ASME
SETTING THE STANDARD

Journal of Applied Mechanics

Published Bimonthly by ASME

VOLUME 73 • NUMBER 2 • MARCH 2006

Transactions of the ASME, Journal of Applied Mechanics (ISSN 0021-8936) is published bimonthly (Jan., Mar., May, July, Sept., Nov.) by The American Society of Mechanical Engineers, Three Park Avenue, New York, NY 10016. Periodicals postage paid at New York, NY and additional mailing offices. POSTMASTER: Send address changes to Transactions of the ASME, Journal of Applied Mechanics, c/o THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 22 Law Drive, Box 2300, Fairfield, NJ 07007-2300. **CHANGES OF ADDRESS** must be received at Society headquarters seven weeks before they are to be effective. Please send old label and new address. **STATEMENT from By-Laws.** The Society shall not be responsible for statements or opinions advanced in papers or ... printed in its publications (B7.1, Para. 3).

GUEST EDITORIAL

TECHNICAL PAPERS

(Contents continued on inside back cover)

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

Journal of Applied Mechanics Volume 73, Number 2 MARCH 2006

TECHNICAL BRIEFS

DISCUSSION

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

Journal of Applied Mechanics

Guest Editorial

Vision of the Future of Solid Mechanics

Although mechanics is the oldest field of science, it is still advancing rapidly, driven, in all areas of technology, by the need:

(1) to develop fundamental understanding of material and system behavior at *multiple scales*;
(2) to deduce realistic *mesoscopic and macroscopic models*, and verify and calibrate them experimentally;
(3) to advance the understanding of *failure mechanisms* of materials, structures, and systems;
(4) to achieve *mechanically and functionally superior* performance, ensuring a near-zero incidence of failure and minimal long-term deterioration; and
(5) to advance the understanding and *robustness* of diverse *complex* systems such as those found in biology and nanotechnology.

Advances in solid mechanics are made possible by new superior instrumentation, novel and ingenious experimental methods, better mathematical models, even more powerful computational tools, and new and growing interactions with an ever expanding range of disciplines.

Importance of Mechanics in the Modern World

The successful development of technology in broad fields of activity still crucially depends on advances in solid mechanics and its application. Opportunities can be identified in areas such as electronics, where devices depend on successful utilization of mechanical effects as much as on electrical phenomena, with examples including epitaxy, strain dependent band-gaps and the integrity of electrical leads and connections, dies, and circuit boards. Biology, whether related to medical treatments, involving prostheses, stents, and implants, or regarding the growth, function, adhesion, and motion of cells, the conformation and interaction of proteins, and the ubiquity of DNA and other molecules, is replete with issues such as dynamics, compatibility, and constitutive response that are the bread and butter of solid mechanics. Diverse fields of engineering are still vitalized by advances in solid mechanics, including blast resistant structures, tough, strong, and durable advanced materials, thermal protection systems, and deployable structures.

Frustrations exist that can be remedied by future developments in solid mechanics, whether it is the absence of hypersonic vehicles due to inadequacies in stiffness, strength, and temperature resistance of materials, of the unpredictability of earthquakes, landslides, and avalanches due to inadequate advances in the understanding of large-scale frictional cohesive shear fracture in heterogeneous materials. Also, the insufficient durability of hip replacements and other prostheses requires a much better understanding of long-term progressive damage and frictional wear of composites and micro-porous metals. The poor fatigue resistance and durability of systems in aggressive environments demands better technology to avoid many catastrophic failures and to achieve enormous savings by extension of the lifespan of our nation's infrastructure.

Avoidable failures continue to occur, and ensuring that they are not repeated in the future depends on continued advances in solid mechanics. Consider some examples: both space shuttle disasters have been traced to material failure (leakage of a seal in one case and loss of integrity of structure and thermal protection in the other); the crash of a DC-10 in takeoff from O'Hare Airport in 1979 was caused by a fatigue crack in an engine pylon; the crash of an Airbus shortly after takeoff from JFK Airport in 2001 was probably caused by overload fracture in a large vertical stabilizer made of advanced composites; the World Trade Center collapse was triggered by viscoplastic deformation of columns heated by fire; and the giant Sleipner oil platform would not have sunk, and earthquakes would not have destroyed the viaducts in Kobe, Oakland, or Los Angeles, were quasi-brittle compression-shear fracture and its scaling understood.

Evidently, mechanics of solids is the controlling factor in many advanced technologies, a roadblock to implementing innovative technologies, and the explanation for many catastrophes. Advanced experimental methods and large-scale computer simulations are now rendering realistic simulation and prediction feasible. We have great opportunities for research.

We also face challenges in education. We need to attract and train new generations of solid mechanicians at a time when the preparation of the young in mathematics and science is degrading, while many students perceive other opportunities to be more attractive or glamorous, and most universities are no longer interested in mechanics programs. In the face of this, the tensorial nature of solid mechanics, the nonlinearities of constitutive behavior, and the complexities of damage and scaling necessitate prolonged training in a focused program.

Solid mechanics is a unifying discipline. Cutting across many professions, it is perhaps the most interdisciplinary scientific activity in the leading engineering schools. The scientific field is one, but its interventions span mechanical, aeronautical, aerospace, civil, materials, biomedical, chemical, environmental, nuclear, offshore, naval, arctic, and electrical engineering, as well as the sciences of materials, biology, chemistry, physics, geophysics, and planetology.

Challenges and Opportunities for Research

With no claim for completeness, a diverse set can be assembled:

(1) *Multiscale modeling*, connecting the hierarchy of scales in materials (nano-micro-meso-macro), is a dominant trend. Embedding a discrete model at one scale (e.g., atomistic simulation, simulation of discrete dislocations, simulation of particles or fibers in a matrix, the role of nanopores in concrete) into a continuum model at the next higher scale is a challenge where success can yield superior understanding of composites, polycrystals, and porous or cellular materials.
(2) *Failure scaling and size effects* represent a companion problem—that of finding the laws of transition among re-

gimes whose scaling individually can be characterized by power laws; e.g., the transition from a discrete local scale of fibers or particles embedded in a matrix to a continuum representing a composite; from crystal grains with dislocations to a continuous thin film; from intact rock blocks to a mountain mass intersected by rock joints; or from mile-size ice floes separated by thin ice to the continuous cover of the entire Arctic Ocean.

(3) While *composite materials* had their dawn much before nanotechnology, they still present great opportunities. Their promise for load-bearing aerospace and ship structures resides in their high strength-weight ratio and energy absorption, as well as potentially easier maintenance. Hurdles to overcome exist in processing, fracture and size effect prediction, moisture ingress, and damage detection.

(4) *Nanotechnology* has become a booming field. Design against fracture and debonding of submicrometer metallic *thin films* for electronic components, development of superstiff super-strong *nanocomposites* of low brittleness and wear, exploitation of the symbiotic strength, electronic and thermal properties of carbon *nanotubes*, etc., present unique opportunities for *nanomechanics*.

(5) *Detection of damage* such as cracks and corrosion in aging aircraft, steel bridges, nuclear reactor vessels, ocean structures, etc., is of paramount importance. *Nondestructive* testing requires sophisticated inverse analysis of acoustic wave propagation problems. Despite great advances, much more is needed, not least for fiber and particulate composites where acoustic wave dispersion by inhomogeneities poses hard obstacles to detection.

(6) A related task is the development of sensors, especially *nanosensors*, for "smart" structures and devices that can signal information on their damage, and thus allow automatic structural health monitoring if structural system identification by inverse structural analysis is mastered. *Adaptive smart structures* and devices, capable of controlled expansion, contraction, flexing, and stiffening, will be important for self-deployable space structures, damping of seismic oscillations, etc.

(7) *Chemomechanics*, applied, e.g., to concrete subjected to chemical attack involving diffusion and thermal effects, is a fertile field bound to improve structural durability and serve environmental objectives (e.g., embedding waste glass in concrete). Similar phenomena, such as crystallographic phase conversion, are making polycrystalline *shape-memory* alloys attractive for smart structures. In biology, chemomechanics presents fascinating challenges demanding new multidisciplinary approaches.

(8) *Bio-inspired materials* offer intriguing examples of *mechanical superiority*. The abalone shell achieves its amazing strength, fatigue resistance, and shock resistance by an intricate design and self-assembly of nanoparticles of brittle calcite bound with a small amount of protein-based polymer. Some sponges achieve toughness and robustness by integrating brittle materials over at least seven scales. The combination of strength and deformability of spider thread has not yet been equaled. Such feats demonstrate what is achievable. *Biomechanics*, a long-burgeoning field, appears headed for perpetual growth, with applications to osteoporosis and fracture of bones, large strain of anisotropic blood vessels or soft tissues, etc. This field is today reaching into intriguing questions regarding the cytoskeletons of cells, the conformity and compatibility of biological polymers and proteins, and the mechanical behavior of DNA and related molecules.

(9) *Probabilistic mechanics* and *reliability analysis* have reached a high degree of mathematical sophistication, yet still present great opportunities, especially for extensions of primitive material models to quasi-brittle fracture with localization and scale effects, multiscale and nano-based models, coupling with poromechanics, heterogeneities, diffusion, etc. Advances will especially be required in the understanding of extreme-value statistics of random fields in the context of damage localization, quasi-brittle fracture with size effects, and multiscale models. Because of the large values of empirical safety factors used in mechanical design, rationalizing them with extreme-value based statistical mechanics is an enormously promising prospect.

A host of other challenges could be elaborated on. For all of them, simplicity of modeling will be essential for conquering complexity. Usually, only a few characteristics among many dominate the response of a complex mechanical system. The goal of theory is to identify them and condense them into rationally derived but simple laws describing, at least approximately, the overall behavior of the system. This is what much of materials science has been concerned with.

While simple laws governing the mechanical response of structures will obviously be essential for progress in civil engineering (where thousands of structures, each different, are designed annually), they will be no less important as an optimizing tool in all fields of application of mechanics, including those were only a few new designs appear annually (as in aircraft or automotive engineering). A computer allows a brute-force conquest of complex individual situations, but it is the discovery of a simple mathematical law that lends us general understanding, and thus control.

Progress in developing solid mechanics theories and practical approaches to all these problems is sure to happen eventually, but the question is when and where. An institution, firm, or country that will lead in this pursuit in an aggressive manner will reap many benefits.

Zdeněk P. Bažant

Green's Function for a Closed, Infinite, Circular Cylindrical Elastic Shell

J. G. Simmonds
Professor Emeritus
Fellow ASME
Department of Civil Engineering,
University of Virginia,
Charlottesville, VA 22904
e-mail: jgs@virginia.edu

An acceptable variant of the Koiter–Morley equations for an elastically isotropic circular cylindrical shell is replaced by a constant coefficient fourth-order partial differential equation for a complex-valued displacement-stress function. An approximate formal solution for the associated "free-space" Green's function (i.e., the Green's function for a closed, infinite shell) is derived using an inner and outer expansion. The point wise error in this solution is shown rigorously to be of relative order $(h/a)(1+\sqrt{h/a}|x|)$*, where h is the constant thickness of the shell, a is the radius of the mid surface, and ax is distance along a generator of the mid surface.* [DOI: 10.1115/1.2065627]

To the memory of J. Lyell Sanders, Jr. (1924-1998): friend, colleague, mentor, and shell theorist extraordinary

1 Introduction

Elastically isotropic circular cylindrical shells are basic structural elements and a useful tool in their analysis, e.g., by the boundary element method, is the associated free-space Green's function, i.e., the Green's function for a closed, infinite shell. Herein, we first construct an accurate approximation G_A to the dimensionless Green's function G for the fourth-order, complex-valued, constant-coefficient partial differential equations to which the first-approximation Sanders–Koiter shell equations [1,2] may be reduced for an elastically isotropic material [3]. G_A is expressed as the sum of a *formal* inner solution comprising the shallow shell Green's function developed by Sanders and Simmonds [4] plus a formal outer solution involving easily computed functions, minus an explicit common part.

Next, using Fourier series in the circumferential direction and Fourier transforms in the axial direction, we obtain an integral expression for the nth Fourier component of the *residual* $R \equiv G - G_A$. Then, using simple upper bound estimates, we show that $|R| = \mathrm{O}[(h/a)(1+\sqrt{h/a}|x|)]$. Because G itself is $\mathrm{O}(\sqrt{h/a}|x|)$, the *relative* error in R is $\mathrm{O}(h/a)$.

The present paper presents, in part, a simplification and extension of work by Buchwald [5] who, though not concerned with the construction of a Green's function *per se*, considers the closely related problem of determining the normal mid surface deflection W of an infinite, circular cylindrical shell under single and multiple radial loads. Buchwald solves Morley's eighth-order partial differential equation for W [6,7] by combining Fourier transform methods (following Lighthill [8]) with an inner and outer expansion procedure that takes advantage of the small thickness to radius ratio appearing in Morley's equation. (The *inner region* of a closed, infinite cylindrical shell under a point load is a small neighborhood of the mid surface centered about the point of application of the load; the *outer region* is the remainder of the mid surface.) One of our extensions of Buchwald's work is a new expansion of the outer solution useful for small values of the axial variable.

In place of Morley's equation for W we work with an equally accurate fourth-order partial differential equation for the complex-valued dependent variable $\Psi = W + i\sqrt{A/D}F$, where, conventionally, $A = 1/Eh$, $D = Eh^3/12(1-\nu^3)$, E is Young's modulus, ν is Poisson's ratio, and F is a stress function [3]. This simplifies algebra and the auxiliary equations for stresses and tangential displacements. Our Green's function G satisfied the equation for Ψ with the non-homogeneous term on the right side (that involves both normal and tangential loads) replaced by a certain constant times a two-dimensional delta function.

2 Governing Equations

Let ax and ay denote, respectively, distance along the mid surface in the axial and circumferential directions, let a prime and a dot denote differentiation with respect to x and y, let Δ denote the Laplacian, and let $\{p_x, p_y, p\}$ denote the components of the distributed surface loads in the axial, circumferential, and outward normal directions. Derived from the Sanders–Koiter equations of general first-approximation shell theory [1,2], the basic partial differential equation, given by Eq. (7.1) of [3] with the arbitrary O(1) constant set to zero, reads

$$\mathcal{M}\Psi \equiv \Delta\Delta\Psi + \Psi^{\cdot\cdot} - 4i\lambda^2\Psi'' = (a^3/D)\{\Omega'' + (i/4\lambda^2)[\Delta\Xi - (1+\nu) \times (p_x' + p_y^{\cdot})]\} \tag{1}$$

where $-\infty < x < \infty$, $-\pi < y \leqslant \pi$,

$$4\lambda^2 = a/\sqrt{AD} = \sqrt{12(1-\nu^2)}a/h \tag{2}$$

is a large parameter, and

$$\Omega = \int_0^x \int_0^{\bar{x}} \left(p + \int_0^y p_y d\bar{y} \right) d\hat{x} d\bar{x}, \quad \Xi = \Omega + \int_0^x p_x d\bar{x} + \int_0^y p_y d\bar{y} \tag{3}$$

are load potentials. In obtaining these equations, we have replaced, respectively, the symbols Ψ, F_*, λ, and μ^2 in [3] by $a\Psi$, F, 0, and $2\lambda^2$. Auxiliary equations for stress resultants, stress couples, and displacements, that are needed, for example, to use the boundary element method, may be found in [3] and will not be repeated here.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received November 11, 2004; final manuscript received January 28, 2005. Review conducted by N. Triantafyllides. 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 publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

3 Green's Function

For a closed, infinite, circular cylindrical shell ($-\infty<x<\infty$, $-\pi<y\leqslant\pi$), we define the Green's function $G(x,y;\lambda)$ to be that solution of Eq. (1) with the right side replaced by $4\pi\lambda\delta(x)\delta(y)$ that is 2π periodic in y and that grows, at most, algebraically as $|x|\to\infty$. These conditions determine G uniquely to within rigid-body-like terms. Equivalently, we may consider Eq. (1) in the whole xy plane with the right side replaced by

$$4\pi\lambda\delta(x)\sum_{-\infty}^{\infty}\delta(y-2n\pi)=2\lambda\delta(x)\left(1+2\cos y+2\sum_{2}^{\infty}\cos ny\right). \tag{4}$$

Strictly speaking, this equation must be interpreted in a distributional sense. Alternatively, we could replace the infinite sum on the right by a finite sum from 2 to some positive integer N and at an appropriate stage let $N\to\infty$.

3.1 Physical Interpretation. The two-dimensional delta functions strung out along the y axis *do not* correspond to a simple normal point load. To see exactly what loading could produce G, assume that the in-plane external surface loads are derivable from a load potential such that $p_x=\chi'$ and $p_y=\chi^{\cdot}$. Then from Eq. (1)

$$4\pi\lambda\delta(x)\sum_{-\infty}^{\infty}\delta(y-2\pi n)=(a^3/D)\{\Omega''+(i/4\lambda^2)\Delta[\Xi-(1+\nu)\chi]\}. \tag{5}$$

This equation can be satisfied by taking

$$\Xi=(1+\nu)\chi \text{ and } \Omega=2\pi\lambda(D/a^3)|x|\sum_{-\infty}^{\infty}\delta(y-2n\pi). \tag{6}$$

But $\Xi=\Omega+2\chi$. Thus,

$$(1-\nu)\chi=-2\pi(D/a^3)|x|\sum_{-\infty}^{\infty}\delta(y-2n\pi). \tag{7}$$

The normal pressure is found from the relation $p=\Omega''-\int p_y dy=\Omega''-\chi$. It now follows that $G(x,y;\lambda)$ represents the solution of Eq. (1) for an infinite circular cylindrical shell under the surface loading

$$\begin{Bmatrix}p_x\\p_y\\p\end{Bmatrix}=\frac{2\pi D\lambda}{(1-\nu)a^3}\begin{Bmatrix}-\operatorname{sgn}x\\-|x|\\2(1-\nu)\delta(x)+|x|\end{Bmatrix}\sum_{-\infty}^{\infty}\delta(y-2n\pi). \tag{8}$$

These loads are not self-equilibrating but give rise to a shear force and moment acting over any cross section $x=$constant. This is to be expected given that the dimensionless Green' function for an infinite Euler–Bernoulli beam is $G(x)=(1/12)x^2|x|$, with associated dimensionless shear force $(1/2)\operatorname{sgn}x$ and moment $(1/2)|x|$.

Note that for a 2π periodic of normal point loads along the y axis of the form

$$p(x)=(D/a^3)\delta(x)\sum_{-\infty}^{\infty}\delta(y-2n\pi) \tag{9}$$

the right side of Eq. (1) becomes

$$(1+i/4\lambda^2)\delta(x)\sum_{-\infty}^{\infty}\delta(y-2n\pi)+(i/8\lambda^2)|x|\sum_{-\infty}^{\infty}\delta''(y-2n\pi), \tag{10}$$

which reinforces the point made in the first sentence of this subsection that the Green's function G does not correspond to a simple distribution of normal point loads.

