, T.-H., 2001, "Multiple Collinear Cracks in a Piezoelectric Material," Int. J. Solids Struct., **38**, pp. 9201–9208.

[19] Liu, B., Fang, D.-N., Soh, A. K., and Hwang, K.-C., 2001, "An Approach for Analysis of Poled/Depolarized Piezoelectric Materials With a Crack," Int. J. Fract., **111**, pp. 395–407.

[20] Xu, X.-L., and Rajapakse, R. K. N. D., 2001, "On a Plane Crack in Piezoelectric Solids," Int. J. Solids Struct., **38**, pp. 7643–7658.

[21] Wang, X. D., and Jiang, L. Y., 2002, "Fracture Behavior of Cracks in Piezoelectric Media With Electromechanically Coupled Boundary Conditions," Proc. R. Soc. London, Ser. A, **458**, pp. 2545–2560.

[22] Wang, B. L., and Mai, Y.-W., 2003, "On the Electrical Boundary Conditions on the Crack Surfaces in Piezoelectric Ceramics," Int. J. Eng. Sci., **41**, pp. 633–652.

[23] Wang, B., 1992, "Three-Dimensional Analysis of a Flat Elliptical Crack in a Piezoelectric Material," Int. J. Eng. Sci., **30**, pp. 781–791.

[24] Wang, Z. K., and Zheng, B. L., 1995, "The General Solution of Three-Dimensional Problems in Piezoelectric Media," Int. J. Solids Struct., **32**, pp. 105–115.

[25] Kogan, L., Hui, C. Y., and Molcov, V., 1996, "Stress and Induction Field of a Spheroidal Inclusion of a Penny-Shaped Crack in a Transversely Isotropic Piezoelectric Material," Int. J. Solids Struct., **33**, pp. 2719–2737.

[26] Zhao, M. H., Shen, Y. P., Liu, Y. J., and Liu, G. N., 1997, "Isolated Crack in Three-Dimensional Piezoelectric Solid: Part I-Solution by Hankel Transform," Theor. Appl. Fract. Mech., **26**, pp. 129–139.

[27] Chen, W. Q., and Shioya, T., 1999, "Fundamental Solution for a Penny-Shaped Crack in a Piezoelectric Medium," J. Mech. Phys. Solids, **47**, pp. 1459–1475.

[28] Karapetian, E., Sevostianov, I., and Kachanov, M., 2000, "Penny-Shaped and Half-Plane Cracks in a Transversely Isotropic Piezoelectric Solid Under Arbitrary Loading," Arch. Appl. Mech., **70**, pp. 201–229.

[29] Jiang, L. Z., and Sun, C. T., 2001, "Analysis of Indentation Cracking in Piezoceramics," Int. J. Solids Struct., **38**, pp. 1903–1918.

[30] Yang, J. H., and Lee, K. Y., 2001, "Penny Shaped Crack in Three-Dimensional Piezoelectric Strip Under In-Plane Normal Loadings," Acta Mech., **148**, pp. 187–197.

[31] Wang, B.-L., Noda, N., Han, J.-C., and Du, S.-Y., 2001, "A Penny-Shaped Crack in a Transversely Isotropic Piezoelectric Layer," Eur. J. Mech. A/Solids, **20**, pp. 997–1005.

[32] Lin, S., Narita, F., and Shindo, Y., 2003, "Electroelastic Analysis of a Penny-Shaped Crack in a Piezoelectric Ceramic Under Mode I Loading," Mech. Res. Commun., **30**, pp. 371–386.

[33] Yang, J. H., and Lee, K. Y., 2003, "Penny Shaped Crack in a Piezoelectric Cylinder Surrounded by an Elastic Medium Subjected to Combined In-Plane Mechanical and Electrical Loads," Int. J. Solids Struct., **40**, pp. 573–590.

[34] Ou, Z. C., and Chen, Y. H., 2003, "Discussion of the Crack Face Electric

Boundary Condition in Piezoelectric Fracture Mechanics,'' Int. J. Fract., **123**, pp. L151–L155.

[35] Sneddon, I. N., 1966, *Mixed Boundary Value Problems in Potential Theory*, North-Holland, Amsterdam.

[36] Sneddon, I. N., and Lowengrub, M., 1969, *Crack Problems in the Classical Theory of Elasticity*, Wiley, New York.

[37] Fabrikant, V. I., 1991, *Mixed Boundary Value Problems of Potential Theory and Their Applications in Engineering*, Kluwer Academic Publishers, Dordrecht.

[38] Park, S., and Sun, C. T., 1995, ''Fracture Criteria for Piezoelectric Ceramics,'' J. Am. Ceram. Soc., **78**, pp. 1475–1480.

[39] Fabrikant, V. I., 2003, ''Computation of Infinite Integrals Involving Three Bessel Functions by Introduction of New Formalism,'' Z. Angew. Math. Mech., **83**, pp. 363–374.

T. A. Godfrey
Research Engineer,
Natick Soldier Center,
U.S. Army Research,
Development & Engineering Command,
Natick, MA 01760-5020
Mem. ASME

J. N. Rossettos
Professor,
Department of Mechanical, Industrial &
Manufacturing Engineering,
Northeastern University,
Boston, MA 02115-5096
Fellow ASME

S. E. Bosselman
Research Engineer,
Natick Soldier Center,
U.S. Army Research,
Development & Engineering Command,
Natick, MA 01760-5020

The Onset of Tearing at Slits in Stressed Coated Plain Weave Fabrics

A simple micromechanical model is presented for predicting the onset of tearing at slitlike damage sites in biaxially stressed coated plain weave fabrics. The stress concentration in the first intact yarn at the slit tip is determined as a function of increasing loading, and predictions for the onset of tearing are made under the assumption that tearing initiates through the rupture of the first intact yarn at a characteristic yarn breaking load. Extensive onset of tearing experiments on various coated nylon and polyester fabrics are presented, and the model is shown to capture the onset of tearing in these fabrics quite well over a range of slit lengths. [DOI: 10.1115/1.1794165]

Introduction

Coated woven fabrics are used in a variety of inflatable and tension structures. An emerging application is the use of coated fabrics to create pressure-stabilized (air-inflated) fabric beam and arch structures, commonly referred to as *air-beams*. While the propagation of tears in coated fabric structures has long been a concern, the issue is particularly important in air-beams, where rapid tear propagation initiating at sites of local damage or stress concentration is accompanied by a hazardous explosive release of compressed air. An improved fundamental understanding of the mechanics of damage in coated woven fabrics is needed to provide a rational basis for efficient, damage-tolerant air-beam materials and designs.

Hedgepeth [1] provided the first micromechanical analysis of a damaged filamentary structure. His analysis, based on shear-lag theory, has been applied to fiber/matrix composites, where the matrix transfers the load from broken to unbroken fibers by means of shear. Hedgepeth regarded his model to apply to coated woven fabrics as well, where the coating transfers shear between yarns. Literature experimental results on slit-damaged coated fabrics [2–4] suggest that Hedgepeth's analysis underestimates the remote stress needed to initiate tearing—fabrics are more damage tolerant than predicted by the Hedgepeth model. Inelastic effects, not considered in Ref. [1], may partially explain the discrepancy between the literature experiments and the analysis. The model presented here, while having a similar mathematical structure to Ref. [1], includes a type of inelastic deformation near the slit tip. In this paper, we present the onset of tearing experiments for a variety of coated polyester and nylon fabrics that agree well with the model.

For an extensive review of the literature on mechanics of damage in stressed fabrics, the reader is directed to Ref. [5]. Here we will highlight some recent works not discussed in the prior review. Szostkiewicz and Hamelin [6] observed and analyzed failure mechanisms of a PVC-coated plain-woven polyester fabric with a variety of slit-damage configurations under displacement-controlled uniaxial and biaxial loading conditions. They reported a number of interesting findings; however, their work did not include any theoretical modeling. Experiments on the same polyester fabric with and without the PVC coating demonstrated that the coating significantly lowers the threshold load for the onset of tearing. Davidson et al. [7] consider the role of woven fabric behavior in developing fracture toughness in fabric-reinforced composites. They investigated the ‘‘crack-tip’’ micromechanics of a common plain-woven apparel fabric through optical displacement measurements and finite element modeling. The finite element model represented the interlaced yarns of the weave by a network of pin-jointed beams. When frictional slip between interlaced yarns at crossover points was included in the model through spring elements that allowed relative displacement between the beams at the pinned joints [8], a good fit with the experimental measurements of fabric strains near the crack tip was obtained. In fabric-reinforced composites, Davidson et al. [7] consider the matrix to play a role in limiting the possible fracture toughness of the composite by inhibiting the potential for frictional slip between yarns near the crack tip. This idea is supported by the results mentioned above [6] on fabrics with and without coatings (a coated fabric is essentially fabric-reinforced composite with an elastomeric matrix) and by the present work. In this paper, inelastic deformation involving slip between yarns and permanent deformation of the coating, is key to reducing the stress concentration in the yarn at the slit tip and, therefore, inhibiting the onset of tearing failure.

While the major emphasis in this paper is the comparison of the present model with experiments on coated fabrics, the behavior of the coated-fabric model will also be discussed in the context of shear-lag models for notched-fiber composite sheets where matrix yielding plays a role similar to the yarn-slip mechanism encountered in fabrics [9]. Such models, outgrowths of Hedgepeth's original two-dimensional (2D) shear-lag model [1], have been the subject of extensive study. The onset of inelastic deformation, its growth under increasing load, and the attendant affect on stress concentrations at the notch tip will be briefly discussed in both material systems. This framework provides further insight into the coated-fabric model by pointing out some common aspects of the

---

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 24, 2003; final revision, April 2, 2004. Associate Editor: D. A. Kouris. 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.

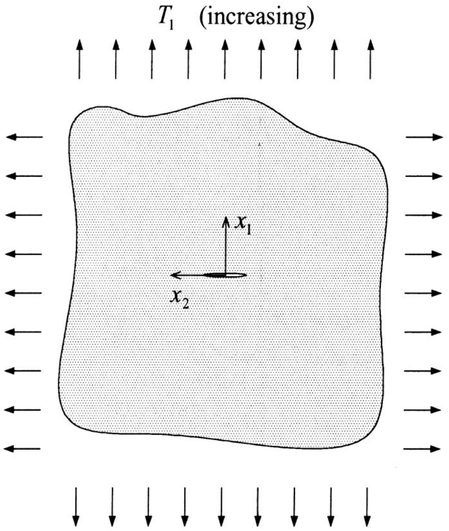


Fig. 1 Configuration of biaxial remote stresses on damaged fabric

behavior of notched-sheet systems and highlighting behavioral features that are unique to the coated-fabric model system.

Micromechanical Model. We adopt a micromechanical modeling approach originally developed in [10]. Essential aspects of the development are summarized here. Consider a coated plain-weave fabric with damage consisting of a slit, parallel to the x_2 coordinate direction at $x_1=0$, where the x_1x_2 coordinate system is aligned with the yarn directions. The slit interrupts the coating and a series of consecutive number one (#1) yarns, referring to yarns parallel to the x_1 and x_2 directions as #1 and number two (#2) yarns, respectively. The plain-weave-unit cell dimensions are y_{01} along the x_1 axis (the spacing of the #2 yarns) and y_{02} along the x_2 axis (the spacing of the #1 yarns). Remote biaxial membrane stresses (i.e., having dimensions of force/length) are applied to the fabric such that the stress in the x_2 direction, T_2, is held constant while the stress in the x_1 direction, T_1, is increased quasistatically. The global configuration of the damage and remote loading is shown in Fig. 1. We assume that the elastomeric coating carries only shear so that the direct membrane stresses are supported entirely by tension in the yarns. At the microstructural level, therefore, the membrane stresses are viewed as individual remote yarn tensions such that #2 yarns are under constant remote tensions F_2^* and #1 yarns are under quasistatically increasing remote tensions p (i.e., $T_1=p/y_{02}$, $T_2=F_2^*/y_{01}$). As p increases, the #1 yarns exhibit displacements in the x_1 direction, and the #2 yarns exhibit x_1 direction displacements and small rotations in the fabric plane. With continued increasing loading, a small amount of permanent deformation takes place in a region along the first intact #1 yarns at the slit tips. When p reaches a critical value p_c, tearing begins through the rupture of one of the first intact #1 yarns at either end of the slit. In actual fabric testing, tearing may be catastrophic, where the first yarn rupture coincides with rapid propagation of the tear, or progressive, depending on the particular fabric and test conditions. In this work it is assumed that rupture occurs when the maximum tension in the first intact #1 yarns reaches the yarn ultimate breaking load p_u.

Appropriate differential equations, which describe the equilibrium of yarns in regions where inelastic behavior occurs and in the region where inelastic behavior does not occur, can easily be derived. For instance, in the elastic region, equilibrium of the #1 yarns can be derived by taking into account the load transfer to the #1 yarns that occurs due to the rotated tensioned #2 yarns in the fabric plane and the additional load transfer due to shear of the coating, assuming the coating does not wrinkle. For small rotations, the angles are indicated in Fig. 2(*b*). The component of the #2 yarn tension along the #1 yarn can then be written.

Introduce u_n^j as the x_1 displacement of the *j*th crossover point on the *n*th #1 yarn, where the reference state for displacements are the positions of points on an otherwise identical stressed fabric *without damage*. For sufficiently high values of the loading *p*, the #1 yarns are assumed to be in a nearly straightened out condition (i.e., the out-of-plane ‘‘crimp’’ due to interlacing with the #2 yarns has been taken out by displacement along the x_1 direction, see Ref. [11]), and display an effective constant axial stiffness property $(EA)_{\text{eff}}$ having the dimension of force. In-plane rotations of the #1 yarns may be neglected, and the #1 yarn load-displacement relation is

$$F_{1n}^j = EA_{\text{eff}}\frac{u_n^j - u_n^{j-1}}{y_{01}} + p \tag{1}$$

where the factor $(u_n^j - u_n^{j-1})/y_{01}$ is the additional strain in the yarn due to the damage, and p accounts for the yarn load in the reference state. Rotation of the #2 yarns is represented by relative displacements at points on adjacent #1 yarns. Considering the crossover point unit cell as a free body, Fig. 2(*b*), the component of force in the x_1 direction acting on the #2 yarn entry and exit boundaries will be $F_2^*(u_{n-1}^j - u_n^j)/y_{02}$ and $-F_2^*(u_n^j - u_{n+1}^j)/y_{02}$. Shear in the matrix bays between the *n*th and neighboring $(n-1)$th and $(n+1)$th yarns gives rise to force components $Ghy_{01}(u_{n-1}^j - u_n^j)/y_{02}$ and $-Ghy_{01}(u_n^j - u_{n+1}^j)/y_{02}$, where Gh is the shear stiffness of the coating layer (dimension of force/length), y_{01} is the unit cell length, and shear strains in the coating bays between #1 yarns are identical to the #2 yarn rotations. Therefore, referring to Fig. 2(*b*), equilibrium of the *j*th crossover unit cell in the x_1 direction is written as

$$\frac{EA_{\text{eff}}}{y_{01}}(u_n^{j-1} - 2u_n^j + u_n^{j+1}) + \frac{Ghy_{01}+F_2^*}{y_{02}}(u_{n-1}^j - 2u_n^j + u_{n+1}^j) = 0 \tag{2}$$

where Eq. (1) has been used for the #1 yarn force on either side of the crossover point, i.e., the *j*th and $(j+1)$th F_{1n} terms. Regarding u_n as a continuous function of position x_1, this equilibrium equation can then be written as

$$\frac{d^2u_n}{dx_1^2} + \frac{Gh+F_2^*/y_{01}}{EA_{\text{eff}}y_{02}}(u_{n-1} - 2u_n + u_{n+1}) = 0 \tag{3}$$

Equation (3) is written in dimensionless form as

$$U_n'' + U_{n-1} - 2U_n + U_{n+1} = 0 \tag{4}$$

using the following nondimensionalization:

$$x_1 = \sqrt{\frac{EA_{\text{eff}}y_{02}}{F_2^*/y_{01}+Gh}}\,\xi, \quad u_n = p\sqrt{\frac{y_{02}}{EA_{\text{eff}}\{F_2^*/y_{01}+Gh\}}}\,U_n \tag{5}$$

where primes denote differentiation with respect to ξ.

Regarding F_{1n} as a function of position, introduce a nondimensional yarn load P_n, defined by

$$F_{1n} = pP_n \tag{6}$$

Reforming the #1 yarn load-displacement relation, Eq. (1), into a continuous form [i.e., replacing $(u_n^j - u_n^{j-1})/y_{01}$ with du_n/dx_1] and nondimensionalizing, using Eq. (5), gives the nondimensional load-displacement relation,

$$P_n = U_n' + 1 \tag{7}$$

where we have used the definition given by Eq. (6).

Experiments and analytical modeling of uncoated plain-weave fabrics have shown that, for moderate-sized slits, as p increases, crossover point yarn slip occurs predominantly along the first intact #1 yarn [e.g., yarn 2 in Fig. 2(*a*)] at the slit tip [12]. Crossover

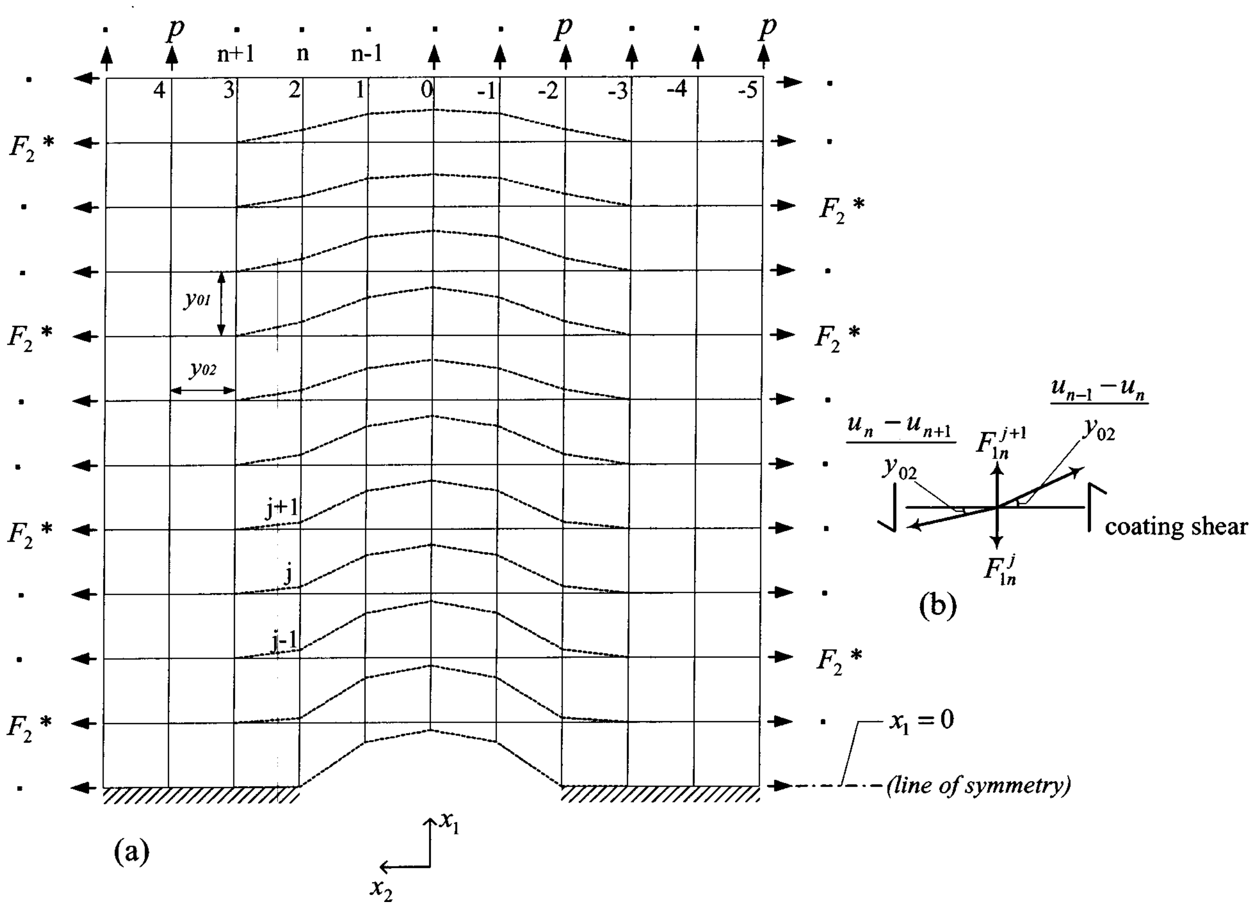


Fig. 2 (*a*). Geometry of damaged fabric indicating the elastic deformation. Breaks in #1 yarns, dashed lines represent deformed #2 yarns. (*b*). Equilibrium of the *j*th cross-over point unit cell under forces due to rotation of the tensioned #2 yarns and shear in the coating.

point forces between the #1 and #2 yarns are a maximum along the first intact-yarn near $x_1 = 0$, therefore it is in this region that the forces first attain a threshold value necessary for frictional slip to occur between the interlaced yarns. In coated fabrics, the coating layer acts to bond the interlaced yarns together and inhibit crossover point slip. As in uncoated fabrics, crossover point forces between the interlaced #1 and #2 yarns will be greatest along the first intact #1 yarn near $x_1 = 0$. Therefore, assuming that the coated fabric cross-over point has some characteristic strength, we expect that crossover point slip or inelastic relative displacement will occur preferentially along the first intact yarn. The inelastic relative displacement between the #1 and #2 yarns involves yielding or separation of the coating, as depicted speculatively in Fig. 3.

In the present model, we assume that inelastic yarn-to-yarn relative displacement occurs in a region along the first intact #1 yarn near the slit line, $0 < x_1 < l$, where l denotes the extent of the inelastic zone. The inelastic deformation is assumed to occur at a constant force of f (perfect plasticity) acting in the $+x_1$ direction on the #1 yarn at the cross-over point. Note that possible wrinkling will have the same effect. We interpret f as an average crossover point force. For fabrics coated on one side only, the actual crossover point slip forces alternate between small values and nearly $2f$ and these forces are regarded as having an average value of f per crossover point. Note that in the development of the equilibrium equation, Eq. (8), the inelastic force will be averaged over the crossover point unit cell to a force per length, i.e., f/y_{01}.

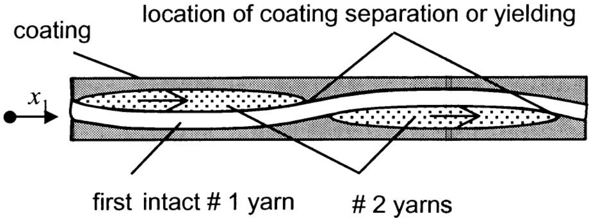


Fig. 3 Cross-section showing first intact #1 yarn at slit tip and location of possible separation or yielding of coating. Arrows indicate motion of #2 yarns.

We consider a finite-width configuration of $2q+1$ #1 yarns with a slit interrupting $2N+1$ yarns centered at the zeroth yarn. Yarns in the positive x_2 half-plane are numbered so that n equals 1 to q. Yarns in the negative x_2 half-plane are numbered -1 to $-q$. The slit is symmetrical about the center yarn, so that, concerning ourselves only with non-negative n, yarns numbered $0 \leq n \leq N$ are severed by the slit, and yarns numbered $n > N$ are intact. The $(N+1)$th yarn is the first intact #1 yarn at the slit tip. Because of symmetry, we need only consider equations for yarns $n=0$ to $n=q$.

Following the previous discussion of inelastic behavior, equilibrium of the $(N+1)$th yarn in the region $0 < x_1 < l$ may be written as

$$\frac{d^2 u_{N+1}}{dx_1^2} + \frac{f}{EA_{\text{eff}} y_{01}} = 0 \qquad (8)$$

where the inelastic force has been averaged over the cross-over point spacing. Equation (8) may be written in dimensionless form as

$$U''_{N+1}+\hat{f}=0 \tag{9}$$

where slip occurs in a region $0<\xi<\hat{l}$ and a dimensionless loading parameter and extent of the inelastic region are introduced,

$$f=p\sqrt{\frac{(F_2^*+Ghy_{01})y_{01}}{EA_{\text{eff}}y_{02}}}\,\hat{f} \tag{10a}$$

$$l=\sqrt{\frac{EA_{\text{eff}}y_{02}}{F_2^*/y_{01}+Gh}}\,\hat{l} \tag{10b}$$

Development of the equilibrium equations in the inelastic region for the nearest-neighbor yarns to the first intact yarn (i.e., yarns N and $N+2$), takes into account the rotation of the #2 yarns, which, in this region, is not coupled to u_{N+1} due to slip. Following a similar procedure to that in Ref. [12], the #2 yarns are treated as classical taut strings with x_1 displacement δ_{N+1} at the crossover points along the $(N+1)$th #1 yarn. Equilibrium of the #2 yarns is used to eliminate δ_{N+1} in the Nth and $(N+2)$th yarn equations. Finally, the equilibrium equations in the region $0<\xi<\hat{l}$ for the Nth and $(N+2)$th yarns are written in dimensionless form as

$$U''_N+U_{N-1}-\frac{3}{2}U_N+\frac{1}{2}U_{N+2}=\frac{\hat{f}}{2} \tag{11a}$$

$$U''_{N+2}+\frac{1}{2}U_N-\frac{3}{2}U_{N+2}+U_{N+3}=\frac{\hat{f}}{2} \tag{11b}$$

The width of the model, specified by the value of q, is assumed to be sufficiently large such that the behavior of the finite width configuration closely approximates that of an infinite fabric with an isolated slit. Similar studies in composite sheets containing matrix yield zones near the slit tip [13,14], where the structure of the equations is the same and the slip zone in fabrics plays an analogous role as the matrix yield zone in composites, have shown exponential decay in the width direction, so that finite width sheets provide effective models for the infinite sheet problems. Therefore, the displacements of yarn q are taken to be those of the undamaged reference, i.e., $U_q(\xi)=0$.

Symmetry about the center yarn and the above assumption regarding the qth yarn lead to the following special forms for yarns 0 and $q-1$, written as:

$$U''_0-2U_0+2U_1=0 \tag{12a}$$

$$U''_{q-1}+U_{q-2}-2U_{q-1}=0 \tag{12b}$$

The boundary value problem for a slit involving a given number of severed #1 yarns requires appropriate boundary conditions on the severed and intact yarns at the line of the slit, $\xi=0$, and at a remote distance from the slit, $\xi=\infty$. Since the broken yarn ends are stress free at the slit, the boundary condition on the broken yarns is $P_n(0)=0$, which using Eq. (7), gives $U'_n(0)=-1$. For intact yarns, symmetry requires that $U_n=0$ at $\xi=0$. Uniform yarn loads of p are assumed far from the breaks, therefore, for all yarns we require that $P_n=1$ and, therefore, $U'_n=0$ at $\xi=\infty$.

Since all yarns are continuous at $\xi=\hat{l}$, continuous yarn displacements and strains lead to continuity conditions on U_n and U'_n at $\xi=\hat{l}$. An additional continuity condition arises from the assumption that cross-over point forces required to initiate inelastic deformation at the boundary of the slipping region are approached in a continuous manner [10].

The system of equations for regions I, $0<\xi<\hat{l}$, and II, $\xi\geqslant\hat{l}$, are written in matrix form, and solutions in each region are obtained using an eigenvector expansion technique, as described in detail for a similar boundary value problem in [12]. The solution process is completed by selecting values of the inelastic zone extent $\hat{l}$ and determining the values of the integration constants and parameter $\hat{f}$, such that the boundary and continuity conditions are satisfied. Observing from Eq. (10a) that the parameter $\hat{f}$ is inversely proportional to the applied load p, we introduce a dimensionless applied load $\hat{p}$ that we define as $\hat{p}=\hat{f}^{-1}$. For selected values of the inelastic zone extent, solution of the defined boundary value problem provides the corresponding value of the applied load $\hat{p}$. A key result that can be easily computed from the solution is the stress concentration factor (SCF), defined here as the ratio of the maximum tension (which occurs at $x_1=0$) in the first intact yarn at the tip of the slit to the remote applied load, $F_{1\ N+1}(0)/p$. Using Eqs. (6) and (7) the SCF can be written as $U'_{N+1}(0)+1$.

As p increases, the value of $F_{1\ N+1}(0)$ increases also, to the point of rupture of the $(N+1)$th yarn, which we assume occurs when $F_{1\ N+1}(0)$ attains the value of the ultimate yarn breaking load p_u. Rupture of the first intact yarn is regarded as the onset of tearing. The applied load has the value p_c (*p critical*) at the instant of first yarn rupture, and we denote the value of the SCF at this instant as SCF_t, i.e., the *stress concentration factor at onset of tearing*. From our assumption regarding yarn rupture, we write SCF_t as $\text{SCF}_t=p_u/p_c$. We denote the value of dimensionless applied load at onset of tearing as $\hat{p}_c$; it can be evaluated as $\hat{p}_c=\hat{f}^{-1}|_{p=p_c}$.

Consider the behavior of a particular fabric over a range of slit lengths. For increasing slit lengths (involving greater numbers of severed yarns), the value of p_c will decrease and SCF_t will increase. The product $p_c\,\text{SCF}_t$ should remain approximately constant over the range of slit lengths, however, since $p_c\,\text{SCF}_t=p_u$ (the yarn breaking load), which is independent of the slit size. Nondimensionalizing this relationship, we introduce a parameter $\hat{p}_u=\hat{p}_c\,\text{SCF}_t$, which may be used to characterize the behavior of specific fabrics over a range of slit lengths. The parameter $\hat{p}_u$ may be regarded as a material property of the coated fabric, since it contains micromechanical constitutive parameters, such as yarn strength, yarn stiffness, unit cell dimensions, and the crossover point inelastic force. Results for SCF_t versus $\hat{p}_c$ (i.e., SCF versus $\hat{p}$ obtained from solution of the boundary value problem) are used to tabulate SCF_t versus $\hat{p}_u$ for slits involving various numbers of severed yarns. Selecting values of $\hat{p}_u$ and extracting from the tabulation the corresponding SCF_t for various slit sizes provides the behavior of specific fabrics versus slit length.

Effects of Inelastic Behavior—Comparison With a Fiber Composite Sheet Model

As previously mentioned, shear-lag models for fiber composite sheets (monolayers) containing a central notch have a similar mathematical structure to the present model for a slit-damaged coated fabric. For a comprehensive exposition, including both 2D and 3D shear lag models for fibrous composite materials, connection to statistical strength theories, and an extensive account of the literature, the reader is referred to Ref. [15]. A review with particular attention to 2D sheets is given by Rossettos and Godfrey [5].

A number of investigators have considered extensions to Hedgepeth's model incorporating local plastic deformation of the matrix. Here we refer to Beyerlein and Phoenix [9] (B-P) in particular. As in [1], the 2D composite consists of infinite sheet of equispaced parallel elastic fibers in an elastic matrix, where the fibers are parallel to the x axis, uniformly loaded at $x=\pm\infty$ by fiber loads p. The matrix is viewed as divided into discrete bays separating neighboring fibers. A central notch interrupts a series of consecutive fibers and the associated matrix bays along the line $x=0$. Fibers are assumed to deform in simple tension, and the matrix is assumed to deform in simple shear. It is assumed that the

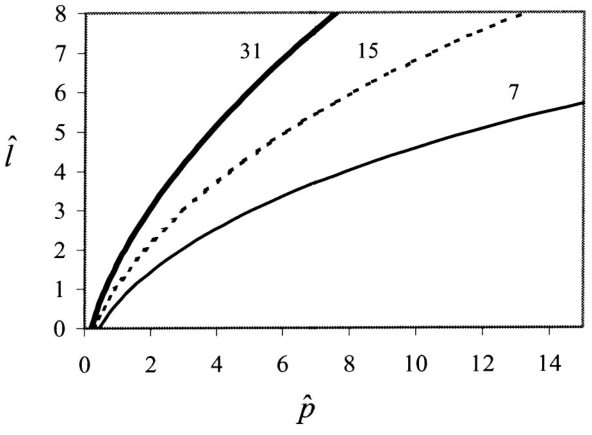


Fig. 4 Analytical results for inelastic zone extent, $\hat{l}$, with increasing applied load, $\hat{p}$, for slits involving seven, fifteen, and 31 severed yarns (# breaks indicated next to curves)

matrix yields in an elastic-perfectly plastic sense only in the matrix bay between the last fiber break and the first intact fiber at the tip of the notch.

Beyerlein and Phoenix [9] (B-P) provide extensive results for this configuration, including calculations for notch sizes (number of contiguous breaks) similar to the slit sizes considered in the present work. The nondimensional load P defined by B-P, is essentially equivalent to our $\hat{p}$, where a matrix bay yield force/length $\tau_y h$ (yield stress times sheet thickness), may be regarded as playing the same role as a quantity f/y_{01} in our model. As the load is increased, the composite deforms elastically until a threshold is reached at which the matrix bay begins to yield at $x=0$. B-P give the applied load at first yield as $P_Y=0.33$ for 15 breaks and 0.225 for 31 breaks. For the coated fabric model, comparable nondimensional loads at the onset of inelastic deformation are found to be 0.294 and 0.204 for 15 and 31 breaks, respectively, which is slightly lower than in the fiber composite, but very similar in dependence on number of breaks, with increasing slit size leading to decreasing loads at onset of inelasticity. In both material system models, an increasing load beyond the yield/inelastic threshold leads to the growth of a yielded region along the first intact fiber or yarn. The inelastic zone extent $\hat{l}$, with increasing applied load $\hat{p}$ for the coated fabric model is exhibited in Fig. 4. The inelastic zone extent is seen to grow at a higher rate with applied load for larger slit sizes. B-P find the same trend with the fiber composite model, however, as exhibited in Fig. 7 of [9], the extent of the yielded region increases approximately linearly with applied load, whereas growth is seen to be nonlinear in the coated fabric model (Fig. 4), which leads to smaller inelastic zone extents at similar load values. In the fiber composite model, the entire matrix bay yields in the region $0 \leqslant x \leqslant b$, where b is the dimensional yield region extent, corresponding to the nondimensional extent β. The yielded region, therefore, passes a total shear load of $\tau_y hb$ to the first intact fiber and the remaining intact sheet (at one notch tip in a quarter symmetric portion of the sheet). Since only a small, limited load can be transferred elastically, increasing the applied load beyond yield requires that the shear transferred through the yielded region $\tau_y hb$ increase in proportion to the applied load, leading to the nearly linear increase in β. In the coated fabric model, inelasticity limits load transfer to the first intact yarn through slip between the yarn and the coating/#2 yarns, but no gross yielding of the coating "bay" occurs. Therefore, load transfer to the remaining undamaged fabric is not as severely limited, and the inelastic zone extent grows at a rate less than linear with increasing applied load.