4 Approximate Green's Function

4.1 Inner Solution. With the change of variables

$$\bar{x}=\lambda x,\quad \bar{y}=\lambda y, \tag{11}$$

Eq. (1) takes the form

$$(\bar{\Delta}\bar{\Delta}+\lambda^{-2}\boldsymbol{\partial}^2/\boldsymbol{\partial}\bar{y}^2-4i\boldsymbol{\partial}^2/\boldsymbol{\partial}\bar{x}^2)G=4\pi\lambda^{-1}\delta(\bar{x})\sum_{-\infty}^{\infty}\delta(\bar{y}-2n\lambda\pi),\quad -\infty<\bar{x},\bar{y}<\infty, \tag{12}$$

where $\bar{\Delta}=\boldsymbol{\partial}^2/\boldsymbol{\partial}\bar{x}^2+\boldsymbol{\partial}^2/\boldsymbol{\partial}\bar{y}^2$. If we replace the right side of Eq. (12) by $\lambda\delta(\bar{x})\delta(\bar{y})$ and let $\lambda\to\infty$, we obtain an equation whose solution follows from Eq. (64) of [4] as

$$\Phi=4\pi\lambda w=i^{3/2}[T(\bar{x},\bar{y})+\bar{y}Q(\bar{x},\bar{y})],\quad -\infty<\bar{x},\bar{y}<\infty, \tag{13}$$

where

$$Q=(1/2)[P(\bar{r},\theta)+P(\bar{r},\pi-\theta)] \tag{14}$$

and

$$T=\bar{x}\sinh(i^{1/2}\bar{x})K_0(i^{1/2}\bar{r})+\bar{r}\cosh(i^{1/2}\bar{x})K_1(i^{1/2}\bar{r}). \tag{15}$$

In these equations $i^{1/2}=e^{i\pi/4}$, $i^{3/4}=e^{3i\pi/4}$, $\bar{r}\equiv\sqrt{\bar{x}^2+\bar{y}^2}$, K_0 and K_1 are modified Bessel functions of the second kind (that may be expressed in terms of the real-valued ker and kei functions), and P is a complex-valued function described in detail in the section of [4] labeled **Function P** that begins with Eq. (77). We now define the *approximate* inner Green's function as

$$\bar{G}(\bar{x},\bar{y};\lambda)\equiv\lambda^{-1}\sum_{-\infty}^{\infty}\Phi(\bar{x},\bar{y}-2n\lambda\pi). \tag{16}$$

That Eq. (16) is a formal approximate solution is of no real consequence as we shall show *rigorously* in Sec. 6 that the various approximate Green's functions can be pieced together to approximate the exact Green's function to within a point wise error of $O[(h/a)(1+\sqrt{h/a}|x|)]$.

4.2 Outer Solution. Let

$$x=2\lambda\tilde{x} \text{ and } G=\tilde{G}(\tilde{x},y;\lambda). \tag{17}$$

Then Eq. (1) may be given the form

$$\tilde{G}''''+\tilde{G}''+(1/2)\lambda^{-2}\partial^2\tilde{G}''/\partial\tilde{x}^2+(1/16)\lambda^{-4}\partial^4\tilde{G}/\partial\tilde{x}^4-i\partial^2\tilde{G}/\partial\tilde{x}^2=\delta(\tilde{x})\left(1+2\cos y+2\sum_{2}^{\infty}\cos ny\right). \tag{18}$$

Assume the Fourier expansion

$$\tilde{G}(\tilde{x},y;\infty)=\sum_{0}^{\infty}O_n(x)\cos ny, \tag{19}$$

where the "O" is a mnemonic for "outer." Then for $n=0$ and $n=1$, Eq. (18) yields the algebraic solutions

$$O_0=b_0+(i/2)|\tilde{x}|,\quad O_1=b_1+i|\tilde{x}|, \tag{20}$$

where b_0 and b_1 are arbitrary constants that represent rigid-body-like terms and will be set to zero. If $n\geqslant 2$, Eq. (18) yields the exponentially decaying solutions

$$O_n=-\frac{i^{3/2}\exp(i^{3/2}n\sqrt{n^2-1}|\tilde{x}|)}{n\sqrt{n^2-1}}. \tag{21}$$

Thus,

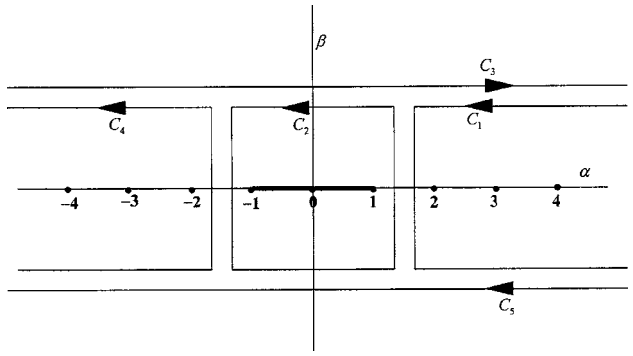


Fig. 1 Diagram for converting the infinite series in Eq. (22) to a finite sum of contour integrals

$$\widetilde{G}(\tilde{x},y;\infty)=i(1/2+\cos y)|\tilde{x}|-i^{3/2}\sum_{2}^{\infty}\frac{\exp(i^{3/2}n\sqrt{n^2-1}|\tilde{x}|)\cos ny}{n\sqrt{n^2-1}}. \tag{22}$$

The series converges absolutely but slowly if $|\tilde{x}|$ is small. We therefore now develop an alternative expansion for $\widetilde{G}(\tilde{x},y;\infty)$ that converges rapidly if $|\tilde{x}|\ll 1$.

4.3 Alternative Expansion for $\widetilde{G}(\tilde{x},y;\infty)$. Using a well-known device, we replace the integer n in the series in Eq. (22) by the complex variable $\gamma=\alpha+i\beta$, multiply the summand by $-(1/2)i\csc\gamma\pi$, and integrate the resulting meromorphic function around a contour C_1 enclosing the points $\gamma=2,3,\ldots$, as shown in Fig. 1. Because the residue of $\pi\csc\gamma\pi$ at $\gamma=n$ is $(-1)^n$, we eliminate this alternating factor by setting

$$\cos ny=(-1)^n\cos n(\pi-|y|). \tag{23}$$

Thus,

$$-i^{3/2}\sum_{2}^{\infty}\frac{\exp(i^{3/2}n\sqrt{n^2-1}|\tilde{x}|)\cos ny}{n\sqrt{n^2-1}}=\int_{C_1}f(\tilde{x},y;\gamma)d\gamma, \tag{24}$$

where

$$f=-\frac{i^{1/2}\exp(i^{3/2}\gamma\sqrt{\gamma^2-1}|\tilde{x}|)\cos\gamma(\pi-|y|)}{2\gamma\sqrt{\gamma^2-1}\sin\gamma\pi}. \tag{25}$$

We make $\sqrt{\gamma^2-1}$ single valued (and analytic) by introducing a cut in the γ plane from $\gamma=-1$ to $\gamma=1$ and choosing that branch that approaches γ as $|\gamma|\to\infty$, as illustrated in Fig. 1. This makes $\gamma\sqrt{\gamma^2-1}$ and f, respectively, *even* and *odd functions* of γ.

Now consider the five distinct, non-overlapping contours, $C_1,\ldots,C_5$ shown in Fig. 1, where C_4 is the mirror image of C_1 about the imaginary axis and C_5 is the negative mirror image of C_3 about the real axis. Because f is an analytic function of γ at all points of the cut γ plane except at $\gamma=\pm n, n\geqslant 2$, we may distort these contours so that parts of them coincide in such a way that

$$\int_{C_1+\cdots+C_5}fd\gamma=0. \tag{26}$$

But, f is odd in γ. Thus,

$$\int_{C_4}f(\gamma)d\gamma=\int_{C_1}f(-\gamma)d(-\gamma)=\int_{C_1}fd\gamma \text{ and } \int_{C_5}f(\gamma)d\gamma$$
$$=-\int_{C_3}f(-\gamma)d(-\gamma)=\int_{C_3}f(\gamma)d\gamma. \tag{27}$$

This yields

$$\int_{C_1}fd\gamma=-(1/2)\int_{C_2}fd\gamma-\int_{C_3}fd\gamma. \tag{28}$$

We now replace the exponential in Eq. (25) by the sum of a hyperbolic cosine and sine, which makes $\sinh(i^{3/2}\gamma\sqrt{\gamma^2-1}|\tilde{x}|)/\gamma\sqrt{\gamma^2-1}$ an analytic function of γ. By the residue theorem

$$\frac{i^{1/2}}{4}\int_{C_2}\frac{\sinh(i^{3/2}\gamma\sqrt{\gamma^2-1}|\tilde{x}|)\cos\gamma(\pi-|y|)}{\gamma\sqrt{\gamma^2-1}\sin\gamma\pi}d\gamma=-i(1/2+\cos y)|\tilde{x}|, \tag{29}$$

so that from Eqs. (22), (28), and (29) we obtain the representation

$$\widetilde{G}(\tilde{x},y;\infty)=\frac{i^{1/2}}{4}\int_{C_2}\frac{\cosh(i^{3/2}\gamma\sqrt{\gamma^2-1}|\tilde{x}|)\cos\gamma(\pi-|y|)}{\gamma\sqrt{\gamma^2-1}\sin\gamma\pi}d\gamma$$
$$-\int_{C_3}f(\tilde{x},y;\gamma)d\gamma. \tag{30}$$

Recalling that the object of this subsection is to obtain a useful representation of $\widetilde{G}(\tilde{x},y;\infty)$ for small $|\tilde{x}|$, we now expand the first integral in Eq. (30) in powers of $\tilde{x}$. Thus,

$$\widetilde{G}(\tilde{x},y;\infty)=i^{-1/2}\sum_{0}^{\infty}a_n(y)(-i\tilde{x}^2)^n/(2n)!-\int_{C_3}f(\tilde{x},y;\gamma)d\gamma, \tag{31}$$

where

$$a_0(y)=\frac{i}{4}\int_{C_2}\frac{\cos\gamma(\pi-|y|)}{\gamma\sqrt{\gamma^2-1}\sin\gamma\pi}d\gamma \tag{32}$$

and

$$a_n(y)=\int_0^1\alpha^{2n-1}(1-\alpha^2)^{n-1/2}\csc\alpha\pi\cos\alpha(\pi-|y|)d\alpha,\quad n\geqslant 1. \tag{33}$$

Note that the $a_n(y)$ are real, even functions of y. Note also that $\int_{C_3}f(0,y;\gamma)d\gamma=0$. To see why, let $\gamma=\alpha+ib$, where $b>0$ is a constant, use the double triangle inequality $\|z_1|-|z_2\|\leqslant|z_1+z_2|\leqslant|z_1|+|z_2|$, the exponential representations for the sine and cosine, and the inequality $|\gamma\sqrt{\gamma^2-1}|\geqslant b(\alpha^2+b^2)$ to conclude that

$$\left|\int_{C_3}fd\gamma\right|\leqslant\frac{e^{-b|y|}(1+e^{-2b(\pi-|y|)})}{2b(1-e^{-2b\pi})}\int_{-\infty}^{\infty}\frac{d\alpha}{\alpha^2+b^2}\to 0 \text{ as } b\to\infty. \tag{34}$$

The easiest way to compute $a_0(y)$ is to return to Eq. (22) and set $\tilde{x}=0$ so that

$$a_0(y)=\sum_{2}^{\infty}\frac{\cos ny}{n\sqrt{n^2-1}}. \tag{35}$$

To improve the convergence of this series and to display explicitly terms up to $O(y^2)$, we set $\cos ny=1-2\sin^2 ny/2$, add and subtract the series $\Sigma_2^\infty(1/n^2)\sin^2 ny/2$, and use series 683 of [9]. Thus,

$$a_0=1+\sum_{2}^{\infty}\frac{1}{n\sqrt{n^2-1}}-\cos y-\frac{\pi|y|}{2}+\frac{y^2}{4}$$
$$-2\sum_{2}^{\infty}\frac{n-\sqrt{n^2-1}}{n^2\sqrt{n^2-1}}\sin^2 ny/2. \tag{36}$$

Because $\sin^2\theta/2\leqslant(1/4)\theta^2$,

Table 1 Taylor series coefficients

y/π	$a_0(y)$	$a_1(y)$	$a_2(y)$	$a_3(y)$
0.00	0.694 22	−0.336 58	−0.047 47	−0.009 24
0.05	0.462 02	−0.301 41	−0.041 26	−0.007 96
0.10	0.260 44	−0.257 93	−0.034 26	−0.006 55
0.15	0.091 14	−0.206 79	−0.026 55	−0.005 02
0.20	−0.045 19	−0.148 81	−0.018 24	−0.003 40
0.25	−0.148 78	−0.084 97	−0.009 46	−0.001 70
0.30	−0.220 78	−0.016 37	−0.000 33	0.000 05
0.35	−0.263 15	0.055 78	0.009 01	0.001 82
0.40	−0.278 59	0.130 16	0.018 43	0.003 59
0.45	−0.270 45	0.205 41	0.027 77	0.005 34
0.50	−0.242 55	0.280 13	0.036 89	0.007 03
0.55	−0.199 09	0.352 93	0.045 65	0.008 65
0.60	−0.144 49	0.422 43	0.053 92	0.010 18
0.65	−0.083 22	0.487 30	0.061 55	0.011 58
0.70	−0.019 65	0.546 30	0.068 44	0.012 84
0.75	0.042 08	0.598 30	0.074 46	0.013 94
0.80	0.098 20	0.642 29	0.079 53	0.014 86
0.85	0.145 45	0.677 43	0.083 56	0.015 60
0.90	0.181 17	0.703 03	0.086 48	0.016 13
0.95	0.203 41	0.718 59	0.088 25	0.016 45
1.00	0.210 95	0.72381	0.088 85	0.016 56

$$\left|\frac{y^2}{4}-2\sum_{2}^{\infty}\frac{n-\sqrt{n^2-1}}{n^2\sqrt{n^2-1}}\sin^2 ny/2\right| \leqslant (1/4)\left(1+2\sum_{2}^{\infty}\frac{1}{\sqrt{n^2-1}(n+\sqrt{n^2-1})}\right)y^2 \equiv Ky^2. \quad (37)$$

For $n\geqslant 2, a_n(y)$ can be computed from Eq. (33) using standard numerical integration formulas. (For $n=1$, use an equation such as (**25.4.36**) of [10] to handle the integrable singularity at $\alpha=1$.) Above is Table 1 for $a_0(y)$ to $a_3(y)$.

To complete our representation of $\tilde{G}(\tilde{x},y;\infty)$ for small $\tilde{x}$, we consider the integral of f over C_3. Since $\mathcal{I}\gamma>0$ on C_3, a part of f may be rewritten as follows:

$$i\csc\gamma\pi\cos\gamma(\pi-|y|)=(e^{i\gamma|y|}+e^{i\gamma(2\pi-|y|)})\sum_{0}^{\infty}e^{i2\gamma n\pi}=e^{i\gamma|y|}+\sum_{1}^{\infty}e^{i\gamma(2n\pi+y)}+\sum_{1}^{\infty}e^{i\gamma(2n\pi-y)}. \quad (38)$$

Thus we may write

$$\int_{C_3} f\,d\gamma = S(\tilde{x},y)+\sum_{1}^{\infty}S(\tilde{x},2n\pi+y)+\sum_{1}^{\infty}S(\tilde{x},2n\pi-y), \quad (39)$$

where

$$S(\tilde{x},y)=\frac{i^{3/2}}{2}\int_{C_3}\frac{\exp(i^{3/2}\gamma\sqrt{\gamma^2-1}|\tilde{x}|+i\gamma|y|)}{\gamma\sqrt{\gamma^2-1}}d\gamma. \quad (40)$$

The form of the right side of Eq. (39) shows that Eq. (40) represents the contribution of the delta function at the origin to the integral of f over C_3.

To obtain an expansion of $S(\tilde{x},y)$ in powers of $\tilde{x}$, we introduce the change of variables

$$\gamma=\frac{i^{1/4}\zeta}{\sqrt{|\tilde{x}|}},\quad \mu=\frac{i^{1/4}|y|}{2\sqrt{|\tilde{x}|}}, \quad (41)$$

write the argument of the exponent in Eq. (40) as

$$i^{3/2}\gamma\sqrt{\gamma^2-1}|\tilde{x}|+i\gamma|y|=-\zeta^2\sqrt{1+i^{3/2}|\tilde{x}|/\zeta^2}+i2\mu\zeta=-\zeta^2+\zeta^2(1-\sqrt{1+i^{3/2}|\tilde{x}|/\zeta^2})+i2\mu\zeta \quad (42)$$

and use the expansion

$$\frac{\exp[\zeta^2(1-\sqrt{1+i^{3/2}|\tilde{x}|/\zeta^2})]}{\sqrt{1+i^{3/2}|\tilde{x}|/\zeta^2}}=1-\frac{i^{3/2}}{2}\left(1+\frac{1}{\zeta^2}\right)|\tilde{x}|-\frac{i}{8}\left(1+\frac{3}{\zeta^2}+\frac{3}{\zeta^4}\right)\tilde{x}^2+\cdots. \quad (43)$$

Insertion of Eq. (43) into Eq. (40) leaves us with integrals to evaluate of the form

$$I_{2n}(\mu)\equiv\int_{C_3'}\frac{e^{-\zeta^2+i2\mu\zeta}}{\zeta^{2n}}d\zeta, \quad (44)$$

where C_3' is the image of C_3 in the ζ plane.

Because $I_{2n}(\infty)=0$ for all n and because Eq. (44) may be differentiated n times under the integral sign, $I_{2n}(\mu)$ may be expressed as an n-fold integral. In the standard way, this iterated integral may be replaced by a single integral that may be expressed in closed form

$$I_{2n}(\mu)=\frac{(2i)^{2n}}{(2n-1)!}\int_{\mu}^{\infty}(t-\mu)^{2n-1}\left(\int_{C_3'}e^{-\zeta^2+i2\mu\zeta}d\zeta\right)dt = \frac{4^n(-1)^n\sqrt{\pi}}{(2n-1)!}\int_{\mu}^{\infty}(t-\mu)^{2n-1}e^{-t^2}dt = 2^{2n-1}(-1)^n\mathrm{i}^{2n-1}\operatorname{erfc}\mu, \quad (45)$$

where $\mathrm{i}^n\operatorname{erfc}\mu$ is the nth repeated integral of the complementary error function. (See equation (**7.2.3**) of [10] and note that $\mathrm{i}\neq i$.) Hence,

$$S(\tilde{x},y)=-i^{5/4}\pi\sqrt{|\tilde{x}|}[\mathrm{i}\operatorname{erfc}\mu-(1/2)i^{3/2}(\mathrm{i}\operatorname{erfc}\mu-4\mathrm{i}^3\operatorname{erfc}\mu)|\tilde{x}|-(1/8)i(\mathrm{i}\operatorname{erfc}\mu-12\mathrm{i}^3\operatorname{erfc}\mu+48\mathrm{i}^5\operatorname{erfc}\mu)\tilde{x}^2+\cdots]. \quad (46)$$

To use existing tables to evaluate $S(\tilde{x},y)$, it is necessary to express each iterated complementary error function in terms of $\operatorname{erfc}\mu$ and $e^{-\mu^2}$. By Eqs. (**7.2.1**), (**7.2.5**), and (**7.2.6**) of [10],

$$S(\tilde{x},y)=-i^{5/4}\pi\sqrt{|\tilde{x}|}\left\{\frac{e^{-\mu^2}}{\sqrt{\pi}}-\mu\operatorname{erfc}\mu-\frac{i^{3/2}}{6}\left[\frac{e^{-\mu^2}}{\sqrt{\pi}}(1-2\mu^2)+2\mu^3\operatorname{erfc}\mu\right]|\tilde{x}|+\frac{i}{80}\left[\frac{e^{-\mu^2}}{\sqrt{\pi}}(1+\mu^2-2\mu^4)-\mu(5-4\mu^4)\operatorname{erfc}\mu\right]\tilde{x}^2+\cdots\right\} \quad (47)$$

Because $\mathrm{i}^n\operatorname{erfc}\mu$ decays rapidly as $\mu=i^{1/4}|y|/2\sqrt{|\tilde{x}|}$ grows, it is clear from Eqs. (39) and (46) that $\int_{C_3}f\,d\gamma\approx S(\tilde{x},y)$, i.e., the contribution of the two-dimensional delta function at the origin dominates all the others strung out along the y axis.