In both material systems, the SCF is found to decrease with increasing applied load, as load transfer tends to occur over a larger portion of the sheet surrounding the notch/slit, due to the increasing yield/inelastic region extent. Results for small slits of 5–13 breaks in the coated fabric were given in [10]. For a notch of 31 fiber breaks in the composite sheet, B-P give the SCF as starting at 5.03 for the elastic case, and falling to roughly 3.9, 3.5, and 3.0 for P values of 0.620, 0.820, and 1.160, respectively. Corresponding yield region extents are $\beta=3$, 5, and 9. Note that, for the elastic case (i.e., the yield/inelastic region extent is taken to be zero), both models reduce to Hedgepeth's original model [1], and the solution methods faithfully reproduce the SCF result given in [1]. For comparison, we find for 31 breaks in the coated fabric, the SCF falls from the elastic value (5.03) to 4.2, 3.95, and 3.6 at $\hat{p}$ values of 0.62, 0.82, and 1.16, respectively. The decrease in SCF is seen to be less steep in the coated fabric model due to the relatively restricted growth in inelastic extent (e.g., $\hat{l}=1.37$ at $\hat{p}=0.82$).

For the notched fiber composite sheet, B-P treat debonding in addition to yielding and provide a wealth of information on shear stress and shear strain distributions in the matrix, load profiles along fibers, stress concentrations on fibers ahead of the crack, relationships to damage region extent, and applied load and deal with the question of yielding in multiple adjacent matrix bays. A similar depth of study of the coated fabric model is beyond the present scope and will be the subject of a future paper. Phoenix and Beyerlein [15] connect the fiber composite sheet results to linear elastic fracture mechanics (LEFM) theory for an orthotropic planar continuum and provide additional theoretical results including a variety of approximations for fiber stress concentration factors, shear stress concentrations, etc., as well as the previously mentioned comprehensive treatment of statistical strength theory.

Experiments

ILC Dover provided samples of five coated fabrics for experimental investigation. One of the samples, a silicone coated plain weave Vectran fabric, could not be effectively tested using our apparatus and was dropped from the study. Characteristics of the remaining four fabrics are given in Table 1. The yarn size is given for fabric a and fabric *d* in the traditional textile unit of denier. A yarn's denier count is defined as the mass in grams for a length of the yarn measuring 9000 meters. Yarn deniers were not available for fabric *b* and *c*. Three additional fabrics were obtained from stock of the Natick Soldier Center's (NSC) tent prototype shop

Table 1 Sample fabrics furnished by ILC Dover

sample	coating	base fabric	construction No. warp yarns/cm×No. weft yarns/cm, yarn denier	specified strength warp, weft direction, N/cm
a	urethane, one side	nylon	26×20, 210	315, 300
b	urethane, both sides	nylon	16×16, -	220, 190
c	PVC, both sides	polyester	7.4×8.7, -	900, 900
d	urethane, one side	Kevlar	14×14, 400	790, 790

Table 2 Sample fabrics furnished by NSC tent shop

sample	application	coating	base fabric	construction No. warp yarns/cm×No. weft yarns/cm
e	Soldier Crew Tent Fly	heat set urethane, both sides	nylon	31×23
f	TEMPER tent fabric	PVC, both sides	polyester	18×13
g	tent liner fabric	-, one side	-	27×20

and are described in Table 2. All fabrics are plain woven with the exception of fabric *c*, which is a weft-inserted warp knit with a laid-in warp system. In such fabrics, warp and weft direction yarns are not interlaced, as in a woven, but are laid in straight and joined via a much lighter weight thread that knits the warps to the wefts at the crossover points, yielding a fabric where the yarns are closely packed and do not exhibit the out-of-plane crimp characteristic of woven fabrics. Areal densities were not obtained for any of the fabrics, however, fabric c, which is typical of a high performance architectural fabric, is significantly heavier with a much greater coating thickness than the other ILC Dover fabrics. The

Table 3 Yarn ultimate breaking loads p_u

sample	p_u, N	No. observations
a	8.97 (0.245)	4
b	14.4 (0.754)	6
c	117 (3.47)	4
d	76.9 (4.26)	4
e	15.2 (0.252)	8
f	22.4 (0.179)	4
g	5.41 (0.338)	4

Standard deviations shown in parentheses.

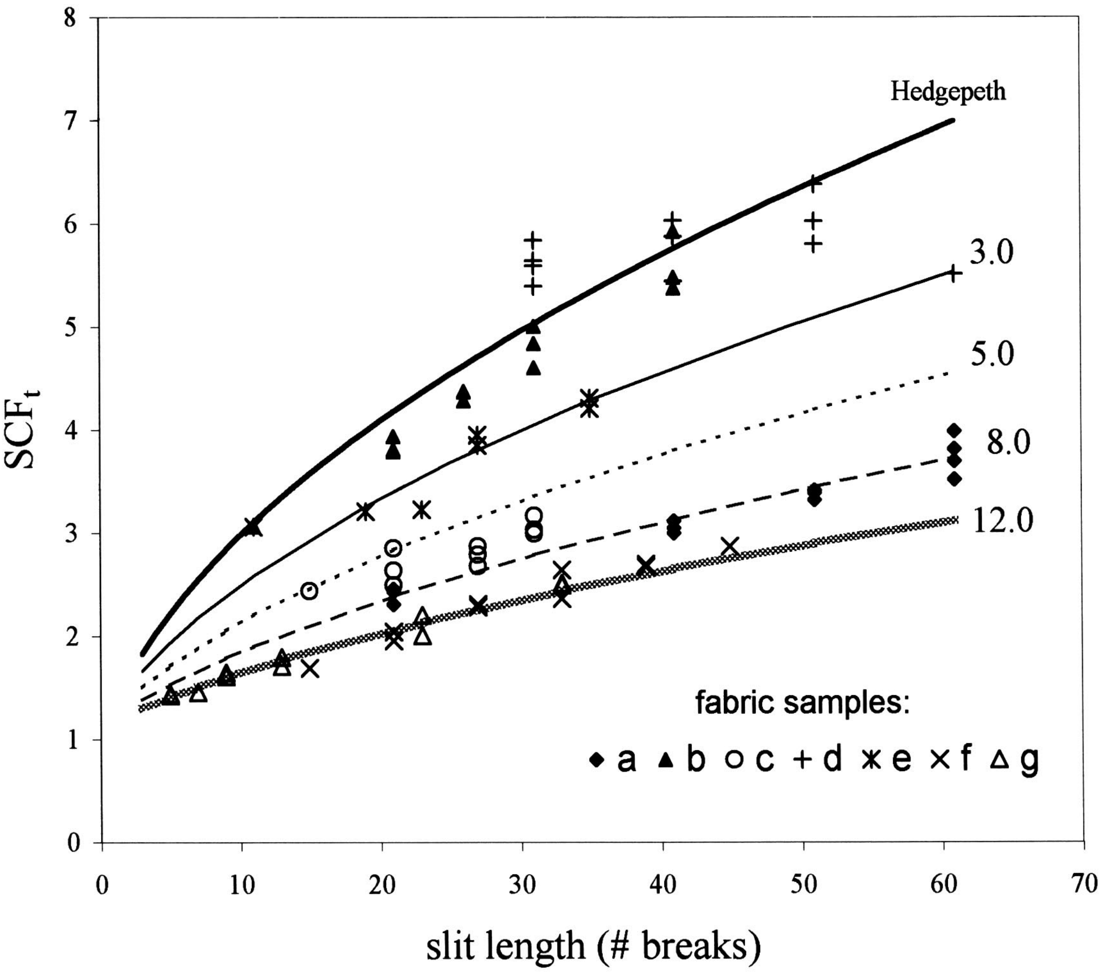


Fig. 5 Experimental (symbols) and analytical results (curves) for SCF$_t$ versus slit length. Top-most curve, labeled "Hedgepeth," shows SCF$_t$ from Ref. [1]. Remaining curves indicate results for various $\hat{p}_u$ (values indicated next to curves).

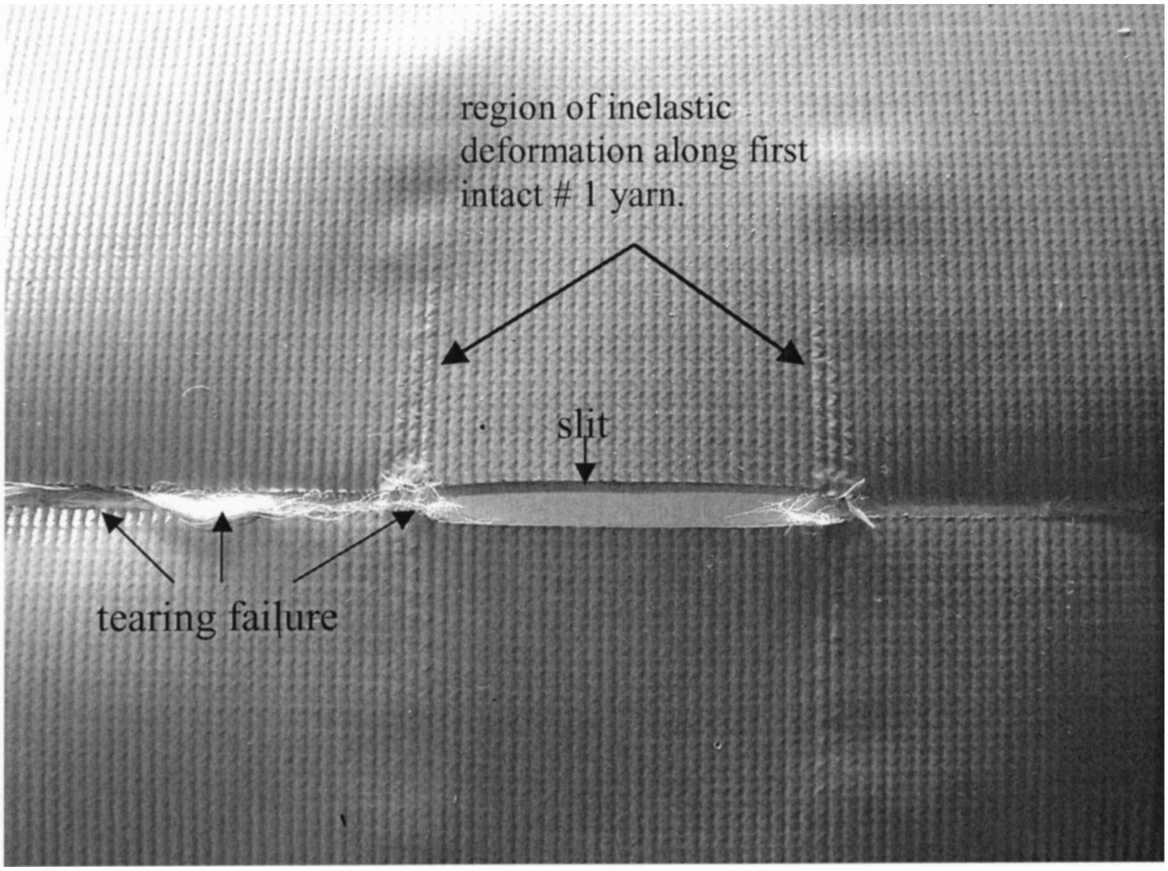


Fig. 6 Post-test condition of fabric c specimen

three military tent fabrics from the NSC shop include a lightweight tent fabric e, a medium-weight tent fabric f, and a tent liner fabric g.

Onset of tearing tests were performed on the sample fabrics using biaxially stressed initially slit cruciform specimens following a procedure similar to that described in Ref. [12]. The central stressed region of the specimen is 15 cm×15 cm and overall dimensions are nominally 36 cm top-to-bottom and 43 cm side-to-side. Each cross arm is 15 cm wide, and a 2.5 cm long section of the end of each arm is retained in the test fixture grips. The initial test slit is cut in the center of the central region with a sharp razor, running parallel to the side-to-side direction of the specimen. Typically, four "stress relief" slits are cut in the side cross arms, and two are cut in the top/bottom cross arms to mitigate the buildup of in-plane shearing stresses along the boundaries of the central region. The test setup uses a lightweight frame incorporating air cylinders to apply a constant tension in the x_2 direction (specimen side-to-side direction), while an increasing tension is applied in the x_1 direction using an Instron test machine. Usually the tests were performed under load control, however, the necessity to change test machines during the test program required that some of the samples be tested under displacement control conditions. Each sample fabric was tested over a range of initial slit lengths. The yarn ultimate breaking loads p_u exhibited in Table 3, were measured through tension tests on narrow strips of the coated fabrics using a loading rate approximately the same as that used in the onset of tearing tests. Generally the gage length for these tests was 5.08 cm, although fabric *c* and d required that capstan grips be used and involved a longer and less well-defined gage length. From four to eight observations were made for each fabric.

The failure process in some of the fabrics was sudden, with a rapid tear propagating across the full center section of the specimen and an associated sudden drop in the applied load. In other fabrics, a few yarns at the tip of the slit rupture while the load is still generally increasing on the specimen. The progression of failure is regarded as a complex dynamic event possibly involving a variety of factors: the response characteristics of the material, statistical aspects of yarn strength, specimen size and geometry, displacement versus load control, compliance of the test fixtures and test machine, etc. As such, investigation of failure progression is not an objective of the present study, and we are concerned only with the conditions leading to the onset of tearing failure at the slit damage site. For the tests where initial yarn ruptures occurred prior to maximum load, the specimen load-displacement curves were carefully studied to reveal the point at which a localized load drop occurs associated with the first yarn rupture at the slit tip. For tests involving sudden catastrophic tearing, the load associated with first yarn rupture coincides with the maximum load, and so the maximum load value is used in the calculation of p_c. Once the proper load value is identified, the value of p_c is determined by assuming the applied load on the specimen's gripped top/bottom cross-arm ends is distributed uniformly among the #1 yarns.

Comparison of Experimental and Analytical Results

The experimental and analytical results for the SCF_t $(=p_u/p_c)$ behavior of particular fabrics over a range of slit lengths are exhibited in Fig. 5. Each individual test result is plotted, where the SCF_t is calculated using the mean value of p_u (Table 3), and the particular value of p_c observed in that test. In some cases, only a single observation was made for a fabric at a particular slit length, though generally two to four replicate tests were made.

Analytical SCF_t curves are plotted in Fig. 5 for various constant values of $\hat{p}_u$, where the value is indicated next to the curve. As shown previously, $\hat{p}_u$ may be regarded as the single dimensionless parameter that governs the SCF_t versus slit-length behavior of particular fabrics. As such, experimental SCF_t—slit length data for individual fabric samples are expected to follow a constant $\hat{p}_u$ curve. Increasing values of $\hat{p}_u$ result in decreasing values of SCF_t for given slit lengths; since fabrics with lower SCF_t values can tolerate higher loading relative to yarn strength $\hat{p}_u$ may be regarded as a measure of a fabric's tolerance to slit damage. Godfrey and Rossettos [16] introduced a somewhat similar parameter as a screening measure for slit-damage tolerance in uncoated woven fabrics.

Hedgepeth's stress concentration factors [1] are exhibited by the curve marked ''Hedgepeth'' in the figure. As previously mentioned, if we assume an inelastic zone extent of length zero, the present model essentially reduces to Hedgepeth's, and the SCF calculated using the present method closely approximate those given by his result.

With the exception of fabric d, the experimental data are seen to be in good agreement with the analytical curves for constant $\hat{p}_u$. While the data for fabrics c and e are somewhat scattered, data for fabrics a, f, and g follow the analytical curves for constant $\hat{p}_u$ quite well. Results for fabric b lie close to the Hedgepeth curve, suggesting that deformation near the damage site in that fabric is primarily elastic up to onset of tearing. Consistent with results for coated fabrics in the literature [2–4], the slit-damaged fabrics sustain higher remote loadings than would be anticipated based on Hedgepeth's analysis (the experimental SCF_t are lower than the result given in [1]).