5 Matching the Inner and Outer Solution

To study how well the inner and outer solutions, $\bar{G}$ and $\tilde{G}$, approximate the exact Green's function G, we introduce the *intermediate variables* x and $\hat{y}$ in terms of which, by Eqs. (11), (17), and (38)

$$\bar{x}=\lambda x,\quad \bar{y}=\lambda^{1/2}\hat{y},\quad \tilde{x}=\lambda^{-1}x/2,\quad y=\lambda^{-1/2}\hat{y}\ \text{and}\ \mu=i^{1/2}\hat{y}/2\sqrt{x}. \quad (48)$$

Assume that G has an expansion of the form

$$G(x,y;\lambda)=G(x,\lambda^{-1/2}\hat{y};\lambda)=\lambda^{-1/2}[G_0(x,\hat{y})+\lambda^{-1}G_1(x,\hat{y})+O(\lambda^{-2})]. \tag{49}$$

Then

$$\hat{\mathcal{L}}G_0\equiv G_0^{####}-4iG_0''=4\pi\delta(x)\sum_{-\infty}^{\infty}\delta(\hat{y}-2\lambda^{1/2}n\pi)$$

$$\text{and } \hat{\mathcal{L}}G_1=-(2G_0^{''##}+G_0^{##}), \tag{50}$$

where $\#=\partial/\partial\hat{y}$.

Now consider the inner Green's function $\bar{G}(\bar{x},\bar{y};\lambda)$ defined by Eq. (16). If we introduce the intermediate variables x and $\bar{y}$ defined in Eq. (48) and set

$$\bar{G}(\lambda x,\lambda^{1/2}\hat{y};\lambda)=\lambda^{-1/2}[\bar{G}_0(x,\hat{y})+\lambda^{-1}\bar{G}_1(x,\hat{y})+O(\lambda^{-2})], \tag{51}$$

then, by Eq. (12) and Eq. (50)

$$\hat{\mathcal{L}}\bar{G}_0=4\pi\delta(x)\sum_{-\infty}^{\infty}\delta(\hat{y}-2\lambda^{1/2}n\pi) \text{ and } \hat{\mathcal{L}}\bar{G}_1=-2\bar{G}_0^{''##}. \tag{52}$$

Further, assume that the approximate outer solution $\tilde{G}(\bar{x},y;\infty)$, when written in terms of the intermediate variables, has an expansion of the form

$$\tilde{G}(\lambda^{-1}x/2,\lambda^{-1/2}\hat{y};\infty)=\lambda^{-1/2}[\tilde{G}_0(x,\hat{y})+\lambda^{-1}\tilde{G}_1(x,\hat{y})+O(\lambda^{-2})]. \tag{53}$$

Then by Eq. (4) and Eq. (18),

$$\hat{\mathcal{L}}\tilde{G}_0=4\pi\delta(x)\sum_{-\infty}^{\infty}\delta(\hat{y}-2\lambda^{1/2}n\pi) \text{ and } \hat{\mathcal{L}}\tilde{G}_1=-\tilde{G}_0^{##}. \tag{54}$$

Comparing Eq. (50) with Eqs. (52) and (54), we see that $G_0=(1/2)(\bar{G}_0+\tilde{G}_0)$ and $G_1=\bar{G}_1+\tilde{G}_1$. Moreover, from Eqs. (13)–(15) and Eqs. (86) and (87) of [4], Eqs. (31), (36), (37), and (47),

$$\bar{G}_0(x,\hat{y})=\tilde{G}_0(x,\hat{y})=\frac{i^{3/2}\pi}{2}|\hat{y}|+i^{5/4}\pi\sqrt{\frac{|x|}{2}}\left(\frac{e^{-\mu^2}}{\sqrt{\pi}}-\mu\,\text{erfc}\,\mu\right)$$

$$\equiv\hat{\Phi}(x,\hat{y}),\quad |\hat{y}|\leqslant\lambda^{1/2}\pi. \tag{55}$$

Thus, $G=\lambda^{-1/2}\{[\bar{G}_0+\lambda^{-1}\bar{G}_1+O(\lambda^{-2})]+[\tilde{G}_0+\lambda^{-1}\tilde{G}_1+O(\lambda^{-2})]\}-\hat{G}$, where

$$\hat{G}(x,\hat{y};\lambda)\equiv\lambda^{-1/2}\sum_{-\infty}^{\infty}\hat{\Phi}(x,\hat{y}-2\lambda^{1/2}n\pi). \tag{56}$$

That is,

$$G(x,y;\lambda)=\bar{G}(\lambda x,\lambda y;\lambda)+\tilde{G}(\lambda^{-1}x/2,y;\lambda)+\hat{G}(x,\lambda^{1/2}y;\lambda)+O(\lambda^{-2}). \tag{57}$$

In the next section we prove that this formal error estimate is rigorous.

6 Rigorous Estimate of the Error in the Approximate Green's Function

Let $R\equiv G-\bar{G}-\tilde{G}+\hat{G}$ and $\mathcal{L}\equiv\partial^4/\partial y^4-4i\lambda^2\partial^2/\partial x^2$. Then the various partial differential equations satisfied by G, $\bar{G}$, $\tilde{G}$, and $\hat{G}$ may be summarized as

$$\mathcal{L}\{G,\bar{G},\tilde{G},\hat{G}\}+2\{G,\bar{G},0,0\}^{'''}+\{G,\bar{G},0,0\}^{''''}+\{G,0,\tilde{G},0\}^{\cdot\cdot}$$

$$=4\pi\lambda\,\delta(x)\sum_{-\infty}^{\infty}\delta(y-2n\pi). \tag{58}$$

By Eqs. (1), (4), and (58)

$$\mathcal{M}R=\bar{G}^{\cdot\cdot}+2\tilde{G}^{''\cdot\cdot}+\tilde{G}^{''''}-\hat{G}^{\cdot\cdot}-2\hat{G}^{''\cdot\cdot}-\hat{G}^{''''}. \tag{59}$$

Now assume that all functions have Fourier series expansions of the form

$$R=R_0+R_1\cos y+\sum_2^{\infty}R_n\cos ny,\ \text{etc.} \tag{60}$$

Then from Eq. (59)

$$R_n''''-2n^2R_n''+n^4R_n-n^2R_n-4i\lambda^2R_n''=-n^2\bar{G}_n-2n^2\tilde{G}_n''+\tilde{G}_n''''+n^2\hat{G}_n+2n^2\hat{G}_n''-\hat{G}_n''''. \tag{61}$$

Let $f(\xi)=\int_{-\infty}^{\infty}e^{i\xi x}F(x)dx$ and $F(x)=(1/2\pi)\int_{-\infty}^{\infty}e^{-ix\xi}f(\xi)d\xi$ denote, respectively, the Fourier transform of a function and its inverse. Taking the Fourier transform of both sides of Eq. (61), we obtain

$$D(\xi,n;\lambda)r_n=-n^2\bar{g}_n+2n^2\xi^2\bar{g}_n+\xi^4\bar{g}_n+n^2\hat{g}_n-2n^2\xi^2\hat{g}_n-\xi^4\hat{g}_n, \tag{62}$$

where

$$D=(\xi^2+n^2)^2-n^2+4i\lambda^2\xi^2. \tag{63}$$

Taking the Fourier transform of both sides of Eq. (58) with the right side replaced by the right side of Eq. (4), we have

$$\bar{D}(\xi,n;\lambda)\bar{g}_n=4\lambda,\quad \tilde{D}(\xi,n;\lambda)\tilde{g}_n=4\lambda,\quad \hat{D}(\xi,n;\lambda)\hat{g}_n=4\lambda, \tag{64}$$

where

$$\bar{D}=(\xi^2+n^2)^2+4i\lambda^2\xi^2,\quad \tilde{D}=n^4-n^2+4i\lambda^2\xi^2,\quad \hat{D}=n^4+4i\lambda^2\xi^2. \tag{65}$$

If we substitute Eqs. (64) and (65) into Eq. (62), we find after a bit of algebra that

$$r_n=\frac{4\lambda n^2N}{D\bar{D}\tilde{D}\hat{D}}, \tag{66}$$

where

$$N=\xi^2(\xi^2+2n^2)(D+\hat{D}). \tag{67}$$

Thus, by the triangle inequality

$$|R_n|=\left|(4/\pi)\lambda n^2\int_0^{\infty}e^{-ix\xi}\frac{N}{D\bar{D}\tilde{D}\hat{D}}d\xi\right|$$

$$\leqslant(4/\pi)\lambda n^2\int_0^{\infty}\frac{\xi^2(\xi^2+2n^2)}{|\bar{D}||\tilde{D}|}\left(\frac{1}{|\hat{D}|}+\frac{1}{|D|}\right)d\xi. \tag{68}$$

Because $|a+ib|=\sqrt{a^2+b^2}\geqslant(1/2)(|a|+|b|)$ and $n^2(n^2-1)\geqslant(3/4)n^4$ if $n\geqslant2$,

$$|D|\geqslant(1/2)[\xi^4+2\xi^2n^2+n^2(n^2-1)+4\lambda^2\xi^2]\geqslant(3/8)[n^4+(16/3)\lambda^2\xi^2]\geqslant(3/8)(n^4+4\lambda^2\xi^2). \tag{69}$$

Moreover,

$$|\hat{D}|\geqslant(1/2)(n^4+4\lambda^2\xi^2). \tag{70}$$

Thus,

$$|R_n|\leqslant\left(\frac{56}{3\pi}\right)\lambda n^2\int_0^{\infty}\frac{\xi^2(\xi^2+2n^2)d\xi}{|\bar{D}||\tilde{D}|(n^4+4\lambda^2\xi^2)},\quad n\geqslant2. \tag{71}$$

Consider the integral

$$I_1 \equiv \int_0^\infty \frac{\xi^4 d\xi}{|\bar{D}||\tilde{D}|(n^4+4\lambda^2\xi^2)}. \tag{72}$$

Because

$$|\bar{D}| \geqslant \xi^2(\xi^2+4\lambda^2) \text{ and } |\tilde{D}| \geqslant 4\lambda^2\xi^2,$$

$$I_1 \leqslant \frac{1}{4\lambda^2 n^4}\int_0^\infty \frac{d\xi}{\xi^2+4\lambda^2} = \frac{\pi}{16\lambda^3 n^4}, \quad n \geqslant 2. \tag{73}$$

Now consider the integral

$$I_2 \equiv \int_0^\infty \frac{\xi^2 d\xi}{|\bar{D}||\tilde{D}|(n^4+4\lambda^2\xi^2)}. \tag{74}$$

Because

$$|\bar{D}| \geqslant (1/2)(n^4+4\lambda^2\xi^2) \text{ and}$$

$$|\tilde{D}| \geqslant (1/2)[(3/4)n^4+4\lambda^2\xi^2] \geqslant (3/8)(n^4+4\lambda^2\xi^2) \tag{75}$$

if $n \geqslant 2$

$$I_2 \leqslant \frac{16}{3}\int_0^\infty \frac{\xi^2 d\xi}{(n^4+4\lambda^2\xi^2)^3} = \frac{2}{3\lambda^3 n^6}\int_0^\infty \frac{u^2 du}{(1+u^2)^3} = \frac{\pi}{8\lambda^3 n^6}, \quad n \geqslant 2. \tag{76}$$

Altogether, from Eqs. (71)–(76), we have

$$|R_n| \leqslant \frac{115}{24\lambda^2 n^2}, \quad n \geqslant 2. \tag{77}$$

From Eqs. (4) and (58), the axisymmetric Fourier components $G_0=\bar{G}_0$ and $\tilde{G}_0=\hat{G}_0$, so $R_0=0$. In the residual $R_1=G_1-\bar{G}_1-\tilde{G}_1+\hat{G}_1$, $|G_1-\tilde{G}_1|$ can be easily computed. Thus

$$G_1 = -\frac{1}{1+2i\lambda^2}\left(\frac{e^{-\lambda p|x|}}{p}+\lambda|x|\right), \quad p = 2i^{1/2}[1-(i/2)\lambda^{-2}]^{1/2} \tag{78}$$

and

$$\tilde{G}_1 = (i/2)\lambda^{-1}|x| \tag{79}$$

so that

$$|G_1(x)-\tilde{G}_1(x)| \leqslant \frac{1}{4\lambda^2}+\frac{|x|}{4\lambda^3}. \tag{80}$$

Although the last term on the right is not uniformly $O(\lambda^{-3})$, it is *uniformly small* of $O(\lambda^{-2})$ compared to $|G_1(x)|$ which is $O(\lambda^{-1}|x|)$.

The easiest way to obtain an upper bound on $|\hat{G}_1-\bar{G}_1|$ is to consider the inverse Fourier transform. Thus

$$|\hat{G}_1(x)-\bar{G}_1(x)| \leqslant (1/\pi)\int_0^\infty |\hat{g}_1(\xi)-\bar{g}_1(\xi)|d\xi$$

$$\leqslant \frac{4\lambda}{\pi}\int_0^\infty \frac{\xi^4+2\xi^2}{|(\xi^2+1)^2+4i\lambda^2\xi^2||1+4i\lambda^2\xi^2|}d\xi$$

$$\equiv I_3+I_4. \tag{81}$$

Take $|1+4i\lambda^2\xi^2| \geqslant 4\lambda^2\xi^2$ in both I_3 and I_4. In I_3, I_4, respectively, take

$$|(\xi^2+1)^2+4i\lambda^2\xi^2| \geqslant (1/2)\xi^2(\xi^2+4\lambda^2) \tag{82}$$

and

$$|(\xi^2+1)^2+4i\lambda^2\xi^2| \geqslant (1/2)(1+4\lambda^2\xi^2). \tag{83}$$

Thus

$$I_3 \leqslant \frac{2}{\pi\lambda}\int_0^\infty \frac{d\xi}{\xi^2+4\lambda^2} = \frac{1}{2\lambda^2} \text{ and } I_4 \leqslant \frac{4}{\pi\lambda}\int_0^\infty \frac{d\xi}{1+4\lambda^2\xi^2} = \frac{1}{\lambda^2}. \tag{84}$$

Altogether,

$$|R(x)| \leqslant |R_1(x)| + \sum_2^\infty |R_n| \leqslant \frac{1}{4\lambda^2}\left[7+\frac{|x|}{\lambda}+\frac{115}{6}\left(\frac{\pi^2}{6}-1\right)\right]$$

$$\leqslant \frac{4.84\cdots}{\lambda^2}+\frac{|x|}{\lambda^3}. \tag{85}$$

References

[1] Sanders, J. L., Jr., 1959, "An Improved First-Approximation Theory for Thin Shells," NASA Report No. 24.

[2] Koiter, W. T., 1959, "A Consistent First Approximation in the Theory of Elastic Shells," *Proceedings of the Symposium on the Theory of Thin Elastic Shells*, W. T. Koiter ed., North–Holland, Amsterdam, pp. 12–33.

[3] Simmonds, J. G., 1966, "A Set of Simple, Accurate Equations for Circular Cylindrical Elastic Shells," Int. J. Solids Struct., **2**, pp. 525–541.

[4] Sanders, J. L., Jr., and Simmonds, J. G., 1970, "Concentrated Forces on Shallow Cylindrical Shells," ASME J. Appl. Mech., **37**, pp. 367–373.

[5] Buchwald, V. T., 1967, "Some Problems of Thin Circular Cylindrical Shells, I," J. Math. Phys., **46**, pp. 237–252.

[6] Morley, L. S. D., 1959, "An Improvement on Donnell's Approximation for Thin-Walled Circular Cylinders," Q. J. Mech. Appl. Math., **12**, pp. 89–99.

[7] Morley, L. S. D., 1960, "The Thin-Walled Circular Cylinder Subjected to Concentrated Radial Loads," Q. J. Mech. Appl. Math., **13**, pp. 24–37.

[8] Lighthill, J., 1960, *Fourier Analysis and Generalized Functions*, Cambridge University Press, Cambridge.

[9] Jolley, L. B. W., 1961, *Summation of Series*, 2nd revised ed., Dover, New York.

[10] Abramowitz, M., and Stegun, I., eds., 1965, *Handbook of Mathematical Functions*, National Bureau of Standards.

Rayleigh's Quotient of Linear Piezoelectricity and Its Use in the Approximate Calculation of the Natural Frequencies of Piezoelectric Continua

Piotr Cupiał
Assistant Professor
Cracow University of Technology,
Institute of Applied Mechanics,
Al. Jana Pawla II-go 37,
31-864, Krakow, Poland
e-mail: cupial@mech.pk.edu.pl

The paper discusses an extension to the linear theory of piezoelectricity of the Rayleigh quotient used in the analysis of the properties and the approximate calculation of the natural frequencies of elastic continua. It is shown that for a piezoelectric continuum an infinite number of equivalent expressions can be obtained which generalize the classical Rayleigh quotient. The stationarity conditions of any of these quotients under additional constraints imposed by the Gauss equation of electrostatics and the prescribed natural electrical boundary condition are shown to result in the complete set of the governing equations of the free vibration problem of a piezoelectric continuum. The general results discussed in the paper are illustrated by the approximate calculation of the natural frequencies of a piezoelectric rod by the Rayleigh–Ritz method. Unlike in the case of elastic structures, no monotonic convergence of the approximate frequencies is guaranteed for a piezoelectric continuum, the property which can be explained using the introduced Rayleigh quotients. [DOI: 10.1115/1.2065667]

1 Introduction

Since its introduction by Lord Rayleigh [1], the Rayleigh quotient has proved a very valuable concept used in the study of the properties of the natural frequencies of vibrating multiple degree-of-freedom and continuous elastic systems. In particular, the Rayleigh quotient can be used in the approximate calculation of the natural frequencies of elastic systems using the Rayleigh–Ritz method. The properties of the Rayleigh quotient for elastic continuous systems have been discussed extensively by Courant and Hilbert [2], Mikhlin [3], Strang and Fix [4] and Washizu [5]. It is known from these studies that the approximate natural frequencies calculated using the Rayleigh–Ritz method approach the exact values monotonically and the approximate frequencies are always higher than the exact values.

This paper will discuss an extension of classical Rayleigh's concept to the case of the free vibrations of piezoelectric continua. The theory of piezoelectricity has been discussed at length by Cady [6] and Mason [7]. Extensive information about the historical development of the theory can be found in Ref. [6]. Both Refs. [6,7] provide solutions of some vibration problems for piezoelectric elements such as rods, beams and plates. Even though the solutions discussed in Refs. [6,7] have proved very useful in the practical design of piezoelectric resonators, no full set of coupled equations of piezoelectricity has been used in these references (no explicit use has been made of the Gauss equation of electrostatics). The consistent formulation of the free vibration problem of an infinite piezoelectric plate has been considered by Lawson [8], followed by more extensive studies by Mindlin and Tiersten. The derivation of the full set of the governing equations of piezoelectricity and some solutions of coupled electromechanical problems are discussed by Tiersten [9].

When studying the vibrations of piezoelectric continua, for which the mechanical and electrical phenomena are coupled, energy concepts are of much help. Energy concepts were first used in the theory of piezoelectricity to define the electromechanical coupling factor, which was defined by Mason [7] for the case of forced vibrations excited by an applied voltage as the "square root of the ratio of the energy stored in mechanical form, for a given type of displacement, to the total input electrical energy obtained from the input battery." The variational formulation of the dynamical theory of linear piezoelectricity has been discussed by Tiersten [9] and Allik and Hughes [10]. Various formulations of the variational principles of the static theory of piezoelectricity can be found in [11]. The variational principles of linear dynamic thermo-piezoelectricity are discussed by Nowacki [12] and He [13]. Yang and Batra [14] consider an interesting problem of obtaining the conservation laws of linear piezoelectricity using the variational formulation and the Noether theorem.

One particular form of the Rayleigh quotient has been used in the calculation of the group velocity in functionally graded piezoelectric plates by Liu and Xi [15] and in cylinders by Han and Liu [16]. The expression for the Rayleigh quotient was obtained there using the discretized matrix equations of wave propagation. Two equivalent expressions of the Rayleigh quotient were derived in the author's paper [17] using both the equations of motion of a piezoelectric continuum and the balance of energy, and they were then used to verify the natural frequencies and mode shapes of a rectangular piezoelectric plate. In the present paper it is shown that an infinite number of equivalent expressions of the Rayleigh quotient can be obtained in linear piezoelectricity. The stationarity of these quotients alone is not equivalent to the defining equations

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received November 15, 2004; final manuscript received May 20, 2005. Review conducted by I. Mezic. 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 this paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

of the free vibration problem of a piezoelectric continuum. It is shown in the paper that the governing equations and the respective natural boundary conditions can be obtained using the stationarity condition of any of the introduced quotients with the Gauss equation of electrostatics and the prescribed natural electrical boundary condition added as constraints. The theoretical results will be illustrated by the approximate calculation of the natural frequencies of the longitudinal vibrations of a piezoelectric rod by the Rayleigh–Ritz method and the study of the character of convergence of the frequencies.