Results for fabric d do not exhibit the trend of increasing SCF_t with increasing slit size clearly shown in the analytical and other experimental results. Noting that the Kevlar yarns in fabric d are significantly stiffer than the nylon and polyester yarns in the other fabric samples, we investigated how the fabric's elastic properties affect the relationship between the actual experiments and the theoretical model. The model treats an isolated slit in an infinite fabric, whereas the experiment assumes that a 15 cm by 15 cm region of the specimen is *approximately* infinite. It turns out that, due to the high stiffness EA_{eff} of the Kevlar yarns, the characteristic length scale (i.e., the physical length along x_1 associated with a unit length along the dimensionless position coordinate ξ) in fabric d is about an order-of-magnitude larger than in the other fabric samples. Calculations made using the model showed that, for fabric d, the edge of the central region of the specimen remained well within the local region significantly affected by the presence of the slit for typical slit sizes. Therefore, a much larger specimen size is needed for a proper measurement of the SCF_t in fabric d, and the present experimental results for d should be disregarded. For fabrics with properties typical of the nylon and polyester samples, calculations showed that the central region was large enough to approximate an infinite fabric, and therefore, the specimen size used here is adequate.

Post-test examination of fabric c specimens supports the assumption made on the configuration of inelastic deformation. Figure 6 exhibits a typical c test specimen in the vicinity of the initial slit. The coating surface is seen to be disturbed along lines running some distance perpendicular from the slit tips. These lines follow the first intact #1 yarns. Although fabric c is not a plain weave, a similar mechanism is at work, whereby the coating yields or separates to permit relative motion between the warp and weft direction yarns. No clear evidence of coating disturbance could be detected in casual examination of the other fabric samples.

Conclusion

A simple micromechanical model for predicting the onset of tearing at slitlike damage sites in coated woven fabrics has been presented. A crucial aspect is the treatment of inelastic deformation involving yielding or separation of the coating and relative displacement between the interlaced yarns near the slit tip. Results for the onset and growth of inelastic deformation under increasing load and the associated decrease in stress concentrations at the slit tip have been discussed in connection with similar results for a notched fiber composite sheet model in the literature. Experimental results on a variety of coated nylon and polyester fabrics have been shown to corroborate the model and demonstrate the usefulness of a parameter $\hat{p}_u$ as a measure of slit-damage tolerance. The effect of experimental specimen size on obtaining realistic results consistent with the present analytical model was discussed.

Acknowledgments

We gratefully acknowledge the efforts of Robert Lingo, Thad Fredrickson, and David Zetune, ILC Dover, Inc. in selecting and supplying fabric samples for our experimental program. We thank Professor Julie Chen, University of Massachusetts-Lowell, for her interest and for her help in recruiting one of us (Bosselman) to assist in this research. The helpful suggestions of an anonymous referee that led to a significant improvement in the text are also acknowledged.

References

[1] Hedgepeth, J. M., 1961, ''Stress Concentrations in Filamentary Structures,'' NASA Technical Note D-882, NASA Langley Research Center, Langley Field, VA.
[2] Topping, A. D., 1973, ''The Critical Slit Length of Pressurized Coated Fabric Cylinders,'' J. Coated Fabrics, **3**, pp. 96–110.
[3] Ko, W. L., 1975, ''Fracture Behavior of a Nonlinear Woven Fabric Material,'' J. Compos. Mater., **9**, pp. 361–369.
[4] Wardle, M., 1978, ''Aramid Fibers for High Performance Coated Fabrics,'' J. Coated Fabrics, **7**, pp. 334–356.
[5] Rossettos, J. N., and Godfrey, T. A., 1998, ''Damage Analysis in Fiber Composite Sheets and Uncoated Woven Fabrics,'' Appl. Mech. Rev., **51**(6), pp. 373–385.
[6] Szostkiewicz, C., and Hamelin, P., 2000, ''Stiffness Identification and Tearing Analysis for Coated Membranes Under Biaxial Loads,'' J. Coated Fabrics, **30**(2), pp. 128–145.
[7] Davidson, D. L., Nicolella, D. P., and Spigel, B. S., 2000, ''Fracture Toughness of Fabrics as Composite Reinforcements,'' Exp. Mech., **40**(4), pp. 408–414.
[8] Davidson, D. L., and Nicolella, D. P., 1999, ''Fracture Micromechanics as Influenced by Environment in Textile Reinforced Ceramic Matrix Composites,'' Technical Report 18-7943-001, Southwest Research Institute, San Antonio, TX.
[9] Beyerlein, I. J., and Phoenix, S. L., 1996, ''Stress Concentrations Around Multiple Fiber Breaks in an Elastic Matrix With Local Yielding or Debonding Using Quadratic Influence Superposition,'' J. Mech. Phys. Solids, **44**, pp. 1997–2039.
[10] Godfrey, T. A., and Rossettos, J. N., 2002, ''A Micromechanical Model for Slit Damage in Coated Plain Weave Fabric,'' *Developments in Theoretical and Applied Mechanics, Vol. XXI* (Proc. of 21st SECTAM), A. J. Kassab, D. W. Nicholson, and I. Ionescu, eds., Rivercross Publishing, Orlando, pp. 315–324.
[11] Godfrey, T. A., and Rossettos, J. N., 1998, ''Damage Growth in Prestressed Plain Weave Fabrics,'' Text. Res. J., **68**(5), pp. 359–370.
[12] Godfrey, T. A., and Rossettos, J. N., 1999, ''The Onset of Tear Propagation at Slits in Stressed Uncoated Plain Weave Fabrics,'' ASME J. Appl. Mech., **66**(4), pp. 926–933.
[13] Rossettos, J. N., and Shishesaz, M., 1987, ''Stress Concentration in Fiber Composite Sheets Including Matrix Extension,'' ASME J. Appl. Mech., **54**, pp. 723–724.
[14] Rossettos, J. N., and Olia, M., 1995, ''On the Hybrid Effect and Matrix Yielding at Fibre Breaks in Hybrid Composite Sheets,'' Mech. Comp. Mater. Struct., **2**, pp. 275–280.
[15] Phoenix, S. L., and Beyerlein, I. J., 2000, ''Statistical Strength Theory for Fibrous Composite Materials,'' *Comprehensive Composite Materials*, A. Kelly, C. Zweben, and T. W. Chou, eds., Vol. 1, Pergamon, pp. 559–639.
[16] Godfrey, T. A., and Rossettos, J. N., 1999, ''A Parameter for Comparing the Damage Tolerance of Stressed Plain Weave Fabrics,'' Text. Res. J., **69**(7), pp. 503–511.

Concept and Fundamentals of Temporal-Spatial Pulse Representation for Dislocation Source Modeling

Ray Ruichong Zhang
Associate Professor
Division of Engineering,
Colorado School of Mines,
Golden, CO 80401
e-mail: rzhang@mines.edu

For far-field wave-motion response to a point dynamic dislocation source, the temporal and spatial features of the source mechanism are characterized, respectively, by two factors, i.e., a source time function for dislocation growth and a combination of nine couples of impulse forces that is equivalent to the final dislocation. The mathematical representation for each of the couples, referred to as spatial couples, is a couple of impulses acting in opposing directions with an infinitesimal separation distance or, in the limit, by the derivative of the impulse with respect to the separation-distance parameter. This study proposes a temporal-spatial pulse representation for the nine couples, referred to as temporal-spatial couples, and subsequently for the dislocation source modeling. Each temporal-spatial couple consists of two impulses acting in opposite directions with both an infinitesimal separation distance and an infinitesimal time delay. By examining dynamite source modeling, this study shows that the proposed representation can intrinsically integrate the spatial and temporal features of the dislocation sources from the response point of view. This study also shows an example of a point, shear-slip seismic source modeling using traditional and proposed pulse representations for far-field wave motion. Discussion is finally provided for the implications of the proposed representation in broad applications. [DOI: 10.1115/1.1794712]

1 Introduction

Material crack, seismic rupture source, and the like, share the same source mechanism, i.e., time-dependent dislocation over a finite fault area, or simply the finite dislocation source. The finite dislocation source can be modeled as the summation (or integration in the limit) of point sources, each of which accounts for the evolutionary dislocation over a discretized subarea triggered at different time instants. Therefore, truthfully characterizing the point source is a key in understanding of nature of the dislocation source. More important, it has broad-based applications, exemplified as diagnosing crack damage in structural health monitoring, improving understanding of seismic source and subsequently using the source model to simulate earthquake ground motion for seismic design and retrofit, and conducting seismic survey for oil/gas exploration.

The mechanism of the above point source is typically modeled as the product of a source time function characterizing the dislocation growth (e.g., a ramp function), a factor combining nine couples of impulse forces that is equivalent to the unit dislocation, and a scaling factor or magnitude $(=\text{final dislocation} \times \text{material rigidity} \times \text{fault area})$. Each couple can be represented mathematically by two impulses acting in opposite directions, with an infinitesimal separation distance either along or perpendicular to the impulse direction or, in the limit, by the derivative of the impulse with respect to the separation-distance parameter. This study refers to the above approach as the *spatial pulse representation* for couples, or simply *spatial couples*. The combination of the nine spatial couples can be theoretically proved to be equivalent to any kind of a dislocation (normal and shear) in an arbitrary orientation from an elastodynamic approach, which uses the generalized Betti reciprocal relation by introducing a time parameter, as first described in [1]. In this approach, the Green's function (i.e., displacement response to a unit force) is assumed to be continuous everywhere, even on both sides of a fault as the dislocation never happens (for details, see [2]). Alternatively and more generally, Bakus and Mulcahy [3,4] use the concept of *stress glut* to derive the same results. That derivation uses the assumption that the earthquake and indigenous sources are considered to be the result of a localized, transient failure of the linearized elastic constitutive relation, which leads to the stress glut as a function of the dislocation quantity (for details, see [5]).

The above traditional approach basically characterizes the spatial and temporal features of the dislocation mechanism separately through two factors, i.e., the spatial couples and the source time function. It provides an equivalent description of far-field wave motion due to the point source, in which the features of the spatial couples are closely related to the type of the physical source. For example, the couples with separation-distance orientation along the impulse direction are usually related to the explosive source (or mode I in fracture mechanics), while the so-called double-couples (or combination of modes II and III), in which each couple has the separation-distance orientation perpendicular to the impulse direction, are associated with the seismic shear-dislocation source. The combination of the above two types of couples can be used to explain the volcanic earthquake source mechanism, among others (see details in [2,6–9]).

While the above point-source characterization and subsequent use for the finite dislocation source modeling have been well developed and widely used in study of seismology (with some applications in fracture mechanics, e.g., see [10,11]), the separation of temporal and spatial features in modeling the point dislocation mechanism needs to be reexamined.

In particular, the aforementioned traditional approach builds on

---

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 3, 2004; final revision, July 13, 2004. Associated 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.

the assumption that the two impulses in a spatial couple must *simultaneously* exist with an infinitesimal separation distance, thus generating no net force in the far field. Note that since the dislocation is equivalent to a combination of nine couples of impulses, only couples/impulses are used here and subsequently for explanation of the underlying physical phenomena. With the same reason of no net force generated by the spatial couples in the far field, *temporal pulse representation* for couples or *temporal couples* can be introduced, each of which is modeled by the two impulses acting in the opposite directions with no separation distance but with an infinitesimal time delay or, in the limit, by the derivative of the impulse with respect to time.

The objectives of this study are to examine the validation and implications of the alternative temporal couples for dislocation modeling. Moreover, the study will propose a concept of combined *temporal-spatial pulse representation* for the nine couples, or *temporal-spatial couples,* for dislocation source modeling. Each temporal-spatial couple consists of two impulses acting in opposite directions with both an infinitesimal separation distance and an infinitesimal time delay. This study will then provide the fundamentals and explore the underlying physics and implications of the proposed representation.

2 Spatial and Temporal Representations for a Couple

Without loss of generality, this study first investigates the far-field displacement wave motion due to a spatial couple of impulses in the x_j direction. Shown in Fig. 1, this couple is represented as two impulses at time τ acting in opposite directions along j with an infinitesimal separation distance h_k in direction k. A mathematical representation of the impulses exerted by this couple is

$$\begin{aligned}\vec{F}_j^s(\vec{x},t) &\equiv \vec{F}_j^{s+}(\vec{x},t)+\vec{F}_j^{s-}(\vec{x},t)\\ &= \frac{1}{h_k}[\delta(\vec{x}-\vec{\xi})-\delta(\vec{x}-\vec{\xi}+h_k\vec{e}_k)]\delta(t-\tau)\vec{e}_j\\ &\xrightarrow[h_k\to 0]{} \frac{\partial}{\partial \xi_k}\delta(\vec{x}-\vec{\xi})\,\delta(t-\tau)\vec{e}_j \quad \text{or}\\ &\xrightarrow[h_k\to 0]{} -\frac{\partial}{\partial x_k}\delta(\vec{x}-\vec{\xi})\,\delta(t-\tau)\vec{e}_j \qquad (1)\end{aligned}$$

where $\equiv$ denotes "by definition" and $\vec{F}_j^s(\vec{x},t)$ is the force representation of the couple that consists of the two impulses, $\vec{F}_j^{s+}(\vec{x},t)$ and $\vec{F}_j^{s-}(\vec{x},t)$. The superscript s is used to emphasize that this is a spatial pulse representation. In Eq. (1), t represents time, δ is the Dirac delta function, and $\vec{e}_i$ is the unit vector in the x_i direction where $i=j$ or k. Further, $\vec{x}=x_1\vec{e}_1+x_2\vec{e}_2+x_3\vec{e}_3$ represents the observation site in space, where $(\vec{e}_1,\vec{e}_2,\vec{e}_3)$ are the three orthogonal unit vectors, and $\vec{\xi}=\xi_1\vec{e}_1+\xi_2\vec{e}_2+\xi_3\vec{e}_3$ is the location of one source impulse. Note that $\vec{e}_j$ and $\vec{e}_k$ are not necessarily orthogonal or parallel to each other, or coincident with one of the three orthogonal unit vectors $(\vec{e}_1,\vec{e}_2,\vec{e}_3)$. Since the separation distance is infinitesimally small in comparison with the source-to-observation distance, the spatial pulse representation for the couple in Eq. (1) generates no net force at the locations far away from the source.

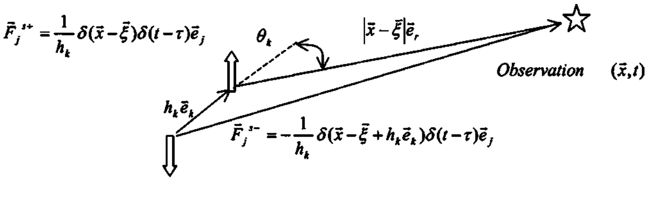


Fig. 1 Spatial pulse representation for a couple with an infinitesimal separation distance

Each of the two impulses will produce three additive terms of motion response, i.e., two far-field motions of pure P and pure S waves, respectively, and one near-field coupled P-S wave motion. This can be seen, for example, in [7], from the classic *Stokes* solution that is the displacement response to a force in a homogeneous medium. The far-field terms of the pure P- and pure S-wave motions are proportional to the geometrical spreading factor 1/R where R is the source-to-observation distance, while the near-field term for the coupled P-S wave motion is proportional to $1/\mathrm{R}^2$. As far as the far-field motion is concerned in this and other pertinent studies, the near-field motion term can be neglected, as illustrated in [7], among others. To this end and also due to the additive response feature, this study can examine the pure P- and S-wave motions in the far field separately.

The impulse response function $G_{ij}(\vec{x},t;\vec{\xi},\tau)$ is denoted as the P- or S-wave response in the direction i at $(\vec{x},t)$ due to a unit impulse in direction j at $(\vec{\xi},\tau)$, i.e., $\delta(\vec{x}-\vec{\xi})\delta(t-\tau)\vec{e}_j$. This study defines the unit impulse having the dimension of moment (force-length) and the impulse response function having the dimension of length squared. The far-field displacement of P- or S-waves at the observation site due to the couple $U_{ij}(\vec{x},t)$ can then be found as the superposition

$$\begin{aligned}U_{ij}(\vec{x},t) &= \frac{1}{h_k}[G_{ij}(\vec{x},t;\vec{\xi},\tau)-G_{ij}(\vec{x},t;\vec{\xi}-h_k\vec{e}_k,\tau)]\\ &\xrightarrow[h_k\to 0]{} \frac{\partial}{\partial \xi_k}G_{ij}(\vec{x},t;\vec{\xi},\tau) \qquad (2)\end{aligned}$$

Alternatively, Eq. (2) can be written as

$$\begin{aligned}U_{ij}(\vec{x},t) &= \frac{1}{h_k}[G_{ij}(\vec{x},t;\vec{\xi},\tau)-G_{ij}(\vec{x}+h_k\vec{e}_k,t;\vec{\xi},\tau)]\\ &\xrightarrow[h_k\to 0]{} -\frac{\partial}{\partial x_k}G_{ij}(\vec{x},t;\vec{\xi},\tau) \qquad (3)\end{aligned}$$

Equations (2) and (3) indicate that the derivatives of the impulse response function, with respect to the spatial parameters at the source and observation locations, are equivalent except for a sign change. This is a direct consequence of spatial reciprocity.