2 Extension of Rayleigh's Quotient to Linear Piezoelectricity and the Stationarity Principle

2.1 Definition of Rayleigh's Quotient of Linear Piezoelectricity. The equations of the free vibration of a piezoelectric continuum are discussed in full detail by Tiersten [9]. They consist of the classical equations of free vibrations $(1)_1$ and the Gauss equation (charge equation) of electrostatics $(1)_2$:

$$\sigma_{ij,j} = \rho \ddot{u}_i, \quad D_{i,i} = 0. \tag{1}$$

Here, σ_{ij} are the components of the symmetric stress tensor, u_i are the components of the displacement vector, ρ is the mass density and D_i stand for the components of the electric displacement vector (a full list of symbols used in the paper is provided in Nomenclature). The spatial partial derivative is denoted by a comma and the time derivative by a dot. The summation convention is used throughout the paper and summation over the range 1–3 is implied for a repeated index. The physical equations of linear piezoelectricity are discussed in Refs. [6,7,9,12], and have the following form:

$$\begin{cases} \sigma_{ij} = c_{ijk\ell}\varepsilon_{k\ell} - e_{kij}E_k \\ D_i = e_{ik\ell}\varepsilon_{k\ell} + \kappa_{ik}E_k \end{cases}. \tag{2}$$

Here $c_{ijk\ell}$ are the components of the elastic stiffness tensor, κ_{ik} stand for the dielectric constants and $e_{ik\ell}$ are the piezoelectric constants that couple the mechanical and electric fields. The results discussed in the present paper are also valid in the case when material properties change from point to point within the piezoelectric body and are functions of the spatial coordinates $x = (x_1, x_2, x_3)$ (nonhomogeneous case which is of importance, e.g., in the study of functionally graded piezoelectric elements), even though in order to make formulas more concise this dependence will not be written explicitly. In Eq. (2) ε_{ij} are the components of the Cauchy strain tensor calculated as:

$$\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}). \tag{3}$$

In the classical theory of piezoelectricity a quasi-static assumption is used, which is discussed in more detail in Ref. [9] and which relies on the fact that the mechanical wavelengths are much shorter than the lengths of the electromagnetic waves of the same frequency. According to the quasi-static approximation, the components of the electric field vector E_k can be expressed in terms of the electrostatic potential ϕ:

$$E_k = -\phi_{,k}. \tag{4}$$

The components of the elastic stiffness tensor $c_{ijk\ell}$, piezoelectric tensor e_{kij}, and dielectric constants κ_{ik} satisfy the following symmetry conditions ([9]):

$$c_{ijk\ell} = c_{jik\ell} = c_{ij\ell k} = c_{k\ell ij}, \quad e_{kij} = e_{kji}, \quad \kappa_{ik} = \kappa_{ki}. \tag{5}$$

Combining Eqs. (1)–(5), one obtains the following governing equations of the free vibration problem expressed in terms of displacements and the electrostatic potential (which are discussed in more detail by Tiersten [9] and Nowacki [12]):

$$\begin{cases} (c_{ijk\ell}u_{k,\ell} + e_{kij}\phi_{,k})_{,j} = \rho \ddot{u}_i \\ (e_{ik\ell}u_{k,\ell} - \kappa_{ik}\phi_{,k})_{,i} = 0 \end{cases} \quad \text{in } V. \tag{6}$$

Equations (6) are solved under the appropriate boundary conditions prescribed on the boundary ∂V, which for the free vibration problem take the form:

$$\begin{cases} u_i = 0 (\text{on } \partial V_1), & \sigma_{ij}n_j = 0 (\text{on } \partial V_2), & \partial V_1 \cap \partial V_2 = \varnothing, & \partial V_1 \cup \partial V_2 = \partial V, \\ \phi = 0 (\text{on } \partial V_3), & \mathrm{D}_i n_i = 0 (\text{on } \partial V_4), & \partial V_3 \cap \partial V_4 = \varnothing, & \partial V_3 \cup \partial V_4 = \partial V. \end{cases} \tag{7}$$

Using the language of the calculus of variations, Eqs. $(7)_{1,3}$ ($u_i = 0$, $\phi = 0$) define the so-called essential boundary conditions, whereas the remaining two conditions $(7)_{2,4}$ are the natural boundary conditions.

The governing Eqs. (6) with the boundary conditions (7) can be used to extend to linear piezoelectricity the classical concept of the Rayleigh quotient. To this end, let us assume that the piezoelectric continuum undergoes free vibration, in which case:

$$u_i(x,t) = U_i^{(n)}(x)\sin(\omega_n t + \alpha), \quad \phi(x,t) = \Phi^{(n)}(x)\sin(\omega_n t + \alpha), \tag{8}$$

where α is an arbitrary phase shift. Using Eq. (8) in Eq. $(6)_1$, multiplying this equation by $U_i^{(n)}$ and integrating over the volume of the piezoelectric body one obtains:

$$\int_V (c_{ijk\ell}U_{k,\ell}^{(n)} + e_{kij}\Phi_{,k}^{(n)})_{,j}U_i^{(n)}dV = -\omega_n^2 \int_V \rho U_i^{(n)}U_i^{(n)}dV. \tag{9}$$

Equation (9) can be written in the following equivalent form:

$$\int_V [(c_{ijk\ell}U_{k,\ell}^{(n)} + e_{kij}\Phi_{,k}^{(n)})U_i^{(n)}]_{,j}dV - \int_V (c_{ijk\ell}U_{k,\ell}^{(n)} + e_{kij}\Phi_{,k}^{(n)})U_{i,j}^{(n)}dV = -\omega_n^2 \int_V \rho U_i^{(n)}U_i^{(n)}dV. \tag{10}$$

Using the Gauss theorem of calculus, Eq. (10) is further transformed to the form:

$$\int_{\partial V} (c_{ijk\ell}U_{k,\ell}^{(n)} + e_{kij}\Phi_{,k}^{(n)})U_i^{(n)}n_j d\partial V - \int_V (c_{ijk\ell}U_{k,\ell}^{(n)} + e_{kij}\Phi_{,k}^{(n)})U_{i,j}^{(n)}dV = -\omega_n^2 \int_V \rho U_i^{(n)}U_i^{(n)}dV. \tag{11}$$

Making use of boundary conditions $(7)_{1,2}$ the first term is equal to zero and expression (11) becomes

$$\int_V (c_{ijk\ell}U^{(n)}_{k,\ell} + e_{kij}\Phi^{(n)}_{,k})U^{(n)}_{i,j}dV = \omega_n^2 \int_V \rho U_i^{(n)}U_i^{(n)}dV. \quad (12)$$

Let us now turn to Eq. $(6)_2$, which is multiplied by $\Phi^{(n)}$ and integrated over the volume of the piezoelectric body. It is important to point out that the resulting equation can be multiplied by an arbitrary real constant f without changing the result. As a consequence, the following equation holds true:

$$f\int_V (e_{ik\ell}U^{(n)}_{k,\ell} - \kappa_{ik}\Phi^{(n)}_{,k})_{,i}\Phi^{(n)}dV = 0. \quad (13)$$

Proceeding in exactly the same manner as in deriving Eq. (12), Eq. (13) can be transformed to the following final form:

$$f\int_V (e_{ik\ell}U^{(n)}_{k,\ell} - \kappa_{ik}\Phi^{(n)}_{,k})\Phi^{(n)}_{,i}dV = 0. \quad (14)$$

Interchanging the summation indices in the second term of Eq. (12) and subtracting Eq. (14) from Eq. (12), one arrives at the following expression for the square of the circular natural frequency:

$$\omega_n^2 = \frac{\int_V [c_{ijk\ell}U^{(n)}_{i,j}U^{(n)}_{k,\ell} + f\kappa_{ik}\Phi^{(n)}_{,i}\Phi^{(n)}_{,k} + (1-f)e_{ik\ell}\Phi^{(n)}_{,i}U^{(n)}_{k,\ell}]dV}{\int_V \rho U_i^{(n)}U_i^{(n)}dV}. \quad (15)$$

Equation (15) is the sought generalization to linear piezoelectricity of Rayleigh's quotient of the free vibration of elastic continua. Since f is an arbitrary constant, an infinite number of equivalent expressions of the quotient have been obtained. For $f=1$ the numerator of Eq. (15) is equal to the overall internal energy (the strain energy and the energy of the electric field) of the piezoelectric continuum, which is non-negative ([9,12]). Therefore, the natural frequencies are real.

In the author's paper [17] expressions equal to quotients (15) corresponding to the values of the tracing constant $f=0$ and $f=1$ have been used to verify the consistency of the natural frequencies and mode shapes of a three-dimensional piezoelectric plate calculated by solving the eigenvalue problem. The mode shapes found by solving the eigenvalue problem were introduced into quotient (15) and the natural frequencies where calculated in an independent way. For both values of the tracing constant used, the same frequency value was obtained, in perfect agreement with the frequencies calculated as eigenvalue solutions.

In the present paper the expression of the Rayleigh quotient has been derived using the governing equations. It is also possible to arrive at the corresponding quotient from the energy balance of a piezoelectric continuum, using the approach discussed in the author's paper [17].

2.2 The Stationarity Principle. Rayleigh's quotient is of much help in studying the properties of the natural frequencies of lumped and continuous parameter systems. It is well known that in the case of vibrations of elastic continua the stationarity of Rayleigh's quotient for all variations of displacements which satisfy the essential boundary conditions of the problem is equivalent to solving the eigenvalue free vibration problem. Moreover, Rayleigh's quotient is used in the proof of the monotonic convergence of the approximate frequencies calculated using the Rayleigh–Ritz method, as has been discussed by Courant and Hilbert [2] and Mikhlin [3].

Unlike in the elastic case, except when $f=-1$ the stationarity of quotient (15) alone (for all variations satisfying the essential mechanical and electrical boundary conditions) is not equivalent to solving the free vibration problem of a piezoelectric continuum, but it is possible to formulate the stationarity principle by considering an extended quotient. Assuming that the displacements and electrostatic potential vary harmonically $[u_i(x,t)=U_i(x)\sin(\omega t+\alpha)$ and $\phi(x,t)=\Phi(x)\sin(\omega t+\alpha)]$ the extended quotient is defined in the following manner:

$$\omega^2 = \frac{\text{Num}}{\int_V \rho U_i U_i dV},$$

$$\begin{aligned}\text{Num} = &\int_V [c_{ijk\ell}U_{i,j}U_{k,\ell} + f\kappa_{ik}\Phi_{,i}\Phi_{,k} + (1-f)e_{ik\ell}\Phi_{,i}U_{k,\ell}]dV \\ &- (1+f)\int_V \Phi(e_{ik\ell}U_{k,\ell} - \kappa_{ik}\Phi_{,k})_{,i}dV \\ &+ (1+f)\int_{\partial V} \Phi(e_{ik\ell}U_{k,\ell} - \kappa_{ik}\Phi_{,k})n_i d\partial V. \end{aligned} \quad (16)$$

For conciseness the dependence of the amplitudes U_i and Φ on x is not written explicitly. The first term appearing in the numerator of Eq. (16) coincides with the numerator of quotient (15). The remaining two terms account for the constraints imposed by the Gauss equation of electrostatics $(6)_2$ and the natural electrical boundary condition $(7)_4$, respectively. The functions $-(1+f)\Phi(x)$ and $(1+f)\Phi(x)$ are the Lagrange multipliers, accounting for the constraints imposed, respectively, by Eq. $(6)_2$ and the boundary condition $(7)_4$ (additional details about the constrained variational principles of mechanics and the Lagrange multiplier method are provided by Reddy [18]). Using the extended quotient (16) one can formulate the following:

Stationarity principle:

Given the extended quotient (16):

(1) The variation of quotient (16) is equal to zero if the displacements and electrostatic potential satisfy the governing equations of the free vibration problem (6) and the prescribed boundary conditions of the problem (7).
(2) If the variation of quotient (16) is equal to zero for all displacements and electrostatic potential which satisfy the prescribed essential mechanical and electrical boundary conditions $(7)_{1,3}(u_i=0, \phi=0)$, the governing Eqs. (6) and the natural mechanical and electrical boundary conditions $(7)_{2,4}$ result from the stationarity condition.

An outline of the proof of this principle is given in the following steps:

Integrating by parts the term $-(1+f)\int_V \Phi(e_{ik\ell}U_{k,\ell} - \kappa_{ik}\Phi_{,k})_{,i}dV$ (using the Gauss theorem of calculus in the way discussed in Sec. 2.1), one transforms quotient (16) to the following equivalent form:

$$\omega^2 = \frac{\int_V (c_{ijk\ell}U_{i,j}U_{k,\ell} - \kappa_{ik}\Phi_{,i}\Phi_{,k} + 2e_{ik\ell}\Phi_{,i}U_{k,\ell})dV}{\int_V \rho U_i U_i dV}. \quad (17)$$

It is to be noted that quotient (17) is independent of the arbitrary constant f.

Using the symmetry conditions (5) the variation of quotient (17) is found after some calculations to be equal to:

$$\delta\omega^2 = 2\frac{\int_V (c_{ijk\ell}U_{i,j} + e_{ik\ell}\Phi_{,i})\,\delta U_{k,\ell}dV + \int_V (e_{ik\ell}U_{k,\ell} - \kappa_{ik}\Phi_{,k})\,\delta\Phi_{,i}dV - \omega^2\int_V \rho U_i\delta U_i dV}{\int_V \rho U_i U_i dV}. \tag{18}$$

By integrating the first and the second term of the numerator of Eq. (18) by parts and using the symmetry conditions (5) one can show part (1) of the stationarity principle. For part (2), variation (18) is equal to zero (ω^2 is stationary) if:

$$\int_V (c_{ijk\ell}U_{i,j} + e_{ik\ell}\Phi_{,i})\,\delta U_{k,\ell}dV + \int_V (e_{ik\ell}U_{k,\ell} - \kappa_{ik}\Phi_{,k})\,\delta\Phi_{,i}dV - \omega^2\int_V \rho U_i\delta U_i dV = 0. \tag{19}$$

Integrating Eq. (19) by parts and assuming that the admissible fields satisfy the prescribed essential boundary conditions part (2) of the stationarity principle is proved.

It should also be pointed out that one can arrive at stationarity condition (19) in a different way by making use of Hamilton's principle discussed in Ref. [9].

3 Application to the Approximate Calculation of the Natural Frequencies of a Piezoelectric Rod

In order to demonstrate the application of quotient (16) [or its equivalent simpler form (17)] to the approximate calculation of the natural frequencies and the character of their convergence, let us consider the case of the longitudinal vibrations along the z axis of a prismatic piezoelectric rod shown in Fig. 1. The case when the lateral surface of the rod is free from surface tractions and is not covered by electrodes is considered. Only the z component of the displacement is restrained at the left end so that the transverse deformation is not restrained. It is assumed that the rod is poled along the z axis and is made of material which is transversely isotropic in the plane perpendicular to the z axis (such material symmetry holds, e.g., if the rod is made of a piezoelectric ceramic poled along the z axis). In this case the physical Eqs. (2) can be written in the following matrix form:

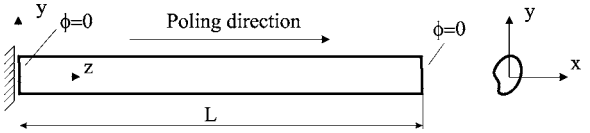


Fig. 1 The piezoelectric rod under consideration

$$\begin{Bmatrix}\sigma_x\\ \sigma_y\\ \sigma_z\\ \sigma_{yz}\\ \sigma_{xz}\\ \sigma_{xy}\end{Bmatrix} = \begin{bmatrix} c_{1111} & c_{1122} & c_{1133} & 0 & 0 & 0\\ c_{1122} & c_{1111} & c_{1133} & 0 & 0 & 0\\ c_{1133} & c_{1133} & c_{3333} & 0 & 0 & 0\\ 0 & 0 & 0 & c_{2323} & 0 & 0\\ 0 & 0 & 0 & 0 & c_{2323} & 0\\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}(c_{1111}-c_{1122}) \end{bmatrix} \times \begin{Bmatrix}\varepsilon_x\\ \varepsilon_y\\ \varepsilon_z\\ 2\varepsilon_{yz}\\ 2\varepsilon_{xz}\\ 2\varepsilon_{xy}\end{Bmatrix} - \begin{bmatrix} 0 & 0 & e_{311}\\ 0 & 0 & e_{311}\\ 0 & 0 & e_{333}\\ 0 & e_{113} & 0\\ e_{113} & 0 & 0\\ 0 & 0 & 0 \end{bmatrix}\begin{Bmatrix}E_x\\ E_y\\ E_z\end{Bmatrix},$$

$$\begin{Bmatrix}D_x\\ D_y\\ D_z\end{Bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & e_{113} & 0\\ 0 & 0 & 0 & e_{113} & 0 & 0\\ e_{311} & e_{311} & e_{333} & 0 & 0 & 0 \end{bmatrix}\begin{Bmatrix}\varepsilon_x\\ \varepsilon_y\\ \varepsilon_z\\ 2\varepsilon_{yz}\\ 2\varepsilon_{xz}\\ 2\varepsilon_{xy}\end{Bmatrix} + \begin{bmatrix} \kappa_{11} & 0 & 0\\ 0 & \kappa_{11} & 0\\ 0 & 0 & \kappa_{33} \end{bmatrix}\begin{Bmatrix}E_x\\ E_y\\ E_z\end{Bmatrix}. \tag{20}$$

For a more detailed discussion of different ways of writing the constitutive equations of linear piezoelectricity (including the so-called compressed notation) the reader is referred to Refs. [7,9,12].

For such a piezoelectric rod and the assumed material symmetry one can obtain the approximate one-dimensional equations of longitudinal vibrations from the general three-dimensional formulation, by the direct extension of the corresponding derivation for an isotropic elastic rod discussed, e.g., by Landau and Lifshitz [19]. Since the rod lateral surface is free from surface tractions and is not covered by electrodes, it follows from the boundary conditions $(7)_{2,4}$ that $\sigma_{xx}=\sigma_{yy}=\sigma_{yz}=\sigma_{xz}=\sigma_{xy}=0$ and $D_x=D_y=0$ on the lateral surface, so that only σ_z and D_z are nonzero there. For a slender rod a uniaxial stress state $\sigma_z \neq 0$ can be used inside the rod, D_z can be taken for the only nonzero component of the electric displacement vector and assuming that the material properties are constant or are functions of z alone, σ_z and D_z are functions of z only. As a result, the problem of the free vibration of a ceramic piezoelectric rod with unelectroded lateral surface is described by the following governing equations and boundary conditions:

$$\begin{cases} \dfrac{\partial}{\partial z}\left[c_K\dfrac{\partial u}{\partial z} + e_K\dfrac{\partial \phi}{\partial z}\right] = \rho\dfrac{\partial^2 u}{\partial t^2} \\ \dfrac{\partial}{\partial z}\left[e_K\dfrac{\partial u}{\partial z} - \kappa_K\dfrac{\partial \phi}{\partial z}\right] = 0 \end{cases} \quad \text{for } z \in (0,L). \tag{21}$$

$$\begin{cases} u=0 \text{ or } c_K\dfrac{\partial u}{\partial z}+e_K\dfrac{\partial \phi}{\partial z}=0 \\ \phi=0 \text{ or } e_K\dfrac{\partial u}{\partial z}-\kappa_K\dfrac{\partial \phi}{\partial z}=0 \end{cases} \quad \text{for } z=0,L, \tag{22}$$

where $u(z,t)$ is the displacement along the z axis. The coefficients appearing in Eqs. (21) and (22) can be functions of z and are defined using the material coefficients in the constitutive Eqs. (20) in the following way:

$$c_K=c_{3333}-\frac{2(c_{1133})^2}{c_{1111}+c_{1122}}, \quad e_K=e_{333}-\frac{2c_{1133}e_{311}}{c_{1111}+c_{1122}},$$

$$\kappa_K=\kappa_{33}+\frac{2(e_{311})^2}{c_{1111}+c_{1122}}. \tag{23}$$

Equation (21) and (22) have also been obtained by Le [20] from a general theory of the vibrations of curved piezoelectric rods made of a material belonging to any symmetry class. The set of approximate governing equations and the boundary conditions for curved rods have been derived using the variational-asymptotic method. For a special case of a straight piezoelectric rod with the lateral surface not covered by electrodes and made of a material belonging to any symmetry class, Eqs. (21) are given in [20] as a special form of the general theory. It is pointed out that the constitutive Eqs. (23) are only valid for the case of transverse isotropy, and for this case they can be shown to be equivalent to those, obtained from the energy functional expressed in terms of the stress and electric displacement vector components (the so-called elastic energy) discussed in Ref. [20].

In the analyzed case of longitudinal vibrations of a piezoelectric rod quotient (16) assumes the form:

$$\omega^2=\frac{\text{Num}}{\displaystyle\int_0^L \rho[U(z)]^2dz},$$

$$\begin{aligned}\text{Num}=&\int_0^L \{c_K[U'(z)]^2+f\kappa_K[\Phi'(z)]^2+(1-f)e_KU'(z)\Phi'(z)\}dz\\ &-(1+f)\int_0^L \Phi(z)[e_KU'(z)-\kappa_K\Phi'(z)]'dz\\ &+(1+f)[\Phi(z)(e_KU'(z)-\kappa_K\Phi'(z))]|_0^L. \end{aligned}\tag{24}$$

The simpler, equivalent expression (17) in the present case is given by the following formula:

$$\omega^2=\frac{\displaystyle\int_0^L \{c_K[U'(z)]^2-\kappa_K[\Phi'(z)]^2+2e_KU'(z)\Phi'(z)\}dz}{\displaystyle\int_0^L \rho[U(z)]^2dz} \tag{25}$$

and the stationarity condition (19) is given as follows:

$$\begin{aligned}&-\omega^2\int_0^L \rho U(z)\delta U(z)dz+\int_0^L [c_KU'(z)+e_K\Phi'(z)]\delta U'(z)dz\\ &+\int_0^L [e_KU'(z)-\kappa_K\Phi'(z)]\delta\Phi'(z)dz=0. \end{aligned}\tag{26}$$

One can use stationarity condition (26) to calculate the natural frequencies by the Rayleigh–Ritz method. As an example, a rod which is mechanically fixed at the left end and free to move at the right end is considered ($u=0$ for $z=0$, $c_K\partial u/\partial z+e_K\partial\phi/\partial z=0$ for $z=L$). A short-circuit electrical boundary condition is used ($\phi=0$ at $z=0,L$). The analysis will be done in non-dimensional form and the following dimensionless quantities are defined (constant material properties will be assumed in all numerical examples):

$$\bar z=\frac{z}{L}, \quad \bar U_n(\bar z)=\frac{U_n(z)}{L}, \quad \bar\Phi_n(\bar z)=\frac{\Phi_n(z)e_K}{Lc_K}, \quad \bar\omega=\omega L\sqrt{\frac{\rho}{c_K}},$$

$$\bar\kappa_K=\frac{\kappa_Kc_K}{(e_K)^2}. \tag{27}$$

The displacement and electrostatic potential and their variations will be approximated using the trigonometric functions which satisfy the prescribed essential boundary conditions of the problem:

$$\bar U_n(\bar z)=\sum_{i=1}^{I} a_i\sin\left(\frac{i\pi\bar z}{2}\right), \quad \delta\bar U_n(\bar z)=\sum_{i'=1}^{I}\delta a_{i'}\sin\left(\frac{i'\pi\bar z}{2}\right),$$

$$\bar\Phi_n(\bar z)=\sum_{j=1}^{J} b_j\sin(j\pi\bar z), \quad \delta\bar\Phi_n(\bar z)=\sum_{j'=1}^{J}\delta b_{j'}\sin(j'\pi\bar z). \tag{28}$$

Introducing expressions (28) into the dimensionless form of the stationarity condition (26) the following algebraic eigenvalue problem results:

$$-\bar\omega^2\begin{bmatrix}\mathbf{M}^{(11)} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{Bmatrix}\mathbf{a}\\ \mathbf{b}\end{Bmatrix}+\begin{bmatrix}\mathbf{K}^{(11)} & \mathbf{K}^{(12)}\\ (\mathbf{K}^{(12)})^{\mathrm{T}} & \mathbf{K}^{(22)}\end{bmatrix}\begin{Bmatrix}\mathbf{a}\\ \mathbf{b}\end{Bmatrix}=\mathbf{0}, \tag{29}$$

where: $\mathbf{a}=(a_1,a_2,\ldots,a_I)^{\mathrm{T}}$ and $\mathbf{b}=(b_1,b_2,\ldots,b_J)^{\mathrm{T}}$. The elements of the block matrices appearing in Eq. (29) are calculated as follows:

$$\begin{aligned}M^{(11)}_{i',i}&=\int_0^1\sin\left(\frac{i'\pi\bar z}{2}\right)\sin\left(\frac{i\pi\bar z}{2}\right)d\bar z,\\ K^{(11)}_{i',i}&=\frac{\pi^2i'i}{4}\int_0^1\cos\left(\frac{i'\pi\bar z}{2}\right)\cos\left(\frac{i\pi\bar z}{2}\right)d\bar z,\\ K^{(22)}_{j',j}&=-\pi^2\bar\kappa_Kj'j\int_0^1\cos(j'\pi\bar z)\cos(j\pi\bar z)d\bar z,\\ K^{(12)}_{i',j}&=\frac{\pi^2i'j}{2}\int_0^1\cos\left(\frac{i'\pi\bar z}{2}\right)\cos(j\pi\bar z)d\bar z,\\ i,i'&=1\ldots I, \quad j,j'=1\ldots J. \end{aligned}\tag{30}$$

By eliminating the vector $\mathbf{b}$, eigenvalue problem (29) can be transformed to the form (which has been used before in the finite element analysis by Allik and Hughes [10]):

$$-\bar\omega^2\mathbf{M}^{(11)}\mathbf{a}+[\mathbf{K}^{(11)}-\mathbf{K}^{(12)}(\mathbf{K}^{(22)})^{-1}(\mathbf{K}^{(12)})^{\mathrm{T}}]\mathbf{a}=\mathbf{0}. \tag{31}$$

To obtain the numerical results the following values of material constants will be used, which are typical of a piezoelectric ceramic PZT4:

$$\begin{aligned}c_{1111}&=13.2\times10^{10}\left(\frac{\mathrm{N}}{\mathrm{m}^2}\right), \quad c_{1122}=7.1\times10^{10}\left(\frac{\mathrm{N}}{\mathrm{m}^2}\right),\\ c_{1133}&=7.3\times10^{10}\left(\frac{\mathrm{N}}{\mathrm{m}^2}\right), \quad c_{3333}=11.5\times10^{10}\left(\frac{\mathrm{N}}{\mathrm{m}^2}\right),\\ e_{311}&=-4.1\left(\frac{\mathrm{C}}{\mathrm{m}^2}\right), \quad e_{333}=14.1\left(\frac{\mathrm{C}}{\mathrm{m}^2}\right),\\ \kappa_{33}&=5.841\times10^{-9}\left(\frac{\mathrm{F}}{\mathrm{m}}\right), \quad \rho=7.5\times10^{3}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right). \end{aligned}\tag{32}$$

The convergence of the lowest three natural frequencies is shown in Table 1. An equal number of terms ($I=J$) have been used in Eq.

Table 1 Convergence of the lowest three non-dimensional circular frequencies of a fixed-free rod with short-circuited ends obtained using approximation given by Eq. (28)

$I=J$	1	2	3	4	6	8	10	12	14
$\bar{\omega}_1$	1.677	1.639	1.638	1.638	1.638	1.638	1.638	1.638	1.638
$\bar{\omega}_2$	⋯	6.892	6.141	6.146	6.150	6.150	6.151	6.151	6.151
$\bar{\omega}_3$	⋯	⋯	13.988	10.352	10.376	10.382	10.384	10.386	10.386

(28) to approximate the displacement and electrostatic potential. The dots appearing at some locations in the table mean that the approximate value of the respective circular frequency could not be calculated with the number of approximating functions used. The underlined values are the first occurrences when the calculated frequencies do not change to the number of digits shown, for two consecutive columns. It is pointed out that unlike in the elastic case, the convergence is not necessarily monotonic as can be seen from the table. Unlike in the elastic case, the numerator of quotient (16) or its simplified equivalent form (17) is not positive definite, and thus the property which is crucial in the proof of the monotonic convergence in the elastic case is not valid in piezoelectricity. The difference between the elastic and the piezoelectric case can also be seen by regarding Eq. (31). Even though the algebraic eigenvalue problem described by Eq. (31) has the same form as in the elastic case it uses the reduced stiffness matrix.

Using the extension of the Rayleigh quotient discussed in the present paper it is possible to prove monotonic convergence when the approximating functions apart from the prescribed essential boundary conditions satisfy the equations of constraints [the electrostatic equation and the natural electrical boundary conditions (if prescribed)]. For this, consider the numerator of quotient (24) in which the second term disappears when the approximating functions satisfy the electrostatic Eq. $(21)_2$ and the third term is zero when the approximating functions fulfill either the essential or the natural electrical boundary condition (whichever is prescribed). Since the stationarity condition (26) is independent of the value of arbitrary constant f used in Eq. (24), it is convenient to take $f=1$. As a result, when the constraint conditions are satisfied, one obtains the following expression for the quotient:

$$\omega^2 = \frac{\int_0^L \{c_K[U'(z)]^2 + \kappa_K[\Phi'(z)]^2\}dz}{\int_0^L \rho[U(z)]^2 dz}. \tag{33}$$

It is possible to show in an explicit way that when the admissible functions fulfill the constraints, expression (33) is indeed valid irrespective of the value of f used in quotient (24). In fact, by integrating the electrostatic Eq. $(21)_2$ one obtains:

$$e_K U'(z) - \kappa_K \Phi'(z) = A, \tag{34}$$

where A is an arbitrary constant of integration. If the electrical natural boundary condition is prescribed at least at one end of the rod, then A is equal to zero. In the case of short-circuited ends ($\Phi=0$ at both ends) studied in the present example, A is nonzero. Introducing Eq. (34) in Eq. (25) one obtains:

$$\omega^2 = \frac{\int_0^L \{c_K[U'(z)]^2 + \kappa_K[\Phi'(z)]^2 + 2A\Phi'(z)\}dz}{\int_0^L \rho[U(z)]^2 dz}, \tag{35}$$

which upon integrating the last term in the numerator and using the boundary condition results in Eq. (33). The numerator of expression (33) (which is equal to the internal energy of the piezoelectric rod) and the denominator are positive definite for the functions satisfying the constraints and thus monotonic convergence results, by the same argument as in the elastic case discussed in Refs. [2,3].

In order to illustrate the above discussion, the displacement and potential are approximated as follows:

$$\bar{U}(\bar{z}) = \sum_{i=1}^{I} a_i \sin(i\pi\bar{z}) + a_{I+1}\bar{z}, \quad \delta\bar{U}(\bar{z}) = \sum_{i'=1}^{I} \delta a_{i'} \sin(i'\pi\bar{z}) + \delta a_{I+1}\bar{z}$$
$$\bar{\Phi}(\bar{z}) = \frac{1}{\bar{\kappa}_K}\sum_{i=1}^{I} a_i \sin(i\pi\bar{z}), \quad \delta\bar{\Phi}(\bar{z}) = \frac{1}{\bar{\kappa}_K}\sum_{i'=1}^{I} \delta a_{i'} \sin(i'\pi\bar{z}). \tag{36}$$

These functions satisfy the prescribed essential boundary conditions of zero displacement at the left end and zero potential at both ends. Moreover, approximations (33) render the dimensionless form of the electrostatic Eq. $(21)_2$ identity. No natural electrical boundary condition (boundary constraint) is prescribed in this example. It is to be noted that the term $a_{K+1}\bar{z}$ has been added in the approximation of displacement so that the approximating functions form a complete set of functions for a fixed-free rod. Upon introducing series (36) into the dimensionless form of stationarity condition (26) the following algebraic eigenvalue problem results:

$$-\bar{\omega}^2\mathbf{Ma} + \mathbf{Ka} = \mathbf{0}, \tag{37}$$

with the components of the mass and stiffness matrices defined as follows:

$$M_{i',i} = \int_0^1 \sin(i'\pi\bar{z})\sin(i\pi\bar{z})d\bar{z},$$
$$M_{i',I+1} = \int_0^1 \bar{z}\sin(i'\pi\bar{z})d\bar{z}, \quad M_{I+1,i} = \int_0^1 \bar{z}\sin(i\pi\bar{z})d\bar{z},$$
$$M_{I+1,I+1} = \int_0^1 \bar{z}^2 d\bar{z},$$
$$K_{i',i} = \left(1 + \frac{1}{\bar{\kappa}_K}\right)\pi^2 i' i \int_0^1 \cos(i'\pi\bar{z})\cos(i\pi\bar{z})d\bar{z}, \tag{38}$$
$$K_{i',I+1} = \left(1 + \frac{1}{\bar{\kappa}_K}\right) i'\pi \int_0^1 \cos(i'\pi\bar{z})d\bar{z} = 0,$$
$$K_{I+1,i} = \left(1 + \frac{1}{\bar{\kappa}_K}\right) i\pi \int_0^1 \cos(i\pi\bar{z})d\bar{z} = 0,$$
$$K_{I+1,I+1} = 1,$$
$$i,i' = 1 \ldots I$$

The lowest three natural frequencies using approximations (36) are shown in Table 2, from which one can observe that the convergence is monotonic in this case.

Table 3 includes a summary of results in dimensional form obtained by four different approaches: using the characteristic

Table 2 Convergence of the lowest three non-dimensional circular frequencies of a fixed-free rod with short-circuited ends obtained using approximation given by Eq. (36)

I	1	2	3	4	6	8	10	12	14
$\bar{\omega}_1$	1.643	1.639	1.639	1.638	1.638	1.368	1.638	1.638	1.638
$\bar{\omega}_2$	7.044	6.221	6.175	6.162	6.155	6.153	6.152	6.152	6.152
$\bar{\omega}_3$	⋯	12.138	10.535	10.445	10.404	10.395	10.391	10.390	10.389

equation, obtained by the Rayleigh-Ritz method with two different approximations (28) or (36) and calculated by the finite element method. By solving the eigenvalue problem in a similar way as for the free vibration of an elastic rod, one obtains for the natural circular frequencies of the longitudinal vibrations of a piezoelectric rod: $\omega_n = k_n\sqrt{(c_K + e_K^2/\kappa_K)/\rho}$, where k_n are the roots of the characteristic equation, which in the present case of a fixed-free rod with electrically short-circuited ends takes the form:

$$\left(c_K + \frac{e_K^2}{\kappa_K}\right)\frac{k_n L}{tg(k_n L)} - \frac{e_K^2}{\kappa_K} = 0. \tag{39}$$

The natural frequencies in (Hz) ($f_n = \omega_n/2\pi$) are shown in Table 3. The finite element results were calculated using the coupled-field capability of the finite element code Ansys®. The results were obtained by three-dimensional analysis using brick elements. The rod 0.1 (m) long had a square cross section with sides equal to 0.005 (m) and 3×3 elements were used over the cross section and 40 elements longitudinally. In the finite element model only the z component of the displacement was set to zero at the rod fixed left end so that the transverse deformation was not restrained. The three-dimensional analysis used the full form of the constitutive Eqs. (20) and the material constants not specified in Eq. (32) were those used in the author's earlier paper [17].

Even though, as has been pointed out above, monotonic convergence of frequencies cannot be guaranteed unless the equations of constraints are satisfied, it can take place in some cases. This is illustrated in Table 4 where the convergence of non-dimensional natural frequencies of a fixed-free rod is shown for open-circuit electrical boundary conditions ($D_z = 0$ at $z = 0, L$). The results shown were obtained using the following approximations of the longitudinal displacement and electrostatic potential:

$$\bar{U}_n(\bar{z}) = \sum_{i=1}^{I} a_i \sin\left(\frac{i\pi\bar{z}}{2}\right), \quad \bar{\Phi}_n(\bar{z}) = \sum_{j=1}^{J} b_j \sin\left(\frac{j\pi\bar{z}}{2}\right). \tag{40}$$

It is to be remembered that the electrostatic potential is only determined up to an arbitrary constant [the electric field vector is calculated from Eq. (4) using the derivatives of the electrostatic potential]. As a result, in the case of open-circuit electrical boundary conditions and unelectroded lateral surface of the rod (electrically "unrestrained" case), an arbitrary constant can be added to approximation $(40)_2$ without altering the calculated natural frequencies.

4 Conclusions

The paper has discussed an extension of the Rayleigh quotient to the linear theory of piezoelectricity. An infinite number of equivalent expressions of the quotient have been obtained and the stationarity principle of any of these quotients under the additional constraints imposed by the Gauss equation of electrostatics and the prescribed natural electrical boundary condition has been discussed. The stationarity condition of any of the quotients under the additional constraints coincides (in the case of free vibrations) with that resulting from Hamilton's principle of piezoelectricity discussed in Ref. [9]. An important difference in the definition of the Rayleigh quotient compared to the elastic case exists, since for a piezoelectric continuum the stationarity principle has been formulated as a constrained variational problem. As a consequence no monotonic convergence of the natural frequencies is guaranteed using the Rayleigh–Ritz method, unless the approximating functions satisfy the equations of constraints imposed by the Gauss equation of electrostatics and the prescribed electrical natural boundary condition. The convergence properties of the Rayleigh–Ritz method have been demonstrated using an example of the free longitudinal vibrations of a piezoelectric rod. Two different sets of admissible functions have been used, one of which has satisfied the equation of constraint, and the different character of convergence using different base functions has been brought up. For both sets used the final values have been shown to agree very well with the frequencies calculated from the characteristic equation and using three-dimensional finite element analysis.

Even though the calculation of the approximate frequencies of a piezoelectric continuum can be based on Hamilton's principle discussed by Tiersten [9], the advantage of studying the Rayleigh

Table 3 Comparison of the lowest three natural frequencies of longitudinal vibrations of a piezoelectric fixed-free rod of length $L = 0.1$ (m), with short-circuited ends.

Lowest natural frequencies	f_1	f_2	f_3
Solution using the characteristic equation	7526 (Hz)	28262 (Hz)	47722 (Hz)
Solution by the Rayleigh–Ritz method [$I = J = 10$ in Eq. (28)]	7526 (Hz)	28259 (Hz)	47709 (Hz)
Rayleigh–Ritz solution [$I = 10$ in Eq. (36)]	7526 (Hz)	28266 (Hz)	47740 (Hz)
Finite element solution (Ansys®)	7526 (Hz)	28268 (Hz)	47754 (Hz)

Table 4 Convergence of the lowest three non-dimensional circular frequencies of a fixed-free rod with open-circuit ends obtained using approximation given by Eq. (40)

$I=J$	1	2	3	4	5	6	8	10
$\bar{\omega}_1$	2.092	2.092	2.092	2.092	2.092	2.092	2.092	2.092
$\bar{\omega}_2$	⋯	7.167	6.277	6.277	6.277	6.277	6.277	6.277
$\bar{\omega}_3$	⋯	⋯	14.343	10.529	10.462	10.462	10.462	10.462

quotient follows from the fact that it provides additional insight into the character of convergence of the approximate frequencies calculated using the variational method. It has been shown in the paper that some properties of the elastic case do not carry over to piezoelectricity and the study of the properties of the Rayleigh quotient can add to the understanding of the free vibration problems of piezoelectric continua.