The spatial reciprocity of source and observation positions can be related to a temporal reciprocity at the source and observation locations. For a given impulse at $(\vec{\xi},\tau)$, the induced wave motion at $(\vec{x}+h_k\vec{e}_k,t)$ is equivalent to the wave motion at $(\vec{x},t+t_{dk})$, where the direction-dependent time difference t_{dk} is

$$t_{dk}=\frac{[h_k\vec{e}_k]\cdot[\vec{e}_r]}{v}=\frac{h_k}{v}\cos\theta_k \qquad (4)$$

Here the dot denotes an inner product, θ_k is the angle between the source-observation direction $\vec{e}_r$ and separation-distance direction $\vec{e}_k$, and v is the body wave speed (i.e., either P- or S-wave speed). Similarly, for a given wave motion at $(\vec{x},t)$, the causative impulse at $(\vec{\xi}-h_k\vec{e}_k,\tau)$ is equivalent to the impulse at $(\vec{\xi}_k,\tau-t_{dk})$.

With the aid of the above relationships, the first parts of Eqs. (2) and (3) can then be written as

$$\begin{aligned}U_{ij}(\vec{x},t) &= \frac{1}{h_k}[G_{ij}(\vec{x},t;\vec{\xi},\tau)-G_{ij}(\vec{x},t;\vec{\xi},\tau-t_{dk})]\\ &\xrightarrow[t_{dk}\to 0]{} \frac{\cos\theta_k}{v}\frac{\partial}{\partial\tau}G_{ij}(\vec{x},t;\vec{\xi},\tau) \qquad (5)\end{aligned}$$

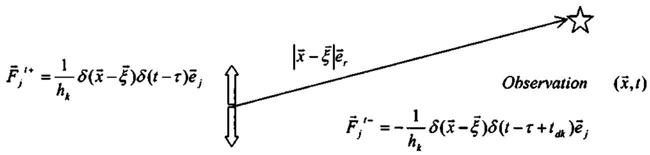


Fig. 2 Alternative temporal pulse representation for a couple with an infinitesimal time difference

$$U_{ij}(\vec{x},t)=\frac{1}{h_k}[G_{ij}(\vec{x},t;\vec{\xi},\tau)-G_{ij}(\vec{x},t+t_{dk};\vec{\xi},\tau)]$$

$$\xrightarrow[t_{dk}\to 0]{} -\frac{\cos\theta_k}{v}\frac{\partial}{\partial t}G_{ij}(\vec{x},t;\vec{\xi},\tau) \qquad (6)$$

where use has been made of Eq. (4). Comparison of Eqs. (2) and (3) with Eqs. (5) and (6) yields

$$\frac{\partial}{\partial \xi_k}G_{ij}(\vec{x},t;\vec{\xi},\tau)=\frac{\cos\theta_k}{v}\frac{\partial}{\partial \tau}G_{ij}(\vec{x},t;\vec{\xi},\tau) \qquad (7)$$

$$\frac{\partial}{\partial x_k}G_{ij}(\vec{x},t;\vec{\xi},\tau)=\frac{\cos\theta_k}{v}\frac{\partial}{\partial t}G_{ij}(\vec{x},t;\vec{\xi},\tau) \qquad (8)$$

The above equations imply that the temporal and spatial derivatives are exchangeable, to within an adjustable factor $\cos\theta_k/v$. These equations also suggest that the spatial pulse representation for a source couple in Eq. (1) can be alternatively replaced by the temporal pulse representation below and in Fig. 2,

$$\vec{F}_j^t(\vec{x},t)\equiv\vec{F}_j^{t+}(\vec{x},t)+\vec{F}_j^{t-}(\vec{x},t)$$

$$=\frac{1}{h_k}\delta(\vec{x}-\vec{\xi})[\delta(t-\tau)-\delta(t-\tau+t_{dk})]\vec{e}_j$$

$$\xrightarrow[t_{dk}\to 0]{} \frac{\cos\theta_k}{v}\frac{\partial}{\partial \tau}\delta(\vec{x}-\vec{\xi})\,\delta(t-\tau)\vec{e}_j \quad \text{or}$$

$$\xrightarrow[t_{dk}\to 0]{} -\frac{\cos\theta_k}{v}\frac{\partial}{\partial t}\delta(\vec{x}-\vec{\xi})\,\delta(t-\tau)\vec{e}_j \qquad (9)$$

where the superscript t in F is used to emphasize that this is a temporal pulse representation, in comparison with the superscript s.

Since the time difference for the temporal couple is infinitesimally small in comparison with the time elapse required for a wave traveling from the source to the far-field observation site, the temporal pulse representation in Eq. (9) introduces no net force in the locations far away from the source. Therefore, both the spatial and temporal pulse representations are conceivable for the indigenous source, the latter of which is referred to as any phenomena occurring within or upon the surface of the earth that does not involve forces exerted by any other bodies, [5].

3 Conventional Models for Dynamite Sources

The physical mechanism of a point dynamite source is highly nonlinear and very complicated (e.g., [12]). Nevertheless, the source for far-field wave motion can be modeled by three orthogonal couples in the spatial pulse representation, which essentially represents the consequence of a sudden pressurization, or forces in all the directions, on a spherical cavity (e.g., see detailed derivation in [13,14]). The far-field displacement in direction i due to the three couples, $u_i^c(\vec{x},t)$, can be found below, with the aid of Eq. (3) in which $k=j$,

$$u_i^c(\vec{x},t)=\sum_{j=1}^{3}U_{ij}(\vec{x},t)=-\sum_{j=1}^{3}\frac{\partial}{\partial x_j}G_{ij}(\vec{x},t;\vec{\xi},\tau) \qquad (10)$$

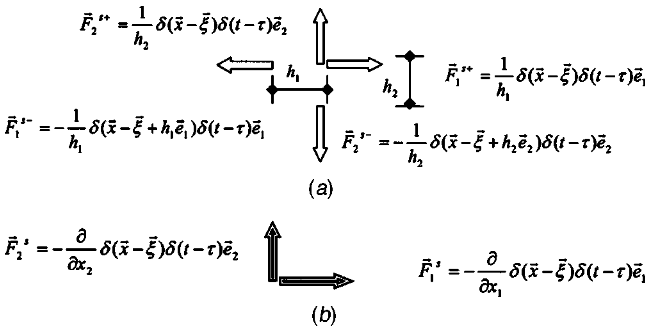


Fig. 3 (*a*) A 2D dynamite source with a spatial pulse representation for couples in a finite version; (*b*) a 2D dynamite source with a spatial pulse representation for couples in a limiting version

Since the dynamite source generates only P waves, the impulse response function in Eq. (10) and pertinent quantities in subsequent equations are related to P-wave motion only. Note that for simplicity, the factor of the source time function, describing the temporal evolution of the dynamite source or forces and scaled by the total magnitude or seismic moment of the source, is not included in Eq. (10). The superscript c is used to indicate that the displacement is obtained by using the *conventional approach* with the spatial pulse representation, as opposed to the new proposed approach to be elaborated on later.

With the aid of Eq. (6) or (8), Eq. (10) can be rewritten as

$$u_i^c(\vec{x},t)=-\sum_{j=1}^{3}\frac{\cos\theta_j}{v}\frac{\partial}{\partial t}[G_{ij}(\vec{x},t;\vec{\xi},\tau)]$$

$$=-\frac{1}{v}\frac{\partial}{\partial t}\left\{\sum_{j=1}^{3}[G_{ij}(\vec{x},t;\vec{\xi},\tau)\cos\theta_j]\right\} \qquad (11)$$

Note that the subscript j in Eqs. (10) and (11) is a dummy index, depending only on the coordinate system selected for describing the three orthogonal couples. Without loss of generality, therefore, one of the three couple directions j $(=1,2,3)$ can be selected to be coincident with the direction $\vec{e}_r$. Equation (11) then becomes

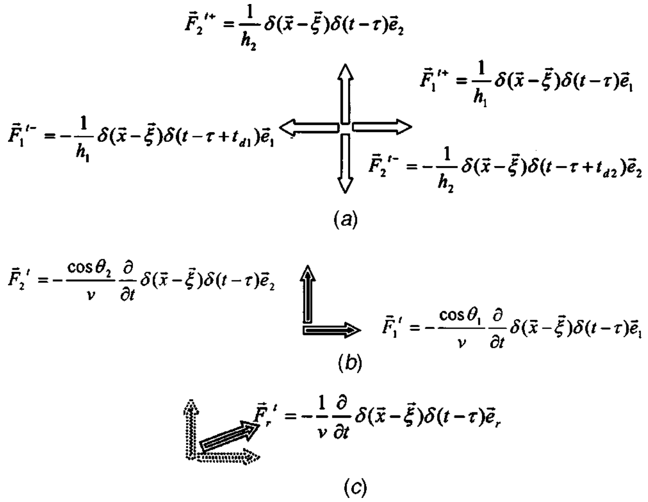


Fig. 4 (*a*) A 2D dynamite source with a temporal pulse representation for couples in a finite version; (*b*) a 2D dynamite source with a temporal pulse representation for couples in a limiting version; (*c*) a 2D dynamite source with a temporal pulse representation for couples in an alternative version

$$u_i^c(\vec{x},t) = -\frac{1}{v}\frac{\partial}{\partial t}G_{ir}(\vec{x},t;\vec{\xi},\tau) \tag{12}$$

Equations (10)–(12) indicate that the far-field displacement, due to a point dynamite source, is composed of fundamental solutions that are proportional to the derivative of the impulse response function with respect to either a spatial or a temporal parameter.

The above equivalence on far-field displacement can also be interpreted from the source point of view. Equation (10) suggests that the three orthogonal spatial couples generate the far-field wave motion. For convenience in illustration, a two-dimensional (2D) dynamite source model is shown in Figs. 3(*a*) and 3(*b*). Since all the impulses in Fig. 3(*a*) develop at the same time τ with infinitesimal separation distances h_1 and h_2, the spatial-based source model can be regarded as to physically capture (or snapshot) the cavity growth of the dynamite process with the cavity size proportional to h_1 and h_2 at time instant τ. Note again that the temporal evolution of the cavity growth is characterized by a source time function that is not included in Eq. (10).

On the other hand, Eq. (11), the alternative form of Eq. (10), implies that the far-field displacement is generated by three orthogonal couples that are described by the temporal pulse representation, as depicted for a 2D source model in Figs. 4(*a*) and 4(*b*). Since all the impulses in Fig. 4(*a*) develop at the same location ($\vec{\xi}$), but with infinitesimal time differences t_{d1} and t_{d2}, this alternative temporal-based source model may feasibly capture the temporal evolution of the impulse generation at a given location ($\vec{\xi}$) for the dynamite source. Similar to Eq. (11) and Figs. 4(*a*) and 4(*b*), Eq. (12) or its 2D case depicted in Fig. 4(*c*), shows an alternative temporal pulse representation for the dynamite source.

4 Implications of Spatial- and Temporal-Based Models

While both spatial and temporal pulse representations for the above dynamite source are consistent each other from the perspective of far-field wave-motion responses, their underlying physics and implications may be different, which is examined below.

The spatial-based model for the dynamite source characterizes the spatial and temporal features of a point dynamite source mechanism separately through the spatial couples and a source time function. The source time function is, however, typically assumed (e.g., a ramp function with a selected risetime for the dislocation progression), which may not have as much intrinsic physical meaning as the spatial couples. In fact, in deriving the spatial pulse representation for the couples that is equivalent to the dislocation, the principles for motion synchronization and compatibility are implicitly used, i.e., the pair of forces or stresses in opposite directions equivalent to the dislocation or discontinuity of strain *must* occur at the same time instant. While the synchronization is widely used in establishing and solving wave motion equations in a continuous medium, it may not truly characterize the dynamic process of a dislocation source that has integral temporal-spatial features.

In particular, from the perspective of far-field wave-motion responses, if one impulse of a couple can be regarded to take place at the source location closer to the observation site than the other impulse, the former impulse should also be viewed similarly as occurring at an earlier time instant than the latter impulse. Note that the above time difference in the generation of the two impulses for the far-field motion is different from the time difference due to the wave motion generated by the latter impulse traveling the infinitesimal separation distance. In other words, the two impulses in each of the couples for modeling a dislocation source have not only the infinitesimal spatial difference (traditional view) but also the infinitesimal temporal difference (new view that is parallel to the traditional view with the time replacing the space parameters).

With the above in mind, the traditional spatial-based source modeling may not capture the essential temporal features of the source in general, and two-impulses formation for each of the nine couples in particular, although the spatial features of the source are well characterized by the spatial couples as shown in Fig. 3(*a*). On the other hand, the alternative temporal-based source model may capture the temporal features of the two-impulses formation for each of the nine couples, while it may not seize upon the spatial characteristics in the two-impulses formation.

The above observations and analyses suggest that the dynamite source may be modeled on the basis of a combined temporal-spatial pulse representation for the couples, each of which has two opposite impulses built at both infinitesimal separation distance and time delay. Accordingly, the temporal and spatial features of the source can be characterized in an integrated way, which may overcome the influences of inappropriate assumption of a source time function in the conventional source modeling.

5 Temporal-Spatial Pulse Representation

Building on the above analysis, a temporal-spatial pulse representation for couples, and consequently for such applications as dynamite source modeling, is proposed. Shown in Fig. 5, the two impulses for a temporal-spatial couple act in opposite directions in direction j with an infinitesimal separation distance h_k in direction k and an infinitesimal time delay t_k. Mathematically, it can be represented as

$$\begin{aligned}\vec{F}_j^n(\vec{x},t) &\equiv \vec{F}_j^{n+}(\vec{x},t) + \vec{F}_j^{n-}(\vec{x},t)\\ &= \frac{1}{t_k h_k}[\delta(\vec{x}-\vec{\xi})\,\delta(t-\tau) - \delta(\vec{x}-\vec{\xi}+h_k\vec{e}_k)\,\delta(t-\tau-t_k)]\vec{e}_j\end{aligned} \tag{13}$$

where the superscript n denotes the *new*, temporal-spatial pulse representation for a couple. Note that the $\vec{F}_j^n(\vec{x},t)$ in Eq. (13) has a different dimension of force such as $\vec{F}_j^s(\vec{x},t)$ in Eq. (1), because the new pulse representation takes into consideration both the spatial and temporal features of the source to be modeled. The expression for $\vec{F}_j^n(\vec{x},t)$ can be adjusted to the force dimension by introducing a constant dimension-related factor in Eq. (13). For convenience and also for conceptual comparison only, however, such a factor is not used and similar terminology, such as "force" for $\vec{F}_j^n(\vec{x},t)$ and its induced wave motion, is still used in the subsequent analysis. With the same reasons used in explaining the spatial and temporal couples, the new pulse representation generates no net force in the locations far away from the source, and is therefore feasible for the indigenous source.

The temporal-spatial pulse representation for the couple [Fig. 5 and Eq. (13)] can be decomposed into one spatial and one temporal pulse representation (see Fig. 6),

$$\begin{aligned}\vec{F}_j^n(\vec{x},t) &= [\vec{F}_j^{ns+}(\vec{x},t) + \vec{F}_j^{ns-}(\vec{x},t)] + [\vec{F}_j^{nt+}(\vec{x},t) + \vec{F}_j^{nt-}(\vec{x},t)]\\ &= \frac{1}{t_k h_k}[\delta(\vec{x}-\vec{\xi})\,\delta(t-\tau) - \delta(\vec{x}-\vec{\xi}+h_k\vec{e}_k)\,\delta(t-\tau)]\vec{e}_j\\ &\quad + \frac{1}{t_k h_k}[\delta(\vec{x}-\vec{\xi}+h_k\vec{e}_k)\,\delta(t-\tau)\\ &\quad - \delta(\vec{x}-\vec{\xi}+h_k\vec{e}_k)\,\delta(t-\tau-t_k)]\vec{e}_j\end{aligned}$$

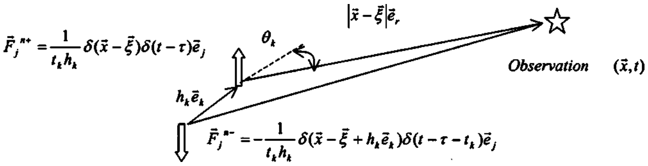


Fig. 5 Temporal-spatial pulse representation for a couple with both a separation distance and a time delay

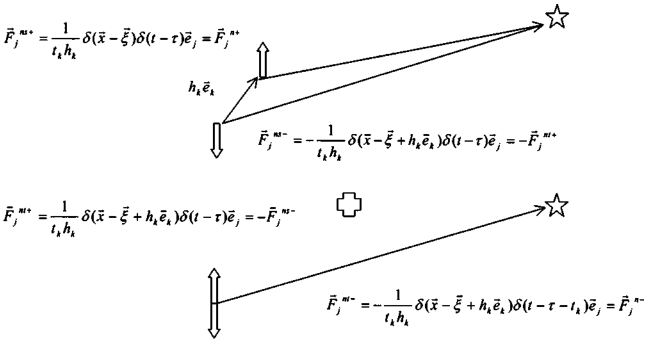


Fig. 6 Decomposition of the temporal-spatial pulse representation for a couple

$$\xrightarrow[t_k,h_k\to 0]{} -\frac{1}{t_k}\frac{\partial}{\partial x_k}\delta(\vec{x}-\vec{\xi})\,\delta(t-\tau)\vec{e}_j+\frac{1}{h_k}\frac{\partial}{\partial t}\delta(\vec{x}-\vec{\xi}+h_k\vec{e}_k)\,\delta(t-\tau)\vec{e}_j \tag{14}$$

The corresponding P- or S-wave displacement, $U_{ij}^n(\vec{x},t)$, is

$$\begin{aligned}U_{ij}^n(\vec{x},t)&=\frac{1}{t_kh_k}[G_{ij}(\vec{x},t;\vec{\xi},\tau)-G_{ij}(\vec{x}+h_k\vec{e}_k,t;\vec{\xi},\tau)]+\frac{1}{t_kh_k}\\&\quad\times[G_{ij}(\vec{x}+h_k\vec{e}_k,t;\vec{\xi},\tau)-G_{ij}(\vec{x}+h_k\vec{e}_k,t-t_k;\vec{\xi},\tau)]\\&\xrightarrow[t_k,h_k\to 0]{}-\frac{1}{t_k}\frac{\partial}{\partial x_k}G_{ij}(\vec{x},t;\vec{\xi},\tau)+\frac{1}{h_k}\frac{\partial}{\partial t}G_{ij}(\vec{x}\\&\quad+h_k\vec{e}_k,t;\vec{\xi},\tau)\end{aligned}\tag{15}$$

In general, the time delay for the impulse-induced waves traveling the infinitesimal separation distance between the two impulses is different from the time difference generated by the two impulses from the perspective of far field, i.e., $t_{dk}\neq t_k$. However, both quantities (t_{dk},t_k) are much smaller than the time required for the impulse-induced waves propagating from the source to the far-field observation site; an assumption can be made that they are approximately equal for the far-field motion. In fact, this is not an assumption if the derivative representation for each of the couples is used. This is similar to the fact that although generally $h_1\neq h_2$ in Fig. 3(*a*), they can be assumed to be the same. Nevertheless, the assumption will not affect the limiting case in Fig. 3(*b*) if the derivative representation is used.

With the above analysis, Eq. (15) then becomes

$$\begin{aligned}U_{ij}^n(\vec{x},t)&=-\frac{1}{h_k}\frac{\partial}{\partial t}G_{ij}(\vec{x},t;\vec{\xi},\tau)+\frac{1}{h_k}\frac{\partial}{\partial t}G_{ij}(\vec{x}+h_k\vec{e}_k,t;\vec{\xi},\tau)\\&\xrightarrow[h_k\to 0]{}\frac{\partial^2}{\partial x_k\partial t}G_{ij}(\vec{x},t;\vec{\xi},\tau)\\&=\frac{\cos\theta_k}{v}\frac{\partial^2}{\partial t^2}G_{ij}(\vec{x},t;\vec{\xi},\tau)\end{aligned}\tag{16}$$

where use is made of Eqs. (4) and (8). The far-field displacement for a dynamite source, $u_i^n(\vec{x},t)$, can then be found by

$$\begin{aligned}u_i^n(\vec{x},t)&=\sum_{j=1}^{3}\frac{\cos\theta_j}{v}\frac{\partial^2}{\partial t^2}G_{ij}(\vec{x},t;\vec{\xi},\tau)\\&=\frac{1}{v}\frac{\partial^2}{\partial t^2}\left\{\sum_{j=1}^{3}G_{ij}(\vec{x},t;\vec{\xi},\tau)\cos\theta_j\right\}\end{aligned}\tag{17}$$

Again, if one of the three couple-directions $j\ (=1,2,3)$ is selected to be coincident with the source-observation direction $\vec{e}_r$, Eq. (17) becomes

$$u_i^n(\vec{x},t)=\frac{1}{v}\frac{\partial^2}{\partial t^2}G_{ir}(\vec{x},t;\vec{\xi},\tau)\tag{18}$$

Equation (18) implies that the far-field displacement for a point dynamite source is proportional to the two derivatives of the impulse response function with respect to time, i.e.,

$$u_i^n(\vec{x},t)\propto\frac{\partial^2}{\partial t^2}G_{ir}(\vec{x},t;\vec{\xi},\tau)\tag{19}$$

Correspondingly, the force representation for the source is then proportional to the two time derivatives of an impulse, i.e.,

$$\vec{F}^n(\vec{x},t)\propto\frac{\partial^2}{\partial t^2}\delta(\vec{x}-\vec{\xi})\,\delta(t-\tau)\vec{e}_r\tag{20}$$

Equation (20) can also be derived from Eqs. (13) and (14) following the same approach used in Eqs. (15)–(19).

The underlying physics of Eq. (20) is now explained. From the perspective of far-field wave-motion response, the point dynamite source mechanism is an evolutionary process of the pressurization on a spherical cavity, which has a diameter infinitesimally small in comparison with the source-to-observation distance. All the same-amplitude forces of the pressurization on the spherical circumference of the cavity are not exerted simultaneously from the response point of view, i.e., each pair of the forces along a certain direction has a time delay that is infinitesimally small in comparison with the time elapse for the force-induced waves propagating from the source to the observation site. Note that all the physical quantities of the cavity formation [e.g., h_k and t_k in Eq. (13)] are also much smaller than the physical quantities of wave motion at the observation site, such as wavelengths of interest. The wave signals generated by the impulse at the cavity end nearest to the observation site will be received first, among all the signals from the source, while the wave signals induced by the impulse at the farthest end of the cavity will be picked up finally. Both the above causative impulses, i.e., a pair of impulses or a couple, are in the line of source-to-site direction. Note again that the time difference in the above pair of wave signals shown in the response is not only caused by the infinitesimal separation distance h_r but also the infinitesimal time delay t_r. The signals caused by the other pairs of impulses or couples in the other directions will be cancelled out by each other due to the symmetry to the source-to-site direction. This leads to a single couple of the impulses in the source-to-site direction in Eq. (20). The two derivatives in Eq. (20) are made of

Table 1 Comparison of conventional, alternative, and proposed pulse representations for the couples at source and wave motions in far field

Description	Conventional	Alternative	Proposed
Force representation for a dislocation source	$\frac{\partial}{\partial x}[\delta(\vec{x}-\vec{\xi})\delta(t-\tau)]$	$\frac{\partial}{\partial t}[\delta(\vec{x}-\vec{\xi})\delta(t-\tau)]$	$\frac{\partial^2}{\partial t^2}[\delta(\vec{x}-\vec{\xi})\delta(t-\tau)]$
Displacement of wave motion to a dislocation source	$\frac{\partial}{\partial x}[G(\vec{x},t;\vec{\xi},\tau)]$	$\frac{\partial}{\partial t}[G(\vec{x},t;\vec{\xi},\tau)]$	$\frac{\partial^2}{\partial t^2}[G(\vec{x},t;\vec{\xi},\tau)]$

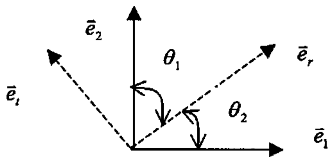


Fig. 7 Two sets of orthogonal unit vectors and their rotational angles for pulse representations of a shear-slip dislocation

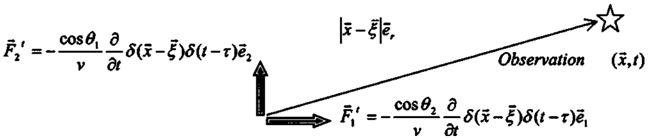


Fig. 9 A point shear-slip seismic source with a temporal pulse representation for couples

one temporal derivative, capturing the temporal features of the couple generation in the source-to-site direction, and another temporal derivative that is equivalent to the spatial derivative, characterizing the spatial features of the couple generation.

In comparison with the traditional source model that is based on the product of a factor for spatial couples and a source time function, the proposed dynamite source model for the wave motion in the far field has an extra derivative of the impulse response function with respect to the time parameter in comparison with the traditional model, which should intrinsically integrate the source temporal feature. Table 1 summarizes the traditional, alternative, and proposed pulse representations for the couples at source and wave motion in far field.

6 Extended Applications

While the above study focuses on the application of the proposed temporal-spatial pulse representation in modeling a point dynamite source, it is straightforward to extend it to modeling other dislocation sources such as seismic shear dislocation and material crack. As an illustration, below is an example of a point, shear-slip seismic source modeling using conventional and proposed pulse representation for in-plane, P- or S-wave motion in far field.

Conventionally, two orthogonal spatial couples or so-called double-couples [2], each of which has the impulse direction perpendicular to the separation-distance orientation, are used to model a point, shear-slip seismic source located at $\vec{\xi}$ and activating at τ, shown in Figs. 7 and 8. The far-field P- or S-wave displacement in direction i in the plane of $\vec{e}_1$ and $\vec{e}_2$ due to the double-couples, $u_i^c(\vec{x},t)$ can be found by

$$u_i^c(\vec{x},t) = -\frac{\partial}{\partial x_2}G_{i1}(\vec{x},t;\vec{\xi},\tau) - \frac{\partial}{\partial x_1}G_{i2}(\vec{x},t;\vec{\xi},\tau) \tag{21}$$

With the two reference frames ($\vec{e}_1$, $\vec{e}_2$) and ($\vec{e}_r$, $\vec{e}_t$) in Fig. 7, one could have the following transformation:

$$\vec{e}_1 = \vec{e}_r\cos\theta_2 - \vec{e}_t\sin\theta_2, \quad \vec{e}_2 = \vec{e}_r\sin\theta_2 + \vec{e}_t\cos\theta_2 \tag{22}$$

$$G_{i1}(\vec{x},t;\vec{\xi},\tau) = G_{ir}(\vec{x},t;\vec{\xi},\tau)\cos\theta_2 - G_{it}(\vec{x},t;\vec{\xi},\tau)\sin\theta_2$$

$$G_{i2}(\vec{x},t;\vec{\xi},\tau) = G_{ir}(\vec{x},t;\vec{\xi},\tau)\sin\theta_2 + G_{it}(\vec{x},t;\vec{\xi},\tau)\cos\theta_2 \tag{23}$$

Note that angle θ_1 (θ_2) is between directions $\vec{e}_r$ and $\vec{e}_2$ ($\vec{e}_r$ and $\vec{e}_1$) because the impulse direction $\vec{e}_1$ ($\vec{e}_2$) of a couple is perpendicular to the direction of separation distance of the couple with $\vec{e}_2$ ($\vec{e}_1$).

With the aid of Eqs. (8) and (23) and $\theta_1+\theta_2=\pi/2$, Eq. (21) for the far-field wave motion can be rewritten as

$$\begin{aligned} u_i^c(\vec{x},t) &= -\frac{\cos\theta_2}{v}\frac{\partial}{\partial t}G_{i1}(\vec{x},t;\vec{\xi},\tau) - \frac{\cos\theta_1}{v}\frac{\partial}{\partial t}G_{i2}(\vec{x},t;\vec{\xi},\tau) \\ &= -\frac{1}{v}\frac{\partial}{\partial t}G_{ir}(\vec{x},t;\vec{\xi},\tau) \end{aligned} \tag{24}$$

Equations (21) and (24) indicate that the far-field P- or S-wave displacement, due to a shear-slip source, can be obtained from the fundamental solutions that are proportional to the derivative of the impulse response function with respect to either a spatial or a temporal parameter.

The above equivalence on far-field displacement can also be interpreted from the source point of view. In particular, Eq. (21) suggests that two orthogonal spatial couples are responsible for the far-field P- or S-displacement wave response, depicted in Fig. 8. On the other hand, the first equality of Eq. (24) implies that two orthogonal temporal couples cause the response shown in Fig. 9. The last equality of Eq. (24) implies an alternative temporal representation for the source, i.e., a temporal couple in the direction $\vec{e}_r$.

With the aid of temporal-spatial pulse representation for couples, the far-field displacement caused by the shear-slip source can be found as

$$\begin{aligned} u_i^n(\vec{x},t) &= -\frac{\partial^2}{\partial x_2\partial t}G_{i1}(\vec{x},t;\vec{\xi},\tau) - \frac{\partial^2}{\partial x_1\partial t}G_{i2}(\vec{x},t;\vec{\xi},\tau) \\ &= -\frac{1}{v}\frac{\partial^2}{\partial t^2}G_{ir}(\vec{x},t;\vec{\xi},\varsigma) \end{aligned} \tag{25}$$

where use has been made of Eqs. (8) and (23). Equation (25) suggests that the far-field P- or S-wave displacement to a shear dislocation source, $u_i^n(\vec{x},t)$, is proportional to the two derivatives of an impulse response function with respect to the time, i.e.,

$$u_i^n(\vec{x},t) \propto \frac{\partial^2}{\partial t^2}G_{ir}(\vec{x},t;\vec{\xi},\tau) \tag{26}$$

Accordingly, the force representation for the source is then proportional to the two time derivatives of an impulse, as shown below and Fig. 10,

$$\vec{F}^n(\vec{x},t) \propto \frac{\partial^2}{\partial t^2}\delta(\vec{x}-\vec{\xi})\delta(t-\tau)\vec{e}_r \tag{27}$$

The underlying physics of impulses for the shear dislocation such as Eq. (27) are not as easily interpreted as those of dynamite source in Eq. (20) (e.g., see [2]), because the double-couples representation for a shear dislocation is derived in the sense that

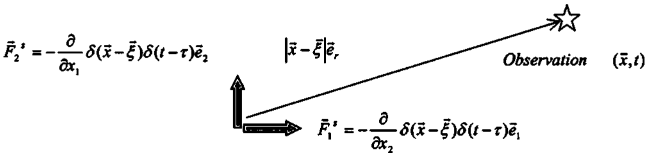


Fig. 8 A point shear-slip seismic source with a spatial pulse representation for couples

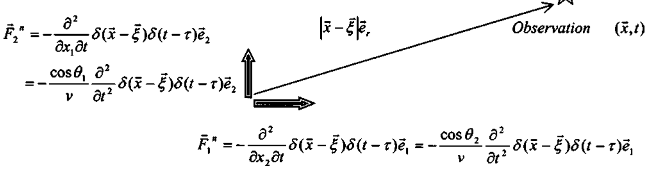


Fig. 10 A point shear-slip seismic source with a proposed temporal-spatial pulse representation for couples

impulse forces and the dislocation are equivalent in mechanics, not explicitly through the simple, physical observations.

7 Conclusions and Discussions

Following a review of the conventional spatial pulse representation for couples, a temporal pulse representation for couples is introduced. This study shows that the temporal and spatial pulse representations are alternative for the source modeling as far as the far-field wave motions are concerned. Furthermore, a temporal-spatial pulse representation for couples is proposed, which is then used to build new models for dislocation sources. The proposed temporal-spatial pulse representation for each couple consists of two impulses acting in opposite directions with an infinitesimal separation distance and an infinitesimal time delay, which can be applied for modeling various types of dynamic dislocation sources.

This study shows that the proposed representation for couples and subsequently for dynamite sources integrates intrinsically the temporal and spatial features of the source mechanism from the response point of view, which is physically sound in comparison with the traditional approach with the use of two separate factors, i.e., one from the spatial couples and the other from a source time function that is typically assumed and thus not fully inherent to the temporal features of the dynamic source mechanism.

While this study presents only the concept and fundamentals of the proposed representation, further development and validation of the representation in terms of the explicit solution expressions and numerical examples will improve the understanding of temporal-spatial mechanism of a class of dynamic dislocation sources. That will be reported in the near future.

The applications of the proposed representation for dislocation sources will be broad, which are briefly discussed now. Since the proposed source modeling suggests that the dislocation-induced signals in responses have one more derivative of an impulse response function than that with the conventional approach, the signature of the dislocation for structural health monitoring and damage detection could be explored most notably in the acceleration recordings if velocity recordings were used with the conventional approach. For seismic applications, the proposed source mechanism can be implemented into earthquake motion models [15,16]. One of the earthquake motion models [15] is assumed as the result of shear dislocation propagating on a fault line that is further discretized into a series of point sources. The activated time, location, and magnitude of each point source is modeled as a random process and field with the aid of the generalized random pulse train theory [17,18]. Accordingly, deterministic and statistical ground motion features can be theoretically obtained from the above model. In addition, with the use of the proposed couples for the source mechanism, the inverse solution for the earthquake source characterization based on the ground motion recordings will be different from the one with the conventional approach [18]. This would improve our understanding of the source mechanism in terms of the role of each source factor, such as slip and rise time in seismic ground motion. It would also help explore other wave propagation effects [19].