Nomenclature

D_i = components of the electric displacement vector
E_i = components of the electric field vector
$\mathbf{K}$ = stiffness matrix
L = length of the rod
$\mathbf{M}$ = mass matrix
T = matrix transpose
$U_i^{(n)}(x)$ = displacements of nth vibration mode
$U_i(x)$ = amplitude of vibrations
V = volume of the piezoelectric body
∂V = boundary of the piezoelectric body
$c_{ijk\ell}$ = components of the elastic stiffness tensor
c_K, e_K, κ_K = constants in equations of longitudinal vibrations of a piezoelectric rod
e_{ijk} = components of the piezoelectric tensor
f = arbitrary tracing constant
f_n = nth natural frequency in (Hz)
k_n = nth root of the characteristic equation
n = index numbering the natural frequencies and eigenmodes
t = time
$u(z,t)$ = displacement of the rod along z axis
$u_i(x,t)$ = components of the displacement vector
$\ddot{u}_i(x,t)$ = components of the acceleration vector
$x=(x_1,x_2,x_3)$ = Cartesian coordinates of the points of the piezoelectric body
$\Phi^{(n)}(x)$ = electrostatic potential for nth vibration mode
$\Phi(x)$ = amplitude of electrostatic potential
ε_{ij} = components of the strain tensor
$\phi(x,t)$ = electrostatic potential
κ_{ij} = components of the dielectric tensor
α = arbitrary phase shift
δ = symbol of the variation
ω_n = nth natural circular frequency
ω^2 = Rayleigh's quotient taking into account the constraints
ρ = mass density
σ_{ij} = components of the stress tensor
$\emptyset$ = empty set
$\cap$ = intersection of sets
$\cup$ = union of sets
$\bar{(\cdot)}$ = non-dimensional quantities
$(\cdot)_{,i}$ = partial derivative $\partial(\cdot)/\partial x_i$
$(\cdot)' = d/dz$ = derivative with respect to z (except when used with indices i', j')

References

[1] Lord, Rayleigh, 1945 (1st ed. 1877), *Theory of Sound*, Dover, New York.
[2] Courant, R., and Hilbert, D., 1937, *Methoden der Mathematischen Physik*, Springer, Berlin. (English translation, *Methods of Mathematical Physics*, 1952 (v.1), 1962 (v.2), Wiley, New York.)
[3] Mikhlin, S. G., 1970, *Variational Methods in Mathematical Physics (in Russian)*, 2nd ed., Nauka, Moscow.
[4] Strang, G., and Fix, G. J., 1973, *An Analysis of the Finite Element Method*, Prentice–Hall, New York.
[5] Washizu, K., 1982, *Variational Methods in Elasticity and Plasticity*, Pergamon, Oxford New York.
[6] Cady, W. G., 1946, *Piezoelectricity*, McGraw–Hill, New York.
[7] Mason, W. P., 1950, *Piezoelectric Crystals and their Application to Ultrasonics*, Van Nostrand, Toronto.
[8] Lawson, A. W., 1942, "The Vibration of Piezoelectric Plates," Phys. Rev., **62**, pp. 71–76.
[9] Tiersten, H. F., 1969, *Linear Piezoelectric Plate Vibrations*, Plenum, New York.
[10] Allik, H., and Hughes, T., 1970, "Finite Element Method for Piezoelectric Vibration," Int. J. Numer. Methods Eng., **2**, pp. 151–157.
[11] He, J.-H., 2001, "Coupled Variational Principles of Piezoelectricity," Int. J. Eng. Sci., **39**, pp. 323–341.
[12] Nowacki, W., 1983, *Electromagnetic Effects in Deformable Solids (in Polish)*, PWN, Warsaw.
[13] He, J.-H., 2001, "Hamilton Principle and Generalized Variational Principle of Linear Thermopiezoelectricity," ASME J. Appl. Mech., **68**, pp. 666–667.
[14] Yang, J. S., and Batra, R. C., 1995, "Conservation Laws in Linear Piezoelectricity," Eng. Fract. Mech., **51**, pp. 1041–1047.
[15] Liu, G. R., and Xi, Z. C., 2002, *Elastic Waves in Anisotropic Laminates*, CRC, Boca Raton, FL.
[16] Han, X., and Liu, G. R., 2003, "Elastic Waves in a Functionally Graded Piezoelectric Cylinder," Smart Mater. Struct., **12**, pp. 962–971.
[17] Cupiał, P., 2005, "Three Dimensional Natural Vibration Analysis and Energy Considerations for a Piezoelectric Rectangular Plate," J. Sound Vib., **283**, pp. 1093–1113.
[18] Reddy, J. N., 1986, *Applied Functional Analysis and Variational Methods in Engineering*, McGraw-Hill, New York.
[19] Landau, L. D., and Lifshitz, E. M., 1993, *Theory of Elasticity (in Polish)*, PWN, Warsaw.
[20] Le, K. C., 1999, *Vibrations of Shells and Rods*, Springer, Berlin.

Correlation Moment Analysis and the Time Dependence of Coherence in Systems Described by Nonlinear Partial Differential Equations

M. R. Belmont
School of Engineering and Computer Science,
University of Exeter,
Exeter EX4 4QF, UK
e-mail: m.r.belmont@exeter.ac.uk

The work presented introduces correlation moment analysis. This technique can be employed to explore the growth of determinism from stochastic initial conditions in physical systems described by non-linear partial differential equations (PDEs) and is also applicable to wholly deterministic situations. Correlation moment analysis allows the analytic determination of the time dependence of the spatial moments of the solutions of certain types of non-linear partial differential equations. These moments provide measures of the growth of processes defined by the PDE, furthermore the results are obtained without requiring explicit solution of the PDE. The development is presented via case studies of the linear diffusion equation and the non-linear Kortweg de-Vries equation which indicate strategies for exploiting the various properties of correlation moments developed in the text. In addition, a variety of results have been developed which show how various classes of terms in PDEs affect the structure of a sequence of correlation moment equations. This allows results to be obtained about the behavior of the PDE solution, in particular how the presence of certain types of terms affects integral measures of the solution. It is also demonstrated that correlation moments provide a very simple, natural approach to determining certain subsets of conserved quantities associated with the PDEs. [DOI: 10.1115/1.2065687]

1 Introduction

A large class of physical systems have the property of producing outputs that are less random than their inputs. Putting this rather loose statement more formally: if the input to such a system is selected randomly from an appropriate ensemble, then the autocorrelation function of the output is always wider than that of the input. This process occurs in both linear and non-linear systems, however for comparable bandwidths it can be much more pronounced in the non-linear case.

One non-linear example of this, drawn from geology, concerns the deformation of layered rocks where an initial localized random folding rapidly transforms into a highly deterministic propagating structure [1]. This entropy reduction process can lead to surprising consequences in unexpected areas. Such a case arises in the simulation of very short term deterministic prediction of ocean swell waves [2]. Here linearized local approximations are made to long period weakly non-linear gravity waves (swell waves) using data gathered of the order of 1 km up-wave of the prediction site [3,4]. The precision of such predictions can be shown to depend strongly upon the degree of frequency domain statistical dependence present within the magnitude and phase spectra of the data.

The usual tools for analyzing such effects are auto and cross-correlation functions. For linear examples these measures can be readily determined using Wiener's developments of linear system theory for stochastic signals. However, hardly surprisingly, there are no such readily applicable general tools for non-linear problems. The direct approach to obtaining quantitative results is via detailed solutions (usually numerical) of the governing equations. An alternative is to treat the systems as stochastic partial differential equations (PDEs) and employ the formulations derived from the Îto integral [5–7], or from stochastic Greens function techniques [8,9]. Both of these two approaches prove to be very demanding in detailed applications. Furthermore, most systems of interest are characterized by deterministic PDEs with stochastic inputs which can be described by band limited (analytic) functions drawn from some parametrized ensemble.

As a consequence, it was felt worthwhile to develop a technique that avoided obtaining explicit solutions to the PDEs of interest. To illustrate the methodology, it is initially applied to the classical linear partial differential equations which describes diffusion. The technique is then used to examine the non-linear Kortweg de-Vries (KdV) equation often used as a model system for solitary waves.

2 Methodology and its Implementation

As stated in the introduction, the initial motivation behind the work presented here was to address the problem of characterizing the evolution of determinism, from stochastic initial conditions, within systems described by partial differential equations. However, the approach employed means that the technique applies equally to the exploration of purely deterministic issues.

The methodology employs an integral operator technique to generate sets of ordinary differential equations (ODEs) from a defining PDE. The dependent variables in these ODEs are correlation moments, i.e., length scale measures of various orders of cross-correlation functions of the solution of the original PDE. Such measures have analogs in the theory of statistical moments. For non-linear PDEs the system of ODEs is coupled together. For

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received September 15, 2004; final manuscript received May 31, 2005. Review conducted by N. Sri Namachchivaya. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

the linear case they are uncoupled. In the examples considered the initial condition becomes the initial form of time dependent spatial correlation functions of various orders.

The presentation is not rigidly formal with the style being that of the physical sciences rather than a mathematical text. In the main the proofs are straightforward and thus have been omitted for clarity of development and for purposes of brevity. The properties of the correlation moments are explored first and then the results obtained are applied to the linear diffusion equations and the non-linear KdV equation. These two explorations demonstrate how the various properties of correlation moments point to a general approach to the application of correlation moment analysis.

A key step in using the correlation moment technique is the selection of weight functions which when taken in conjunction with the various conditions prevailing allow the system of ODEs to be solved for the properties of interest. For linear PDEs that lead to uncoupled ODEs this process is automatic and involves only one set of weight functions. For the case of non-linear PDEs that lead to coupled ODEs, typically a modest number of sets of weights are required. For non-linear PDEs whose terms involve products of integer powers of the dependent variable and their derivatives (as with the KdV equation) the selection of the sets of weights is again straightforward and is essentially prescribed by the correlation moment properties of the partial time derivative term as discussed in Secs. 4.1 and 4.2. This type of PDE is very common in physical applications.

2.1 Correlation Length Scales. The underlying principle behind the present approach is that the width of the spatial cross-correlation function, $R_{u_1,u_2}(\lambda,t)$, between two real-valued functions, $u_1(\zeta,t)$ and $u_2(\zeta,t)$, measures the length scale, $\lambda_w(t)$, [defined in Eq. (8) below] over which u_1 and u_2 are statistically dependent in space. Hence $\lambda_w(t)$ determines the distance over which u_1 and u_2 have a significant degree of interrelationship and thus can be used in both stochastic and wholly deterministic situations.

An unlimited number of different length scale measures of cross-correlation functions can be defined. These moments are denoted by, $\Gamma_{u_1,u_2,r}(t)$ for the cases, $r=0,1,2,3\ldots$ However it will be shown that the measure associated with $r=2$ appears to be the natural choice. These correlation moments are analogous to those associated with the moments of a statistical distribution as is evident from the definition of the rth correlation moment that is proposed here:

$$\Gamma_{u_1,u_2,r}(t)=\int_{-\infty}^{\infty}\lambda^r R_{u_1,u_2}(\lambda,t)d\lambda. \tag{1}$$

2.1.1 Conditions for the Existence of $\Gamma_{u_1,u_2,r}(t)$ and the Class of $R_{u_1,u_2}(\lambda,t)$ Functions. As will be demonstrated in Sec. 3, practical applications of $\Gamma_{u_1,u_2,r}(t)$ typically require the existence of $\Gamma_{u_1,u_2,r}(t)$ only up to modest values of r. However, it is considered useful to examine the boundedness of Eq. (1) for all r.

The Wiener Khintchine theorem states that, $R_{u_1,u_2}(\lambda,t)$, and the cross-spectral density function, $S_{u_1,u_2}(\omega,t)$, are a Fourier transform pair, thus formal r-fold differentiation of $S_{u_1,u_2}(\omega,t)$ with respect to, ω gives:

$$\frac{\partial^r S_{u_1,u_2}(\omega,t)}{\partial\omega^r}=(-j)^r\int_{-\infty}^{\infty}\lambda^r R_{u_1,u_2}(\lambda,t)e^{-j\omega\lambda}d\lambda. \tag{2}$$

Hence, if $\partial^r S_{u_1,u_2}(\omega,t)/\partial\omega^r$ exists at $\omega=0$,

$$\Gamma_{u_1,u_2,r}(t)=j^r\left.\frac{\partial^r S_{u_1,u_2}(\omega,t)}{\partial\omega^r}\right|_{\omega=0}. \tag{3}$$

Hence the conditions for $\Gamma_{u_1,u_2,r}(t)$ to be defined corresponds to the existence of $\partial^r S_{u_1,u_2}(\omega,t)/\partial\omega^r$ at $\omega=0$.

Examination of Eq. (3) shows that the correlation moments are the coefficients of the Taylor series expansion about $\omega=0$ of the cross-power spectral density, $S_{u_1,u_2}(\omega,t)$. This makes it possible to write an explicit expression for $S_{u_1,u_2}(\omega,t)$, in terms of the correlation moments, i.e.:

$$S_{u_1,u_2}(\omega,t)=\sum_{r=0}^{\infty}\frac{(-j\omega)^r}{r!}\Gamma_{u_1,u_2,r}(t). \tag{4}$$

As well as describing the relationship between $\Gamma_{u_1,u_2,r}(t)$ and the cross spectrum, Eq. (4) also provides a link to the formalism of the characteristic function which plays a central role in determining the statistical moments of u_1 and u_2.

Alternatively the boundedness of $\Gamma_{u_1,u_2,r}(t)$ can be examined directly in terms of the definition in Eq. (1). A sufficient, although rather excessively demanding, condition for $\Gamma_{u_1,u_2,r}(t)$ to exist as a function of t is that $\lambda^r R_{u_1,u_2}(\lambda,t)\in L^1(\Re)$ for all t. This is satisfied if $R_{u_1,u_2}(\lambda,t)$ is both bounded and exponentially small for large values of $|\lambda|$. In particular, for the highest order moment, r, of interest it will be required that all orders, p, of derivative present satisfy:

$$\lim_{|\lambda|\to\infty}\left\{\lambda^r\frac{\partial^p R_{u_1,u_2}(\lambda,t)}{\partial\lambda^p}\right\}=0. \tag{5}$$

For such correlation functions it is both natural and convenient to require that u_1 and u_2 are $\in L^2(\Re)$, so that $R_{u_1,u_2}(\lambda,t)$ has the form [10]:

$$R_{u_1,u_2}(\lambda,t)=\int_{-\infty}^{\infty}u_1(\zeta+\lambda,t)u_2(\zeta,t)d\zeta. \tag{6}$$

In this case, $S_{u_1,u_2}(\omega,t)=\hat{u}_1(\omega,t)\overline{\hat{u}_2(\omega,t)}$, so that $S_{u_1,u_2}(\omega,t)\in L^1(\Re)$ as a function of ω. This is different from the definition adopted for what are commonly described as signals of bounded power [11], where

$$R_{u_1,u_2}(\lambda,t)=\lim_{Z\to\infty}\frac{1}{2Z}\int_{-Z}^{Z}u_1(\zeta+\lambda,t)u_2(\zeta,t)d\zeta$$

Using the definition of $R_{u_1,u_2}(\lambda,t)$ provided by Eq. (6) the correlation moment $\Gamma_{u_1,u_2,r}(t)$ is given by:

$$\Gamma_{u_1,u_2,r}(t)=\int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}u_1(\zeta+\lambda,t)u_2(\zeta,t)d\zeta d\lambda. \tag{7}$$

Given the existence of the rth moment, $\Gamma_{u_1,u_2,r}(t)$, of a correlation function, $R_{u_1,u_2}(\lambda,t)$, as introduced in Eq. (1), the various length scales of $R_{u_1,u_2}(\lambda,t)$ are sensibly defined by the normalized integral measure, $\langle\lambda^r,u_1,u_2\rangle$, defined by

$$\langle\lambda^r,u_1,u_2\rangle=\frac{\displaystyle\int_{-\infty}^{\infty}\lambda^r R_{u_1,u_2}(\lambda,t)d\lambda}{\displaystyle\int_{-\infty}^{\infty}R_{u_1,u_2}(\lambda,t)d\lambda}. \tag{8}$$

Expressed in terms of correlation moments for the functions u_1 and u_2, this is

$$\langle\lambda^r,u_1,u_2\rangle=\frac{\Gamma_{u_1,u_2,r}(t)}{\Gamma_{u_1,u_2,0}(t)}. \tag{9}$$

The concept of a length scale over which correlation is significant lies at the heart of the treatment of many stochastic processes such as turbulent flow [12], and time series analysis [10]. In such areas a particularly important quantity is the width of the auto-

correlation function, $R_{u,u}(\lambda,t)$, which is often termed the coherence length or coherence time depending upon the context. Explicit definitions of correlation moments appear to be limited to the case of auto-correlation functions where they have been used as a tool in asymptotic analysis [13].

The auto correlation is an even function of λ and thus has a zero value first moment, consequently the width measure must be obtained from the non-zero second moment, i.e., as a root mean square. Thus it is proposed to adopt this as a definition for the auto-correlation function width, $\lambda_w(t)$, hence:

$$\lambda_w^2(t) = \langle \lambda^2, u, u \rangle. \tag{10}$$

2.2 Nature of the Variables u_1 and u_2. The definition of the correlation function, Eq. (6), is deterministic and so it is important to confirm that it will also prove useful in the stated aim of exploring the evolution of determinism in PDEs with stochastic initial conditions. Consider the case where the functions u_1 and u_2 are members of ensembles which define two random variables. Then Eq. (6) determines the correlation functions by integration over ζ from single members of the ensemble. As the key statistic required, $\lambda_w^2(t)$, is a moment of the spatial auto-correlation function $R_{u,u}(\lambda,t)$ it is necessary that u_1 and u_2, must belong to a class of band limited functions, u, which exhibit spatial ergodicity, at least up to the second order statistic. Consequently, it is necessary to determine whether the restrictions already placed upon u_1 and u_2 are sufficient to ensure this or if further constraints are required.

The functions $u_1, u_2 \in L^2(\Re)$ and so a first order statistic is independent of the origin about which this is estimated. Thus the class of functions containing u_1, u_2 is wide sense stationary. Furthermore, a special case of the requirement that $R_{u_1,u_2}(0,t) < \infty$ and $\lim_{|\lambda|\to\infty} R_{u_1,u_2}(\lambda,t) \to 0$ ensures that: $R_{u,u}(0,t) < \infty$ and $\lim_{\lambda\to\infty} R_{u,u}(\lambda,t) \to 0$ and hence u_1 and u_2 are second order stationary. These properties are sufficient to satisfy the required degree of ergodicity thus the inherent restrictions imposed upon u_1 and u_2 are already adequate.

It is important to emphasize that the required stationarity is spatial, with time, t, serving solely as a parameter. Clearly it is equally possible to develop an equivalent form of correlation moment analysis based around time with spatial variables as parameters.

As will be demonstrated, the present formalism converts PDEs into ODEs whose dependent variables are correlation functions, $R_{u_1,u_2}(\lambda,t)$, thus the initial conditions become initial correlation functions, $R_{u_1,u_2}(\lambda,0)$. Furthermore, it is important to reiterate that the PDEs considered are not treated as stochastic equations. Hence the manipulation follows the rules of conventional differential calculus and not those arising from the Îto integral [5–7]. For example, the simple linear diffusion equation considered in Sec. 3 is treated as a conventional PDE, with the solution taking the physical role of a concentration, and not as in Kolmogorov's forward equation yielding a probability density function.

2.3 Further Conditions on $R_{u_1,u_2}(\lambda,t)$ and $u_1(\zeta,t)$, $u_2(\zeta,t)$. Clearly for Eq. (9) to have meaning, we must require that $\Gamma_{u_1,u_2,0} \neq 0$ which necessitates that the appropriate correlation function $R_{u_1,u_2}(\lambda,t)$ has a non-zero integral. The consequences of this on u_1 and u_2 can be found by first changing the variable in Eq. (7) to give:

$$\Gamma_{u_1,u_2,r}(t) = \int_{-\infty}^{\infty} u_2(\zeta,t) \int_{-\infty}^{\infty} (\beta-\zeta)^r u_1(\beta,t) d\beta d\zeta. \tag{11}$$

This shows that the requirement, $\Gamma_{u_1,u_2,0} \neq 0$, corresponds to the condition:

$$\int_{-\infty}^{\infty} u_1(\zeta,t) d\zeta \neq 0 \text{ and } \int_{-\infty}^{\infty} u_2(\zeta,t) d\zeta \neq 0, \tag{12}$$

which is of course also clear from the spectral form $\Gamma_{u_1,u_2,0}(t) = S(0,t) = \hat{u}_1(0,t)\overline{\hat{u}_2(0,t)}$ in Eq. (3). The condition $\Gamma_{u_1,u_2,0} \neq 0$ prohibits correlation functions with a zero local mean value over the region where $|R_{u_1,u_2}(\lambda,t)|$ is significant. The requirement for u_1 and u_2 to be in L^2 already excludes periodic and almost periodic functions [14,15], together with specific realizations from sets of strictly wide band random functions with infinite support. This additional requirement also disallows L^2 functions formed by windowing zero-mean periodic functions over an integer number of periods.

Equation (11) raises the broader question of the relationship of the correlation moments to the moments of u_1 and u_2. Expansion of Eq. (11) shows that this relationship is:

$$\Gamma_{u_1,u_2,r}(t) = \sum_{k=0}^{r} (-1)^k \frac{r!}{k!(r-k)!} \left(\int_{-\infty}^{\infty} u_1(\beta,t) \beta^{r-k} d\beta \right) \times \left(\int_{-\infty}^{\infty} u_2(\zeta,t) \zeta^k d\zeta \right). \tag{13}$$

While the restrictions placed upon u_1, u_2, exclude strictly periodic and almost periodic functions [12], as well as trajectories of stationary random functions, the constraints are weak enough to encompass a large range of practical cases of interest. These can be qualitatively categorized as propagating wave packets with non-zero mean or pulse-like functions, ensembles of which exhibit band limited stochastic characteristics. As will become evident later, such behavior is typical of diffusion equations and nonlinear wave equations such as the well known class of Kortweg de-Vries equations frequently used to model solitons.

2.4 Correlation Moment Relationships. The route which will be described here for determining $\lambda_w(t)$, without explicitly solving what is typically a non-linear partial differential equation in time and space, is based upon applying the integral defined in Eq. (7) to the system of interest. The effect of the definite integral, when viewed as an operator, is to convert the terms in the partial differential equation defining the system of interest into a set of ordinary differential equations in time, t. This set relates the moments, $\Gamma_{u_1,u_2,r}(t)$, of various orders of cross correlation and the goal is to use these equations to determine the rms length scale, $\lambda_w(t)$, of the auto-correlation function of the solution of the PDE. For purposes of clarity the present treatment involves time, t, and one spatial dimension, ζ, however the extension to higher dimensions is immediate.

2.5 Properties of Correlation Moments. The following properties are used in the development of the correlation moment technique. They can be derived from Eqs. (6) and (7) using elementary techniques.

2.6 Spatial Derivative Properties.

$$\frac{\partial^p R_{u_1,u_2}(\lambda,t)}{\partial \lambda^p} = (-1)^p \int_{-\infty}^{\infty} u_1(\zeta+\lambda,t) \frac{\partial^p u_2(\zeta,t)}{\partial \zeta^p} d\zeta \tag{14}$$

Given that it is required for all r and p of interest, including $p=0$, that the following holds

$$\lim_{|\lambda|\to\infty} \left\{ \lambda^r \frac{\partial^p R_{u_1,u_2}(\lambda,t)}{\partial \lambda^p} \right\} = 0 \tag{15}$$

then using integration by parts produces for $r < p$

$$\int_{-\infty}^{\infty} \lambda^r \frac{\partial^p R_{u_1,u_2}(\lambda,t)}{\partial \lambda^p} d\lambda = 0 \tag{16}$$

and for $r \geqslant p$

$$\int_{-\infty}^{\infty} \lambda^r \frac{\partial^p R_{u_1,u_2}(\lambda,t)}{\partial \lambda^p} d\lambda = (-1)^p \frac{r!}{(r-p)!} \int_{-\infty}^{\infty} \lambda^{r-p} R_{u_1,u_2}(\lambda,t) d\lambda \tag{17}$$

with the usual definition of $0!=1$. The spatial derivative relationship can be summarized directly in terms of u_1 and u_2 as follows:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda^r u_1(\eta+\lambda,t) \frac{\partial^p u_2(\eta,t)}{\partial \eta^p} d\eta d\lambda$$

$$\begin{cases} 0 & r < p \\ \dfrac{r!}{(r-p)!} \Gamma_{u_1,u_2,r-p}(t) & r \geqslant p \end{cases} \tag{18}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda^r u_1(\eta-\lambda,t) \frac{\partial^p u_2(\eta,t)}{\partial \eta^p} d\eta d\lambda$$

$$\begin{cases} 0 & r < p \\ (-1)^r \dfrac{r!}{(r-p)!} \Gamma_{u_1,u_2,r-p}(t) & r \geqslant p \end{cases} \tag{19}$$

2.7 Symmetry Properties.

$$R_{u_1,u_2}(\lambda,t) = R_{u_2,u_1}(-\lambda,t) \tag{20}$$

and hence:

$$\int_{-\infty}^{\infty} \lambda^r R_{u_1,u_2}(\lambda,t) d\lambda = (-1)^r \int_{-\infty}^{\infty} \lambda^r R_{u_2,u_1}(\lambda,t) d\lambda. \tag{21}$$

Furthermore, a rather obvious, but nonetheless useful property is:

$$\int_{-\infty}^{\infty} \lambda^r R_{u_1,u_2}(-\lambda,t) d\lambda = (-1)^r \int_{-\infty}^{\infty} \lambda^r R_{u_1,u_2}(\lambda,t) d\lambda \tag{22}$$

If the indices r and p induce opposite symmetry, i.e., one is an even number while the other is odd, then for the special case of the auto-correlation function:

$$\int_{-\infty}^{\infty} \lambda^r \frac{\partial^p R_{u,u}(\lambda,t)}{\partial \lambda^p} d\lambda = 0 \tag{23}$$

Given that $R_{u_1,u_2}(\lambda,t)$ is an even function of λ the odd order moments of the auto-correlation moments are zero, i.e.

$$\Gamma_{u,u,r}(t) = 0, \quad r \text{ odd} \tag{24}$$

2.8 Special Cases of u_1 and u_2. The treatment up to Eq. (24) applies to correlation moments in general, however in the specific cases considered here a particular class of correlation moments will be explored in detail. Writing the PDE solution as $u(\zeta,t)$, a wide range of PDEs met in applications contain terms of the form $\partial^p u^n(\zeta,t)/\partial t^v \partial \zeta^{p-v}$. This suggests that the following special case of u_1 and u_2 are of interest, i.e.: $u_1=u^m(\zeta,t)$, and $u_2=\partial^p u^n(\zeta,t)/\partial t^v \partial \zeta^{p-v}$. The corresponding rth correlation moment then has the form:

$$\Gamma_{u_1,u_2,r}(t) = \int_{\infty}^{\infty} \lambda^r \int_{-\infty}^{\infty} u^m(\zeta+\lambda,t) \frac{\partial^p u^n(\zeta,t)}{\partial t^v \partial \zeta^{p-v}} d\zeta d\lambda \tag{25}$$

This is evidently related to the form:

$$\Gamma_{u_1,u_2,r}(t) = \int_{-\infty}^{\infty} \lambda^r \int_{-\infty}^{\infty} u^m(\zeta+\lambda,t) u^n(\zeta,t) d\zeta d\lambda \tag{26}$$

in which $u_1=u^m(\zeta,t)$ and $u_2=u^n(\zeta,t)$, suggesting that the following extension of the notation will prove useful:

$$\Gamma_{m,n,r}(t) = \Gamma_{u^m,u^n,r}(t) \tag{27}$$

Thus the version of Eq. (11) which demonstrates the relationship of $\Gamma_{m,n,r}(t)$ to the moments of $u(\zeta,t)$ is:

$$\Gamma_{m,n,r}(t) = \int_{-\infty}^{\infty} u^n(\zeta,t) \int_{-\infty}^{\infty} (\beta-\zeta)^r u^m(\beta,t) d\beta d\zeta \tag{28}$$

that evaluates to:

$$\Gamma_{m,n,r}(t) = \sum_{k=0}^{r} (-1)^k \frac{r!}{k!(r-k)!} \int_{-\infty}^{\infty} u^m(\zeta,t) \zeta^{r-k} d\zeta \int_{-\infty}^{\infty} u^n(\zeta,t) \zeta^k d\zeta \tag{29}$$

The corresponding notation for the correlation functions is:

$$R_{m,n}(\lambda,t) = R_{u^m,u^n}(\lambda,t) \tag{30}$$

A very useful property involving this special class of u_1 and u_2 applies to the first order time derivative of $R_{m,n}(\lambda,t)$:

$$\frac{\partial R_{m,n}(\lambda,t)}{\partial t} = \int_{-\infty}^{\infty} \{ n u^m(\zeta+\lambda,t) u^{n-1}(\zeta,t) + m u^n(\zeta-\lambda,t) u^{m-1}(\zeta,t) \} \frac{\partial u(\zeta,t)}{\partial t} d\zeta \tag{31}$$

It is this relationship that indicates the form of the weight function, i.e., the functions, $\{nu^m(\zeta+\lambda,t)u^{n-1}(\zeta,t)+mu^n(\zeta-\lambda,t)u^{m-1}(\zeta,t)\}$, used to multiply the PDE prior to taking correlation moments and the subsequent generation of the system of ODEs. The spatial derivative forms corresponding to Eqs. (18) and (19) are:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda^r u^n(\eta+\lambda,t) \frac{\partial^p u^m(\eta,t)}{\partial \eta^p} d\eta d\lambda$$

$$\begin{cases} 0 & r < p \\ \dfrac{r!}{(r-p)!} \Gamma_{n,m,r-p}(t) & r \geqslant p \end{cases} \tag{32}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda^r u^n(\eta-\lambda,t) \frac{\partial^p u_m(\eta,t)}{\partial \eta^p} d\eta d\lambda$$

$$\begin{cases} 0 & r < p \\ (-1)^r \dfrac{r!}{(r-p)!} \Gamma_{n,n,r-p}(t) & r \geqslant p \end{cases} \tag{33}$$

2.9 Zeroth Order Moments. As will be shown subsequently, Eqs. (31)–(33), together with the symmetry properties, provide the basic tool kit for implementing correlation moment analysis. A particularly useful outcome of this set of equations is that for PDEs which are first order in time, t, and have no terms that are explicit functions of $u(\eta,t)$ (i.e., all terms contain derivatives of η) then the zeroth order moments, $\Gamma_{m,n,0}(t)$ are actually time independent.

3 Application of the Correlation Moment Technique to a Linear System

To illustrate the correlation moment methodology, the length scale $\lambda_w(t)$ of the auto-correlation function $R_{u,u}(\lambda,t)$ will be determined for the classical linear diffusion equation:

$$\frac{\partial u(\zeta,t)}{\partial t} = D\frac{\partial^2 u(\zeta,t)}{\partial \zeta^2} \tag{34}$$

where D is the diffusion coefficient. Physically the solution $u(\zeta,t)$ is the concentration of some species and hence $u(\zeta,t) \geqslant 0$. The boundary condition is the standard condition applied to $u(\zeta,t)$ throughout the work reported on here, i.e., $\lim_{|\zeta|\to\infty} u(\zeta,t)=0$. At this point the initial conditions are not specified.

The first goal is to convert Eq. (34) into an ODE in terms of the correlation moments, $\Gamma_{1,1,r}(t)$, of the auto-correlation functions $R_{1,1}(\lambda,t)$, which will eventually enable the calculation of $\lambda_w(t)$. This is achieved by recognizing that Eq. (31) links the time derivatives of $R_{1,1,r}(\lambda,t)$ and $u(\zeta,t)$. The lowest order correlations are obtained for $m=1$, $n=1$ thus suggesting multiplying Eq. (34) through by the weight function, $u(\zeta+\lambda,t)+u(\zeta-\lambda,t)$. The correlation moments are then generated by integrating over ζ and λ:

$$\int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}\{u(\zeta+\lambda,t)+u(\zeta-\lambda,t)\}\frac{\partial u(\zeta,t)}{\partial t}d\zeta d\lambda = D\int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}\{u(\zeta+\lambda,t)+u(\zeta-\lambda,t)\}\frac{\partial^2 u(\zeta,t)}{\partial \zeta^2}d\zeta d\lambda \tag{35}$$

Using the $\Gamma_{m,n,r}(t)$ notation from Eq. (27) plus the property defined by Eq. (14) allows Eq. (35) to be rewritten as:

$$\frac{d\Gamma_{1,1,r}(t)}{dt} = D\int_{-\infty}^{\infty}\lambda^r\left\{\frac{\partial^2 R_{1,1}(\lambda,t)}{\partial\lambda^2}+\frac{\partial^2 R_{1,1}(-\lambda,t)}{\partial\lambda^2}\right\}d\lambda. \tag{36}$$

Invoking Eq. (17) on the right hand side (RHS) of Eq. (36) produces for $r \geqslant 0$:

$$\frac{d\Gamma_{1,1,r}(t)}{dt} = Dr(r-1)\int_{-\infty}^{\infty}\lambda^{r-2}\{R_{1,1}(\lambda,t)+R_{1,1}(-\lambda,t)\}d\lambda \tag{37}$$

Writing right hand side of Eq. (37) in terms of the $\Gamma_{m,n,r}(t)$ notation and employing the symmetry properties produces:

$$\frac{d\Gamma_{1,1,r}(t)}{dt} = 2r(r-1)D\Gamma_{1,1,r-2}(t) \text{ for } r \geqslant 2 \tag{38}$$

and from Eqs. (16) and (36)

$$\frac{d\Gamma_{1,1,r}(t)}{dt} = 0 \text{ for } r < 2. \tag{39}$$

Equations (37)–(39) generate a system of relationships between the correlation moments of the auto-correlation function $R_{1,1}(\lambda,t)$. As it is the second moment which defines the rms width, $\lambda_w(t)$, it is the first three that are of interest, i.e.

$$\frac{d\Gamma_{1,1,0}(t)}{dt} = 0 \tag{40}$$

(which is in any case required by the arguments in Sec. 2.9) and

$$\Gamma_{1,1,1}(t) = 0 \tag{41}$$

which is in any case necessitated by Eq. (24) that requires: $\Gamma_{1,1,1}(t)=0$. The relationship for the $r=2$ moment of the auto-correlation function $R_{1,1}(\lambda)$ is thus:

$$\frac{d\Gamma_{1,1,2}(t)}{dt} = 4D\Gamma_{1,1,0}(t) \tag{42}$$

Equation (42) provides an explicit relationship for $\Gamma_{1,1,2}(t)$ in terms of $\Gamma_{1,1,0}$ and given that Eq. (39) shows $\Gamma_{1,1,0}$ is time independent, then Eq. (42) can be divided by $\Gamma_{1,1,0}$ which produces from Eq. (40):

$$d\frac{\Gamma_{1,1,2}(t)}{\Gamma_{1,1,0}}{dt} = 4D \tag{43}$$

Equations (8) and (9) give:

$$\frac{\Gamma_{1,1,2}(t)}{\Gamma_{1,1,0}} = \lambda_w^2(t) \tag{44}$$

thus Eqs. (33) and (44) provide a first order ordinary differential equation for $\lambda_w(t)$ of the form:

$$\frac{d\lambda_w^2(t)}{dt} = 4D \tag{45}$$

Hence the time dependence of the mean square width of the auto-correlation function is given by:

$$\lambda_w^2(t) = 4Dt + \lambda_w^2(0) \tag{46}$$

Note that by Eq. (28)

$$\lambda_w^2(0) = 2\frac{\int_{-\infty}^{\infty}u(\zeta,0)\zeta^2 d\zeta}{\int_{-\infty}^{\infty}u(\zeta,0)d\zeta} - 2\frac{\left(\int_{-\infty}^{\infty}u(\zeta,0)\zeta d\zeta\right)^2}{\left(\int_{-\infty}^{\infty}u(\zeta,0)d\zeta\right)^2} \tag{47}$$

3.1 Initial Conditions. The simplest initial condition is where $u(\zeta,0)$ is a continuous approximation to an initial point source of diffusible material, i.e., $u(\zeta,0)\simeq Q_0\delta(\zeta)$, where Q_0 is the total quantity of material present. This produces $\lambda_w^2(0)\simeq 0$ giving a length scale characterized by $\sqrt{Dt}$ in agreement with the well known result for fundamental solution of Eq. (34) [16].

Given the interest expressed in Sec. 1 over the development of determinism, an initial condition with random characteristics is considered next. A simple example which satisfies the requirement that $u(\zeta,t)\geqslant 0$ is a set of N narrow positive pulse-like functions (continuous in ζ), $p(\zeta-\zeta_k,r_p)$, of unit area and length scale r_p, which are distributed at random locations, ζ_k, over the interval $-\zeta_0 \leqslant \zeta \leqslant \zeta_0$. Thus: $u(\zeta,0)=\Sigma_{k=1}^{N}p(\zeta-\zeta_k,r_p)$. The pulse length scale, r_p, is chosen such $r_p \ll \zeta_0$ so that on the scale, ζ_0, of the initial condition region the pulses can be approximated by delta functions.

If the pulses have a Poisson spatial distribution then the auto correlation of $u(\zeta,0)$ is well approximated by the sum of a triangular pedestal (of peak value $N^2/2\zeta_0$) spanning $-2\zeta_0 \leqslant \lambda \leqslant 2\zeta_0$ and an approximate delta function (the actual width scale is $2r_p$) of weight $N/2\zeta_0$ located at $\zeta=0$. Lee [10], considers the corresponding result for the limiting case where the pulses are true delta functions and the domain is infinite. Evaluation of the correlation length scale at $t=0$ produces a value for the constant $\lambda_w^2(0)$ in Eq. (46) of $\lambda_w^2(0)\simeq 4\zeta_0^2/3N/1+N/2$.

3.2 Large *t* Behavior. At large times the diffusing substance will have spread over large distances compared to the length scale, λ_0, of the initial condition. Thus the effect of the original detail in the initial condition region is lost and the diffusing profile becomes similar to that arising from a delta function initial condition. Under such circumstances the full solutions [16], predict that the diffusion length scales are again defined by $\sqrt{Dt}$, which is in agreement with the form of the large t limit of Eq. (46).

4 Points Arising From the Linear Example

The linear example demonstrates how to derive a system of correlation moment equations from a PDE and reproduce the well known behavior conventionally derived from the explicit solution. In addition, this case also reveals some general features about the manner in which the types of terms present in the PDE affect the

correlation moment equations. As will be discussed in Sec. 4.1, such relationships also provide information directly about properties of the correlation functions and of the PDE solutions in cases where obtaining explicit analytical evaluation of $u(\zeta,t)$ is far more difficult than in the linear case.