In short, the proposed representation would update the existing, practical-use approaches and techniques in modeling various dislocation sources, thus improving the accuracy and efficiency for the broad-based applications such as diagnosing crack damage in structural health monitoring, simulating earthquake ground motion for seismic design and retrofit, conducting seismic surveys for oil/gas exploration, and assessing influences of explosions in structures, among others.

Acknowledgments

The author would like to express sincere gratitude to Dr. Ken Larner at the Center for Wave Phenomena in Colorado School of Mines for providing invaluable advice on the first draft of the paper, and to Dr. Douglas Hart at WesternGeco, Inc. (then) and Regis University (now) for the joint research effort on modeling dynamite sources for seismic survey under the sponsorship of WesternGeco, Inc. and the Colorado Advanced Software Institute. This work was supported by the National Science Foundation with Grant No. 0414363 with Dr. Ken Chong as program director. The opinions, findings, and conclusions expressed herein are those of the author and do not necessarily reflect the views of the sponsors.

References

[1] Burridge, R., and Knopoff, L., 1964, ''Body Force Equivalents for Seismic Dislocations,'' Bull. Seismol. Soc. Am., **54**, pp. 1875–1888.
[2] Aki, K., and Richards, P. G., 1980, ''Quantitative Seismology: Theory and Methods,'' Freeman, San Francisco.
[3] Backus, G. E., and Mulcahy, M., 1976, ''Moment Tensors and Other Phenomenological Descriptions of Seismic Sources–I. Continuous Displacements,'' Geophys. J. R. Astron. Soc., **46**, pp. 341–361.
[4] Backus, G. E., and Mulcahy, M., 1976, ''Moment Tensors and Other Phenomenological Descriptions of Seismic Sources–II. Discontinuous Displacements,'' Geophys. J. R. Astron. Soc., **47**, pp. 301–329.
[5] Dahlen, F. A., and Tromp, J., 1998, *Theoretical Global Seismology*, Princeton University Press, Princeton, NJ.
[6] Ben-Menahem, A., and Singh, S. J., 1981, *Seismic Waves and Sources*, Springer-Verlag, Berlin.
[7] Lay, T., and Wallace, T. C., 1995, *Modern Global Seismology*, Academic Press, New York.
[8] Sheriff, R. E., and Geldart, L. P., 1995, *Exploration Seismology*, second Ed., Cambridge University Press, Cambridge, England.
[9] Koyama, J., 1997, *The Complex Faulting Process of Earthquakes*, Kluwer, Dordrecht.
[10] Freund, L. B., 1998, *Dynamic Fracture Mechanics*, Cambridge University Press, Cambridge, England.
[11] Grosse, C. U., Reinhardt, H. W., and Finck, F., 2003, ''Signal-based Acoustic Emission Techniques in Civil Engineering,'' J. Mater. Civ. Eng., **15**(3), pp. 274–279.
[12] Taylor, S. R., Patton, H. J., and Richards, P. G., 1991, *Geophysical Monograph 65*, American Geophysical Union, Washington, DC.
[13] Achenbach, J. D., 1980, *Wave Propagation in Elastic Solids*, North-Holland, Amsterdam.
[14] Patton, H., 1991, ''Seismic Moment Estimation and the Scaling of the Long-Period Explosion Source Spectrum,'' *Explosion Source Phenomenology*, S. R., Taylor, H. J. Patton, and P. G. Richards, eds Geophysical Monograph 65, American Geophysical Union, Washington, DC, pp. 171–183.
[15] Zhang, R., Yong, Y., and Lin, Y. K., 1991, ''Earthquake Ground Motion Modeling, II: Stochastic Line Source,'' J. Eng. Mech., **117**, pp. 2133–2150.
[16] Zhang, R., Zhang, L., and Shinozuka, M., 1997, ''Seismic Waves in a Layered Medium With Laterally Inhomogeneous Layers, I: Theory,'' ASME J. Appl. Mech., **64**, pp. 50–58.
[17] Lin, Y. K., 1963, ''Applications of Nonstationary Shot Noise in the Study of System Response to a Class of Nonstationary Excitations,'' ASME J. Appl. Mech., **30**(4), pp. 555–558.
[18] Lin, Y. K., 1986, ''On Random Pulse Train and its Evolutionary Spectral Representation,'' J. Probab. Eng. Mech., **1**(4), pp. 219–223.
[19] Zhang, R., Ma, S., and Hartzell, S., 2003, ''Signatures of the Seismic Source in EMD-based Characterization of the 1994 Northridge, California, Earthquake Recordings,'' Bull. Seismol. Soc. Am., **93**, pp. 501–518.

Journal of Applied Mechanics

Brief Notes

Three-Dimensional Steady-State String Motion in a Fluid Flow

Roman Miroshnik
The Israel Electric Corp., R&D Division, P.O. Box 10,
Haifa 31000, Israel

The phenomenon of three-dimensional (3D) steady-state motion of a string traveling along an invariant curve in a flowing medium is studied. Existence conditions are found using a perturbation scheme where a known two-dimensional (2D) solution is used as an initial approximation. [DOI: 10.1115/1.1794713]

1 Introduction

The paper examines the steady-state motion of an inextensible perfectly flexible string along an invariant curve of length L in a flowing medium having constant velocity W. The string, which has a mass density m, travels with constant velocity V between outlet and inlet rollers [1].

The string motion causes a constant tangential drag force, while the fluid induces a normal force. It is assumed that the traveling velocity is much greater than the fluid velocity and the drag resistance does not depend on it. Particular cases of string motion for quiescent [1–3] and flowing medium [4–6] were examined earlier.

2 Three-Dimensional Equations and Solution

Manipulations with string steady-state dimensionless equations [1] in Cartesian coordinates X_i and the Frenet triad $\boldsymbol{\tau}$, v, and b result in

$$\frac{du}{dv}=\frac{q_\tau u-q_{x1}}{q_\tau v-q_{x2}};\ \frac{dp^*}{dv}=-\frac{p^*q_\tau}{q_\tau v-q_{x2}};$$

$$\frac{ds}{dv}=\frac{p^*}{q_\tau v-q_{x2}};\ \frac{dx_1}{dv}=\frac{p^*u}{q_\tau v-q_{x2}};\tag{1}$$

$$\frac{dx_2}{dv}=\frac{p^*v}{q_\tau v-q_{x2}};\ \frac{dx_3}{dv}=\pm\frac{p^*\sqrt{1-u^2-v^2}}{q_\tau v-q_{x2}}.$$

where $x_i=X_i/L$; $s=S/L$; $q_{xi}=Q_{xi}/\text{mg}$; $q_\tau=Q_\tau/\text{mg}$; $p^*=P^*/\text{mgL}$ $(i=1,2,3)$; Q_{xi}, Q_τ are projections of external forces per unit length; S is an arc-length coordinate; $P^*=P-mV^2$ is the fictitious string tension; P is the real tension (P^* may have any sign, while P is always positive); and u and v are the direction cosines of the string tangent vector.

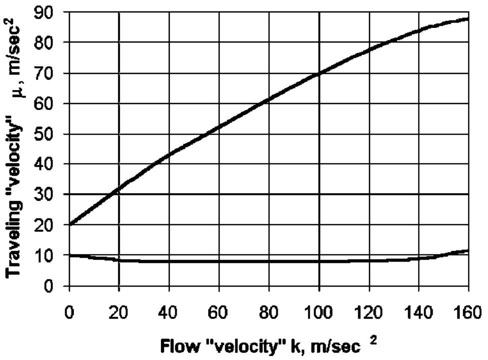


Fig. 1 Three critical domains of the string motion

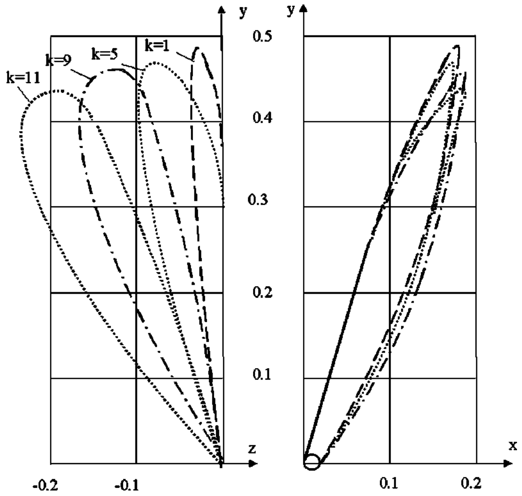


Fig. 2 String modes for different traveling and constant flow velocities

There is a singularity in Eq. (1) at $v = v^*$, where the normal string force vanishes. Because the singularity cannot be crossed uniquely by direct numerical integration, the integration of the equations is carried out in two stages: consecutively from the outlet and from the inlet until the singularity is reached.

The boundary conditions for both integration stages are

$$s(v_o)=0; \quad u(v_o)=u_0; \quad p^*(v_o)=p_o^*;$$

$$x_i(v_o)=0; \quad s(v_e)=0; \quad u(v_e)=u_e;$$

$$p^*(v_e)=p_e^*; \quad x_i(v_e)=x_{i0} \tag{2}$$

where u_o, v_o, u_e, and v_e are direction cosines corresponding to outlet and inlet, and x_{i0} are the coordinates of the inlet.

The boundary conditions (2) contain four unknowns, u_e, v_e, p_o^*, and p_e^*, that are found using a gradient method by successive iterations satisfying the discrepancy equations:

$$x_i^o - x_i^e = 0 \ (i=1,2,3); s^o + s^e - 1 = 0 \tag{3}$$

where x_i^o, s^o, x_i^e, and s^e are the correspondingly values of the variables x_i and s obtained at the end of the first and second integration stages.

Knowledge of an accurate initial approximation of unknowns is necessary to obtain solution convergence. A known solution of the two-dimensional (2D) problem [5] (for the case when medium flow acts on a quiescent string plane) is used as an initial approximation for the present 3D problem. The solution for the appropriate flow direction (not coinciding with initial string plane) is found by advanced movement with sufficiently small angular increments.

The fictitious tension and the radius of curvature ρ either vanish or are infinite at the singularity regardless of the boundary conditions (2). There are two critical velocities: V_{1cr} and V_{2cr}. They determine the minimum velocities for which, correspondingly, the tension and the curvature radius vanish at the singularity. Evidently the string motion exists, only if the tension vanishes.

Three domains of string motion are shown in Fig. 1 (where μ and k are correspondingly the drag force and the fluid resistance per string unit mass):

- Subcritical domain I, when $0 < V < V_{1cr}$; $\rho = \infty$,
- Supercritical one II, when $V_{1cr} < V < V_{2cr}$; $\rho = \infty$,
- Hypercritical domain III, when $V > V_{2cr}$; $\rho = 0$.

3 Results and Discussion

The string modes for different flow velocities and a constant string one are shown in Fig. 2. The experiments with a maximum airflow velocity of 100 m/sec were carried out in a wind tunnel [7]. There is good agreement between theoretical and experimental results.

The stability of steady-state modes is not analyzed here. The experimental results demonstrate the string modes are stable for the analyzed domain of parameters. This enables us to use the obtained results for practical use as a first approach.

References

[1] Miroshnik, R., 2001, "The Phenomenon of Steady-State String Motion," ASME J. Appl. Mech., **68**(4), pp. 568–574.

[2] Perkins, N. C., and Mote, Jr,, C. D., 1989, "Theoretical and Experimental Stability of Two Translating Cable Equilibria," J. Sound Vib., **128**(3), pp. 397–410.

[3] Healey, T. J., and Papadopoulos, J. N., 1990, "Steady Axial Motion of Strings," ASME J. Appl. Mech., **57**(3), pp. 785–787.

[4] Miroshnik, R. A., 1972, "Research of the Ballistic Antenna Stationary Motion in Flat Homogeneous Flow," Izv. Vuz. Mashinostr., **10**, pp. 27–32.

[5] Lemon, G., and Fraser, W. B., 2001, "Steady State Bifurcations and Dynamical Stability of a Whirling Cable Acted on by Aerodynamic Drag," Proc. R. Soc. London, Ser. A, **457**(2009), pp. 1021–1041.

[6] Svetlicky, V. A., and Miroshnik, R. A., 1975, "Research of the Ballistic Antenna Stationary Movement in the Flow of Arbitrary Direction," Soprotivlenie Mater. Teor. Soorujenii, **27**, pp. 36–42.

[7] Kurkin, V. I., and Miroshnik, R. A., 1974, "Influence of wind loading on the form of ballistic antenna," Izve. Vuz. Mashinostr., **5**, pp. 18–21.

Stability of a Rotating Heated Circular Plate With Elastic Edge Support

R. B. Maretic
e-mail: maretic@uns.ns.ac.yu

V. B. Glavardanov
e-mail: vanja@uns.ns.ac.yu
Faculty of Technical Sciences, University of Novi Sad,
21121 Novi Sad, Serbia and Montenegro

[DOI: 10.1115/1.1796448]

1 Introduction

The stability problem of plates has been of considerable importance in engineering. For the case of a thin circular plate described by the von Kármán nonlinear plate theory stability, analysis is given by Wolkowisky [1]. The stability of a clamped annular plate also described by the von Kármán nonlinear plate theory is presented by Machinek and Troger [2], where the Liapunov-Schmidt method is used. For the case of a moderately thick circular plate stability, analysis is given by Raju and Rao [3].

Many papers investigated the influence of temperature on the stability and the stress distribution of the plate. Pal [4] has examined the axisymmetric post-buckling behavior of annular heated plates with both edges clamped or simply supported. Ghosh [5] has shown that the flexural vibration of a spinning disk is strongly affected by the heat flow across the disk and specially by the parabolic type of temperature distribution. Renshaw [6] used the Liapunov method to determine the critical speed of a flexible spinning disk.

Many of the above-mentioned papers mostly deal with the stability of rotating annular plates with the free outer edge. However, there are many cases in practice where the outer edge of the plate is not free. One of them is given in Fig. 1, showing a circular plate of radius R and thickness h, the edge of which is welded to a rigid cylinder rotating at the constant angular speed ω. The plate considered can be the base or the lid of a rotating tank or centrifuge. The elasticity of the welded seam, which will be modeled by torsional springs uniformly distributed on the outer edge of the plate, will be taken into account. The mentioned torsional springs have distributed rotational stiffness k. It is assumed that the constant temperature T is imposed on the plate in comparison to its natural state.

In this paper, an attempt has been made to investigate the influence of the thermal effect on the stability of the above-mentioned circular plate spinning at a high angular speed. The adjacent equilibrium method will be used and only the axisymmetrical deformation will be examined.

2 Mathematical Formulation

In this paper we use von Kármán's theory for axisymmetric thin circular plates and the Duhamel-Neumann theory. The governing equations for a rotating heated plate (see Timoshenko and Woinowsky-Krieger [7] and Nowacki [8]) can be reduced to the following system:

$$\begin{gathered}\frac{dw}{dr}=-\theta,\\ \frac{d\theta}{dr}=k_r,\\ \frac{dk_r}{dr}=\frac{1}{D}N_r\theta-\frac{k_r}{r}+\frac{\theta}{r^2},\\ \frac{dN_r}{dr}=\frac{uEh}{r^2}-\frac{N_r(1-\nu)}{r}-rh\rho\omega^2-\frac{Eh\alpha T}{r},\\ \frac{du}{dr}=-\frac{1}{2}\theta^2-\nu\frac{u}{r}+N_r\frac{1-\nu^2}{Eh}+\alpha T(1+\nu),\end{gathered}\tag{1}$$

where $D=Eh^3/12(1-\nu^2)$ is the bending stiffness, r is the radial coordinate, E is the elastic modulus, ν is the Poisson ratio, α is the thermal expansion coefficient, u and w are the radial and transverse displacements of the middle plane, N_r is the radial membrane force, ρ is the mass density, θ is the angle of the tangent in the radial direction, and k_r is the radial curvature. The boundary conditions are

$$\begin{gathered}u(0)=0,\quad \theta(0)=0,\quad w(R)=0,\quad u(R)=0,\\ kh\theta(R)+D\left[k_r(R)+\frac{\nu}{R}\theta(R)\right]=0.\end{gathered}\tag{2}$$

Introducing the nondimensional variables

$$x=\frac{r}{R},\tag{3a}$$

$$y=\frac{w}{h},\tag{3b}$$

$$\Theta=\frac{\theta R}{h},\tag{3c}$$

$$K=\frac{k_rR^2}{h},\tag{3d}$$

$$U=\frac{uR}{h^2},\tag{3e}$$

$$N=\frac{N_rR^2}{Eh^3},\tag{3f}$$

$$C_T=\frac{kR}{D},\tag{3g}$$

$$\lambda=\frac{\rho\omega^2R^4}{Eh^2},\tag{3h}$$

$$\tau=\frac{\alpha TR^2}{h^2},\tag{3i}$$

and by using Eqs. (1) we obtain

$$\begin{gathered}\dot{y}=-\Theta,\\ \dot{\Theta}=K,\\ \dot{K}=12(1-\nu^2)N\Theta-\frac{K}{x}+\frac{\Theta}{x^2},\\ \dot{N}=\frac{U}{x^2}-\frac{N(1-\nu)}{x}-x\lambda-\frac{\tau}{x},\\ \dot{U}=-\frac{1}{2}\Theta^2+N(1-\nu^2)-\nu\frac{U}{x}+\tau(1+\nu),\end{gathered}\tag{4}$$

where $(\dot{\ })=d(\ )/dx$, subject to

$$U(0)=0,\quad \Theta(0)=0,\tag{5}$$

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, November 26, 2002, final revision, May 10, 2004. Associate Editor: N. Triantafyllidis.

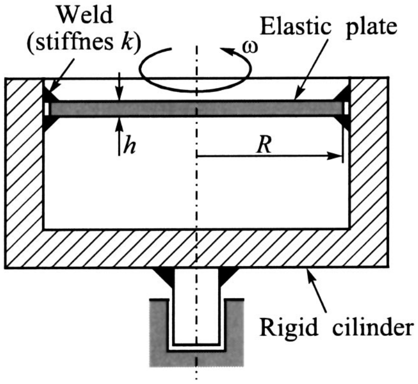


Fig. 1 A circular plate supported elastically around its edge

$$y(1)=0, \quad U(1)=0, \quad (C_T+\nu)\Theta(1)+K(1)=0.$$

The system (4) and (5) has a trivial solution in which the plate remains planar. The trivial solution reads.

$$y_T=0 \tag{6a}$$

$$\Theta_T=0, \tag{6b}$$

$$K_T=0, \tag{6c}$$

$$N_T=\mathcal{N}\lambda-\frac{1}{1-\nu}\tau, \tag{6d}$$

$$U_T=\mathcal{U}\lambda. \tag{6e}$$

where

$$\mathcal{N}=\tfrac{1}{8}[1+\nu-(3+\nu)x^2], \quad \mathcal{U}=\frac{1-\nu^2}{8}(1-x^2)x. \tag{7}$$

We assume the solution of Eqs. (4) in the form

$$\begin{gathered} y=y_T+y^*, \quad \Theta=\Theta_T+\Theta^*, \quad K=K_T+K^*, \\ U=U_T+U^*, \quad N=N_T+N^*. \end{gathered} \tag{8}$$

By using Eqs. (8) in this decomposition and substituting the result into Eqs. (4), we finally obtain

$$\begin{gathered} \dot{y}^*=-\Theta^*, \\ \dot{\Theta}^*=K^* \\ \dot{K}^*=12(1-\nu^2)(N_T+N^*)\Theta^*-\frac{K^*}{x}+\frac{\Theta^*}{x^2}, \\ \dot{N}^*=\frac{U^*}{x^2}-\frac{1-\nu}{x}N^*, \\ \dot{U}=-\frac{1}{2}(\Theta^*)^2+(1-\nu^2)N^*-\frac{\nu}{x}U^*, \end{gathered} \tag{9}$$

subject to

$$U^*(0)=0, \tag{10a}$$

$$\Theta^*(0)=0, \tag{10b}$$

$$y^*(1)=0, \tag{10c}$$

$$U^*(1)=0, \tag{10d}$$

$$(C_T+\nu)\Theta^*(1)+K^*(1)=0. \tag{10e}$$

Note that the system (9) and (10) has the solution $y^*=\Theta^*=K^*=U^*=N^*=0$ for all the values of λ and τ. The nontrivial solutions of Eqs. (9) and (10) correspond to the buckled state (loss of stability) of the plate.

3 Bifurcation Analysys and Critical Angular Speed

To determine the possible bifurcation points of the system (9), (10) we consider the linearization of (9). The result of linearization is

$$\dot{y}^*=-\Theta^*, \tag{11a}$$

$$\dot{\Theta}^*=K^*, \tag{11b}$$

$$\dot{K}^*=12(1-\nu^2)N_T\Theta^*-\frac{K^*}{x}+\frac{\Theta^*}{x^2}, \tag{11c}$$

$$\dot{N}^*=\frac{U^*}{x^2}-\frac{1-\nu}{x}N^*, \tag{11d}$$

$$\dot{U}^*=(1-\nu^2)N^*-\frac{\nu}{x}U^*. \tag{11e}$$

The solution to Eqs. (11*d*), (11*e*), (10*a*), and (10*d*) is

$$U_L^*=0, \quad N_L^*=0, \tag{12}$$

while Eqs. (11*a*)–(11*c*) can be combined to give

$$x^2\dddot{y}^*+x\ddot{y}^*-[12(1-\nu^2)N_Tx^2+1]\dot{y}^*=0. \tag{13}$$

Reducing the order of Eq. (13) and using Eqs. (6*d*) and (7), we obtain

$$x^2\ddot{\Theta}^*+x\dot{\Theta}^*+(bx^4-cx^2-1)\Theta^*=0, \tag{14}$$

where

$$b=\tfrac{3}{2}(1-\nu^2)(3+\nu)\lambda, \quad c=12(1-\nu^2)\left(\frac{1+\nu}{8}\lambda-\frac{\tau}{1-\nu}\right). \tag{15}$$

The solution to the Eq. (14) reads

$$\Theta_L^*=C_1\frac{1}{x}F_0(\eta,\tfrac{1}{2}\sqrt{b}x^2)+C_2\frac{1}{x}G_0(\eta,\tfrac{1}{2}\sqrt{b}x^2), \tag{16}$$

where F_0 is a regular Coulomb wave function of order zero and G_0 is an irregular Coulomb wave function of order zero (Abramowitz and Stegun [9]), C_1 and C_2 are the constants, and

$$\eta=\frac{c}{4\sqrt{b}}. \tag{17}$$

Satisfying the boundary condition (10*b*) we get $C_2=0$, so that Eq. (16) becomes

$$\Theta_L^*=C_1\frac{1}{x}F_0(\eta,\tfrac{1}{2}\sqrt{b}x^2). \tag{18}$$

From Eqs. (11*a*) and (11*b*) it follows that

$$y_L^*=-C_1\int_1^x\frac{1}{x}F_0(\eta,\tfrac{1}{2}\sqrt{b}x^2)dx, \tag{19a}$$

$$K_L^*=C_1\frac{d}{dx}\left[\frac{1}{x}F_0(\eta,\tfrac{1}{2}\sqrt{b}x^2)\right]. \tag{19b}$$

Substituting Eqs. (18) and (19*b*) into the boundary condition (10e) we get

$$C_1[(C_T+\nu-1)F_0(\eta,\tfrac{1}{2}\sqrt{b}x^2)+\dot{F}_0(\eta,\tfrac{1}{2}\sqrt{b}x^2)]|_{x=1}=0. \tag{20}$$

Next, for the case of parameters τ and C_T fixed, solving Eq. (20) the corresponding values of λ are obtained. The choice of the

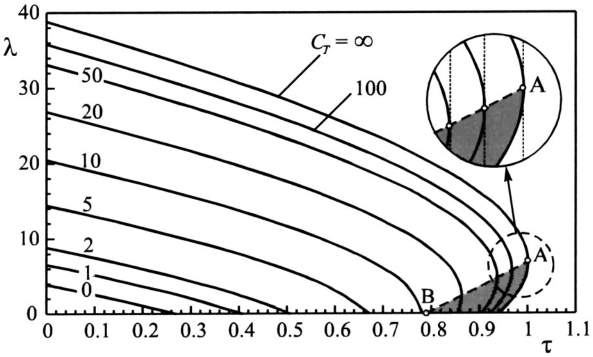


Fig. 2 The critical speed parameter λ_{cr} with super- and subcritical bifurcation regions

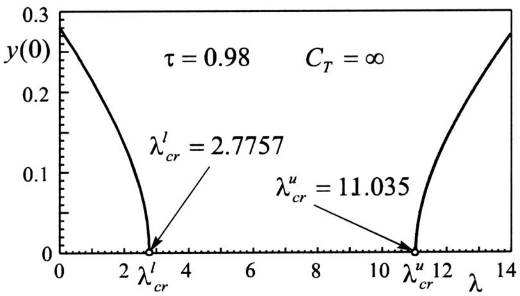


Fig. 3 Maximal transverse displacements

critical value λ_{cr}, between the obtained ones, will be the discussed in Sec. 4. We proceed following the standard Liapunov-Schmidt procedure (see Golubitsky and Schaeffer [10] or Chow and Hale [11]). Therefore we determine the formal adjoint of Eqs. (10) and (11) and then formulate the bifurcation equation in the following form:

$$g(a,\lambda_{cr}+\Delta\lambda)=c_1 a\Delta\lambda+c_3 a^3+O(|a|^2\Delta\lambda,|a|^4), \tag{21}$$

where a is a small real amplitude parameter (see Troger and Steindl [12]), $\Delta\lambda \ll 1$, and

$$c_1=\tfrac{1}{8}\int_0^1\left[\frac{1+\nu}{x}-(3+\nu)x\right]F_0^2(\eta,\tfrac{1}{2}\sqrt{b}x^2)dx, \tag{22}$$

$$c_3=\tfrac{1}{2}\int_0^1\frac{1}{x}\left[W(x)-\frac{1}{1-\nu}\dot{W}(1)-W(1)\right]F_0^2(\eta,\tfrac{1}{2}\sqrt{b}x^2)dx, \tag{23}$$

with

$$W(x)=-\frac{1}{x^2}\int x\int\frac{1}{x^3}F_0^3(\eta,\tfrac{1}{2}\sqrt{b}x^2)dxdx. \tag{24}$$

When writing Eqs. (22) and (23) we assumed that $C_1=1$ without reducing generalization. To study the qualitative behavior of the solution, we can neglect the higher-order terms in Eq. (21). Thus we conclude that Eq. (21) is contact equivalent (Keyfitz [13]) to

$$g(a,\lambda_{cr}+\Delta\lambda)=\mathrm{sgn}(c_1)a\Delta\lambda+\mathrm{sgn}(c_3)a^3. \tag{25}$$

Therefore, Eq. (25) has pitchfork bifurcation, and this type of bifurcation will be studied in the next section. If the constants c_1 and c_2 are of the same sign, the bifurcation is called subcritical; if they are of the opposite sign, the bifurcation is supercritical.

4 Results and Discussion

Figure 2 shows the nondimensional critical angular speed λ_{cr} with respect to the nondimensional temperature τ and the rotational stiffness C_T. The parameter λ_{cr} was determined by solving Eq. (20) given that the Poisson ratio is $\nu=0.3$. In case of $\tau=0$ (not heated plate) the obtained critical values of λ_{cr} are the same as in Maretic [14]. The critical values of λ_{cr} in that paper were determined using the dynamic stability criterion. It can be noticed that overall rule of behavior of the parameter λ_{cr} shows a decrease when the temperature increases and shows an increase when the rotational stiffness increases. However, there is an exception to this rule, since there are certain temperature values that have two corresponding critical values of the parameter λ_{cr}: the upper (λ^u_{cr}) and lower (λ^l_{cr}). In this case they are obtained as the two smallest positive roots of Eq. (20). So, depending on the nondimensional temperature τ and the rotational stiffness C_T, there can be only one or two critical angular speeds. λ_{cr} can be determined as the smallest positive root (only one solution) of Eq. (20) or as the two smallest positive roots (two solutions) of Eq. (20). The lower critical values (λ^l_{cr}) are shown by the shadowed part of the diagram, below the line AB. In that part of the diagram when the temperature increases, the critical angular speed also increases, which is an unexpected behavior. Based on the bifurcation equation (21) it can be established that $c_1>0$ and $c_3>0$ within the shadowed diagram area. This shows that bifurcation is subcritical in this diagram area, while above it (the area not shedowed) $c_1<0$ and $c_3>0$, which shows that bifurcation is supercritical. In the enlarged diagram part around point A in Fig. 2 it can be seen that points on the boundary line AB (which separates different bifurcation types) are determined as points on the curves whose tangents are vertical.

To confirm subcritical bifurcation the numerical integration of the differential equations systems (4) and (5) was performed. For the special case of the clamped plate ($C_T=\infty$) and the parameter value of $\tau=0.98$, Fig. 3 shows the maximal nondimensional transverse displacements $y(0)$ obtained using the numerical integration. For $\lambda=0$ the value of the maximal transverse displacement is $y(0)=0.2804$. With the increase of the angular speed the maximal transverse displacement reduces and for the value of $\lambda=2.7757$ it is $y(0)=0$. So, this value of the parameter λ is the lower critical value (λ^l_{cr}). With the increase of the angular speed the plate remains stable. The plate is not bent and its state is described by the trivial solution (6). The plate retains this state until the angular speed reaches the value of $\lambda=11.035$, which is the upper critical value (λ^u_{cr}). When the parameter λ exceeds the upper critical value, $\lambda>\lambda^u_{cr}$, the plate loses stability again and it bends. The shape of this diagram confirms that there are really two stability boundaries and two bifurcation types, sub- and supercritical.

The numerical integration of the equation system confirms the conclusion of existence of the subcritical bufurcation, as well as the upper and lower critical values of the angular speed. Therefore, the plate will be stable provided that its angular speed is inside the range defined by the lower and upper critical angular speed. For example, if the clamped plate $R=1$ m and $h=30$ mm, made of concrete ($\rho=7850\ \mathrm{kg/m^3}, E=2.1\times10^{11}\ \mathrm{N/m^2}, \alpha=1.26\times10^{-5}\ 1/°\mathrm{C}$), heated at $T=70°\mathrm{C}$ is considered, it will be stable for $\omega^l_{cr}<\omega<\omega^u_{cr}$, where $\omega^l_{cr}=2465$ rpm and $\omega^u_{cr}=4921$ rpm.

5 Conclusions

This paper presents the following results:

(1) The critical values of the angular speed at which the rotating heated circular plate loses its stability have been deter-

mined. From Eq. (3*h*) it follows that the critical angular speed is given by

$$\omega_{\mathrm{cr}}=\frac{h}{R^2}\sqrt{\frac{E\lambda_{\mathrm{cr}}}{\rho}},$$

where λ_{cr} is the root of Eq. (20).

(2) Using the Liapunov-Schmidt method it has been shown that the loss of stability for different values of temperature and stiffness can be sub- and supercritical. Subsequently, there are two values of the critical angular speed, lower and upper. The plate is stable if its angular speed is between the critical values. In case of existence of the supercritical bifurcation only, the plate is stable if its angular speed is less than the critical one.

(3) As an interesting effect we show that under certain circumstances the angular speed can stabilize the plate which lost stability due to heating.

Acknowledgment

This research was supported by the Ministry of Science, Technologies and Development of Republic of Serbia, Project No. 1402.

References

[1] Wolkowisky, J. H., 1967, ‘‘Existence of Buckled States of Circular Plates,’’ Commun. Pure Appl. Math., **XX**, pp. 546–560.

[2] Machinek, A. K., and Troger, H., 1988, ‘‘Postbuckling of Elastic Annular Plates at Multiple Eigenvalues,’’ Dyn. Stab. Syst., **3**, pp. 78–98.

[3] Raju, K. K., and Rao, G. V., 1983, ‘‘Postbuckling Analysis of Moderately Thick Elastic Circular Plates,’’ ASME J. Appl. Mech., **50**, pp. 468–470.

[4] Pal, M. C., 1969, ‘‘Large Deformations of Heated Circular Plates,’’ Acta Mech., **8**, pp. 82–91.

[5] Ghosh, N. C., 1975, ‘‘Thermal Effect on the Transverse Vibration of Spining Disk of Variable Thickness,’’ ASME J. Appl. Mech., **42**, pp. 358–362.

[6] Renshaw, A. A., 1998, ‘‘Critical Speed for Floppy Disks,’’ ASME J. Appl. Mech., **65**, pp. 116–120.

[7] Timoshenko, S., and Woinowsky-Krieger, S., 1959, *Theory of Plates of Shells*, 2nd ed. McGraw-Hill, New York.

[8] Nowacki, W., 1962, *Thermoelasticity*, International Series of Monograph on Aeronautics and Astronautics, **3**, Addison-Wesley, Reading, MA.

[9] Abramowitz, M., and Stegun, I., 1965, *Handbook of Mathematical Functions*, Dover, New York.

[10] Golubitsky, M., and Schaeffer, D. G., 1985, *Singularities and Groups in Bifurcation Theory*, Vol. 1, Springer, New York.

[11] Chow, S. N., and Hale, J. K., 1982, *Methods of Bifurcation Theory*, Springer, New York.

[12] Troger, H., and Steindl, A., 1991, *Nonlinear Stability and Bifurcation Theory: An Introduction for Engineers and Applied Scientisty*, Springer, Wien.

[13] Keyfitz, B. L., 1986, ‘‘Classification of one-state variable bifurcation problems up to codimension seven,’’ Dyn. Stab. Syst., **1**, pp. 1–41.

[14] Maretic, R., 1998, ‘‘Vibration and Stability of Rotating Plates with Elastic Edge Supports,’’ J. Sound Vib., **210**, pp. 291–294.