4.1 Effects of Various Orders of Derivative in the PDE on Zeroth Order Moments. An interesting point arrises as a consequence of Eq. (16) which requires that a spatial derivative of order p will not contribute terms to the moment relations generated for $r \leqslant p$. The first order time derivative term does not have this constraint and makes a contribution to all such equations. Thus, as described in Sec. 2.9, a PDE containing time derivatives of only first order (and without mixed partials) will always lead to a constant zeroth order moment, $\Gamma_{u_1,u_2,0}(t)$, unless an explicit function of $u(\zeta,t)$ is present (explicit in the sense of not being linked to ζ derivatives). This is the correlation function equation equivalent of the well known conservation equation form of a PDE [17]. Under such conditions $\Gamma_{u_1,u_2,0}$ yield so-called conserved quantities for the system, which are also known as the constants of motion.

4.2 Dissipation Terms and Conserved Quantities. A simple illustration of ease with which the correlation moment technique reveals the effect of the nature of the PDE terms on the presence or absence of conserved quantities is given by comparing the results for the diffusion equation as given in Eq. (34) with the following modified form:

$$\frac{\partial u(\zeta,t)}{\partial t} = D\frac{\partial^2 u(\zeta,t)}{\partial \zeta^2} - ku(\zeta,t) \tag{48}$$

The extra term, $ku(\zeta,t)$, describes a first order reaction which removes the diffusing species [16]. Consequently, in addition to the arguments in Secs. 2.9 and 4.1, in this case on physical grounds $u(\zeta,t)$ should not be conserved as compared to the previous example of pure diffusion where it is. Thus the presence in the PDE of this explicit $u(\zeta,t)$, or in fact any explicit function of u that does not contain spatial derivatives, can be viewed as a form of dissipation. Repeating the correlation moment analysis shows the effect of the $ku(\zeta,t)$ term on the zeroth order moment equation, which becomes:

$$\frac{d\Gamma_{1,1,0}(t)}{dt} = -2k\Gamma_{1,1,0}(t) \tag{49}$$

and thus $\Gamma_{1,1,0}(t)=\Gamma_{1,1,0}(0)e^{-2kt}$. Equations (30) yields: $\Gamma_{1,1,0}(0)=\{\int_{-\infty}^{\infty}u(\zeta,0)d\zeta\}^2$, confirming a decay in the total quantity of the diffusing species, which has the explicit form $\int_{-\infty}^{\infty}u(\zeta,t)d\zeta = e^{-kt}\int_{-\infty}^{\infty}u(\zeta,0)d\zeta$.

The ability to determine the time dependence of the total quantity of diffusing material without explicitly solving the PDE is just one of a whole raft of results that potentially arise from the link between the properties of correlation moment equations to the structure of the PDE of interest.

5 Application of the Correlation Moment Technique to a NonLinear System

Having established the correlation moment methodology for the case of a classical linear PDE, the approach will be applied to a non-linear PDE.

5.1 The KdV Equation. The application considered is the non-linear dispersive wave equation usually termed the KdV equation [17–22], which has attracted much attention as a soliton generating system, a useful introduction to which is given in [18]. The equation is presented here in the form:

$$\frac{\partial u(\zeta,t)}{\partial t} - 3\frac{\partial u^2(\zeta,t)}{\partial \zeta} + \frac{\partial^3 u(\zeta,t)}{\partial \zeta^3} = 0 \tag{50}$$

The restrictions imposed here upon $u(\zeta,t)$ naturally induce the boundary condition: $\lim_{|\zeta|\to\infty} u(\zeta,t)=0$.

5.2 Correlation Moment Relationships. The procedure illustrated for the linear diffusion equation is now applied to the non-linear KdV equation. The basic approach is the same, i.e., equations relating correlation moments are generated by multiplying the KdV equation by an appropriate weight function in order to exploit the property in Eq. (31). The correlation moments are then generated by application of the definite integral in Eq. (26). The main difference that arises in the non-linear case is that, while Eqs. (38) and (39) are uncoupled in the sense that they link moments of the same order of correlation function in the linear example, the corresponding equations in the non-linear case contain moments of mixed orders of correlation functions and consequently the ODE system is coupled. In the present example this requires developing two systems of equations rather than one as in the linear case and thus employs two different multiplicative weight functions. The required form for the new weights arrises in a natural manner as the analysis proceeds.

5.3 First System of Correlation Moment Equations. As with the linear example motivated by Eq. (31) the PDE is first multiplied by the factor $\{u(\zeta+\lambda,t)+u(\zeta-\lambda,t)\}$ then integral given by Eq. (26) is applied. This generates an equation that, as will be seen, relates the various moments of the first order auto-correlation function, $R_{1,1}(\lambda,t)$ to those of the cross-correlation function $R_{1,2}(t)$:

$$\begin{aligned}&\int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}\{u(\zeta+\lambda,t)+u(\zeta-\lambda,t)\}\frac{\partial u(\zeta,t)}{\partial t}d\zeta d\lambda\\ &\quad-3\int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}\{u(\zeta+\lambda,t)+u(\zeta-\lambda,t)\}\frac{\partial u^2(\zeta,t)}{\partial \zeta}d\zeta d\lambda\\ &\quad+\int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}\{u(\zeta+\lambda,t)+u(\zeta-\lambda,t)\}\frac{\partial^3 u(\zeta,t)}{\partial \zeta^3}d\zeta d\lambda = 0\end{aligned} \tag{51}$$

Proceeding as in the linear example and invoking the properties defined by Eqs. (14) and (23) causes Eq. (51) to become:

$$\begin{aligned}&\frac{d\Gamma_{1,1,r}(t)}{dt} + 3\{1+(-1)^r\}\int_{-\infty}^{\infty}\frac{\partial R_{1,2}(\lambda,t)}{\partial\lambda}\lambda d\lambda\\ &\quad-\{1+(-1)^r\}\int_{-\infty}^{\infty}\frac{\partial^3 R_{1,1}(\lambda,t)}{\partial\lambda^3}\lambda^r d\lambda = 0\end{aligned} \tag{52}$$

Evaluating Eq. (52) using Eqs. (14)–(19) produces: For $r=0$

$$\frac{d\Gamma_{1,1,0}(t)}{dt} = 0 \tag{53}$$

Equation (53) is confirmed by the fact that the presence of no explicit functions of $u(\eta,t)$ (as opposed to its spatial derivatives) ensures that the $r=0$ moments of all correlation functions, $R_{m,n}(\lambda,t)$, are time independent. For $r=1$

$$\frac{d\Gamma_{1,1,1}(t)}{dt} = 0 \tag{54}$$

which is confirmed by Eq. (24) that requires that the odd moments of auto-correlations are zero. For $r=2$

$$\frac{d\Gamma_{1,1,2}(t)}{dt} = 12\Gamma_{1,2,1}(t) \tag{55}$$

5.4 Second System of Correlation Moment Equations. The presence of moments from two different orders of correlation function require additional systems of equations to explicitly obtain $\lambda_w(t)$. The presence of the $1,2,r$ correlation function in Eq. (55), together with the property that only the first order time derivative will contribute to the $r=0$ moment relationship, requires a further equation system containing the term $\partial R_{1,2}(\lambda,t)/\partial t$. This can be achieved by replacing the multiplying factor $u(\zeta+\lambda,t)+u(\zeta-\lambda,t)$ employed in deriving the first moment equation with the term $u^2(\zeta+\lambda,t)+2u(\zeta-\lambda,t)u(\zeta,t)$.

The equivalent to Eq. (35) then becomes:

$$\int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}\{u^2(\zeta+\lambda,t)+2u(\zeta,t)u(\zeta-\lambda,t)\}\frac{\partial u(\zeta,t)}{\partial t}d\zeta d\lambda - 3\int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}\{u^2(\zeta+\lambda,t)+2u(\zeta,t)u(\zeta-\lambda,t)\}\frac{\partial u^2(\zeta,t)}{\partial\zeta}d\zeta d\lambda + \int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}\{u^2(\zeta+\lambda,t)+2u(\zeta,t)u(\zeta-\lambda,t)\}\frac{\partial^3 u(\zeta,t)}{\partial\zeta^3}d\zeta d\lambda = 0 \tag{56}$$

It is convenient to recast the last left hand side term in Eq. (56) as:

$$\int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}2u(\zeta,t)u(\zeta-\lambda,t)\frac{\partial^3 u(\zeta,t)}{\partial\zeta^3}d\zeta d\lambda = \int_{-\infty}^{\infty}\lambda^r\int_{-\infty}^{\infty}u(\zeta-\lambda,t)\times\left\{\frac{1}{2}\frac{\partial^3 u^2(\zeta,t)}{\partial\zeta^3}-\frac{3}{2}\frac{\partial\left\{\frac{\partial u(\zeta,t)}{\partial\zeta}\right\}^2}{\partial\zeta}\right\}d\zeta d\lambda \tag{57}$$

The factor $\partial\{\partial u(\zeta,t)/\partial\zeta\}^2/\partial\zeta$ will be represented by the ζ derivative of a function $\Phi(\zeta,t)$ satisfying the same conditions as $u(\zeta,t)$, thus:

$$\frac{\partial\Phi(\zeta,t)}{\partial\eta}=\left\{\frac{\partial u(\zeta,t)}{\partial\zeta}\right\}^2 \tag{58}$$

The corresponding cross-correlation function, denoted as, $R_{u,\Phi}(\lambda,t)$, is given by:

$$R_{u,\Phi}(\lambda,t)=\int_{-\infty}^{\infty}u(\zeta+\lambda,t)\Phi(\zeta,t)d\zeta \tag{59}$$

The associated correlation moments are: $\Gamma_{u,\Phi,r}(t)$. The presence of this term may appear to be inconvenient but, as will be shown, its explicit properties are not required in deriving the results of interest.

Using the above Eq. (56) takes the form:

$$\frac{d\Gamma_{2,1,r}(t)}{dt}+\int_{-\infty}^{\infty}\left\{3\frac{\partial R_{2,2}(\lambda,t)}{\partial\lambda}+(-1)^r 4\frac{\partial R_{1,3}(\lambda,t)}{\partial\lambda}-\frac{\partial^3 R_{2,1}(\lambda,t)}{\partial\lambda^3} - (-1)^r\frac{\partial^3 R_{1,2}(\lambda,t)}{\partial\lambda^3}-3(-1)^r\frac{\partial^2 R_{u,\Phi}(\lambda,t)}{\partial\lambda^2}\right\}\lambda^r d\lambda=0 \tag{60}$$

Using Eqs. (14)–(19) to evaluate Eq. (60) produces: For $r=0$

$$\frac{d\Gamma_{2,1,0}}{dt}=0 \tag{61}$$

which conforms to the requirement that for this type of PDE the $r=0$ correlation moments of all correlation functions are time independent. For $r=1$

$$\frac{d\Gamma_{2,1,1}(t)}{dt}=3\Gamma_{2,2,0}-4\Gamma_{1,3,0} \tag{62}$$

Using the symmetry properties of Eqs. (21) and (22) in Eq. (62) produces:

$$\Gamma_{2,1,1}(t)=-\Gamma_{1,2,1}(t) \tag{63}$$

and hence Eq. (64) becomes:

$$\frac{d\Gamma_{1,2,1}(t)}{dt}=-3\Gamma_{2,2,0}+4\Gamma_{1,3,0} \tag{64}$$

Combining Eqs. (55) and (64) gives:

$$\frac{d^2\Gamma_{1,1,2}(t)}{dt^2}=-12\{3\Gamma_{2,2,0}-4\Gamma_{1,3,0}\} \tag{65}$$

Finally, the desired length scale $\lambda_\omega^2(t)=\Gamma_{1,1,2}(t)/\Gamma_{1,1,0}$ is formed producing:

$$\frac{d^2\lambda_\omega^2(t)}{dt^2}=\left\{48\frac{\Gamma_{1,3,0}}{\Gamma_{1,1,0}}-36\frac{\Gamma_{2,2,0}}{\Gamma_{1,1,0}}\right\}\frac{t^2}{2}+\left.\frac{d\lambda_\omega^2(t)}{dt}\right|_{t=0}t+\lambda_\omega^2(0) \tag{66}$$

Using the initial conditions for $d\lambda_\omega^2(t)/dt|_{t=0}$ from Eq. (55) together with the definition of $\lambda_\omega^2(0)$ produces:

$$\lambda_w^2(t)=\left\{48\frac{\Gamma_{1,3,0}}{\Gamma_{1,1,0}}-36\frac{\Gamma_{2,2,0}}{\Gamma_{1,1,0}}\right\}\frac{t^2}{2}+12\frac{\Gamma_{1,2,1}(0)}{\Gamma_{1,1,0}}t+\frac{\Gamma_{1,1,2}(0)}{\Gamma_{1,1,0}} \tag{67}$$

As with the diffusion example by using Eq. (29) the correlation width in Eq. (67) can be readily expressed in terms of the spatial moments of various powers of the initial condition $u(\zeta,0)$.

5.5 Conditions on $\lambda_w(t)$. To be meaningful as a real physical measure of the width of the auto-correlation function $\lambda_w(t)$ as given by Eq. (67) must be both real and $\lambda_w(t)\geqslant 0$. Given that $\Gamma_{1,1,0}\geqslant 0$ and $\Gamma_{1,1,2}\geqslant 0$ these conditions reduce to:

$$\Gamma_{1,2,1}(0)\geqslant 0 \tag{68}$$

and

$$\Gamma_{1,3,0}\geqslant\frac{3}{4}\Gamma_{2,2,0} \tag{69}$$

Inequalities (68) and (69) limit the class of initial conditions for which Eq. (67) is valid and crudely speaking suggest that $u(\zeta,0)$ should be positive for most ζ. This is not altogether surprising as there is a large class of solutions of the KdV equation for which $u(\zeta,t)\geqslant 0$.

5.6 Examples of Initial Conditions. Equation (67) shows three distinct classes of behavior which in general reflect short, medium, and large time-scales but for specific initial conditions may also apply more generally. For small times, or dominance by the factor $\Gamma_{1,1,2}(0)$, the $\lambda_w(t)$ is constant indicating either little evolution or a linear dispersion characteristic which is common in KdV solutions, Refs. [18–20]. For moderate times, or $\Gamma_{1,1,2}(0)$ dominance, $\lambda_w(t)$ grows as $t^{1/2}$, i.e., diffusion-like behavior and for large time, or dominance by the factor $\{48\Gamma_{1,3,0}(0)-\Gamma_{2,2,0}(0)\}/\Gamma_{1,1,0}(0)$, $\lambda_w(t)$ grows linearly with time. The more general quadratic time dependence describes wave packets whose shape evolves with time. A Gaussian initial condition of the form $u(\zeta,0)=e^{-(\zeta/r_p)^2}$ satisfies both the fundamental existence requirements for the correlation moments and the inequality (68) and inequality (69). Specifically inequality (69) evaluates to: $\frac{1}{\sqrt{3}}\geqslant\frac{3}{8}$ and inequality (68) yields: $\Gamma_{1,2,1}(0)=0$. This second result reflects the fact that if $u(\zeta,0)$ is an even function of ζ then for all $m,n\Gamma_{m,n,r}(0)=0$ for r odd. For such an initial condition Eq. (67) produces:

$$\lambda_w(t)=\sqrt{\left\{\left(\frac{4}{\sqrt{3}}-\frac{3}{2}\right)6t^2+r_p^2\right\}} \tag{70}$$

5.7 Conserved Quantities. As stated in Secs. 2.9 and 4.1 for PDEs that are first order in t and have no terms that are explicit functions of $u(\eta,t)$ (i.e., not functions of η derivatives) then the zeroth order correlation moments, $\Gamma_{m,n,0}$ are time independent and hence defined conserved quantities. This suggests that one interpretation of the correlation moments in general (for finite r) is as a natural generalization of a conserved quantity. A discussion of this was initiated in Secs. 4.1 and 4.2.

The two examples considered, i.e., the linear diffusion equation (without source sink terms) and the KdV equation are both of this type, reaffirming the well known property that the KdV system has an infinite number of conserved quantities [17]. However, without considerable extra labor the present techniques only reveal an infinite subset of these constants of motion. Comprehensive investigations of conserved quantities have either relied upon a specialist technique using a combination of the Muira and Gardener transformations [17,21,22], or alternatively the very powerful Lagrangian action integral method exploiting Noether's theorem.

6 Correlation Moment Properties of Certain Classes of Nonlinear Systems

As was made evident in Sec. 5 the key difference between the correlation moment representations of linear and nonlinear systems is that in the linear case the associated ODEs contain moments of the same order of correlation functions while for the nonlinear case the ODEs generally involve mixed orders of correlation functions. This typically requires the use of two different weight functions of the type described in Eq. (31).

However, some *potential properties* of a certain countably infinite sized class of nonlinear PDEs can be determined using a single weight function. The reason for the qualification *potential properties* is because the results discussed in this section only hold if the solution $u(\eta,t)$ satisfies the conditions previously discussed for the correlation moment functions to exist up to the order of interest. For the scope of the systems considered this is a huge question and such confirmation will inevitably be a very incremental process.

The first class considered contain generalizations of the source/sink terms discussed in Sec. 4.3 and is given by:

$$\Psi\{u(\eta,t)\}+\sum_{\ell=1}^{N}\alpha_\ell u^\ell(\eta,t)=\frac{\partial u(\eta,t)}{\partial t} \tag{71}$$

where $\Psi\{\ \}$ is a nonlinear operator which is solely a function of the spatial derivatives.

Applying the methods described in Sec. 3 using the lowest order weight factor, $u(\eta+\lambda,t)+u(\eta-\lambda,t)$, produces the following result:

$$\sum_{\ell=1}^{N}\alpha_\ell\Gamma_{1,\ell,0}(t)=\frac{\partial\Gamma_{1,\ell,0}(t)}{\partial t} \tag{72}$$

Using Eq. (13) this gives:

$$\sum_{\ell=1}^{N}\alpha_\ell\langle u(\eta,t)\rangle\langle u(\eta,t)^\ell\rangle=\frac{\partial\langle u(\eta,t)\rangle^2}{\partial t} \tag{73}$$

Physically this means that in general the spatial mean of $u(\eta,t)$ is not conserved. For the special case of $N=1$ Eq. (73) yields the same form as for the linear diffusion equation

$$\langle u(\eta,t)\rangle^2=\langle u(\eta,0)\rangle^2e^{2\alpha_1 t} \tag{74}$$

The second class of system considered explicitly excludes the generalized source/sink terms and has the form

$$\frac{\partial u(\eta,t)}{\partial\eta}+\sum_{\ell=3}^{N}\alpha_\ell\frac{\partial^\ell u^{s\ell}(\eta,t)}{\partial\eta^\ell}=\beta\frac{\partial u(\eta,t)}{\partial t} \tag{75}$$

Again using the lowest order weight function it can be shown that all such systems have the property that both $\Gamma_{1,1,0}(t)$ and $\Gamma_{1,1,2}(t)$ are constants. The first property is only moderately interesting stating that the spatial mean of $u(\eta,t)$ is conserved. However, the second is the rather surprising result that the width of the correlation function is also conserved, i.e., the length scale of statistical dependence in $u(\eta,t)$ does not change with time. This result occurs because of the effect of Eq. (16) and means that for systems of this kind the coherence length scale is solely controlled by the first spatial derivative.

If the lower index limit in Eq. (75) is reduced from $\ell=3$ to $\ell=2$, the properties of the system change. To discover the changes it is necessary to employ two sets of weight functions. Using $u(\eta+\lambda,t)+u(\eta-\lambda,t)$ generates:

$$\frac{\partial\Gamma_{1,1,2}(t)}{\partial t}=-4\alpha_1\Gamma_{1,1,1}+2\alpha_2\Gamma_{1,s,0} \tag{76}$$

in which, according to Eq. (24), $\Gamma_{1,1,1}=0$. Using the weight $su(\eta+\lambda,t)u^{s-1}(\eta,t)+u^s(\eta,t)$ produces

$$\frac{\partial\Gamma_{1,s,0}(t)}{\partial t}=0 \tag{77}$$

The result of Eqs. (6) and (7) is that:

$$\frac{\partial\Gamma_{1,1,2}(t)}{\partial t}=\text{constant} \tag{78}$$

and hence in general the correlation moment has the linear time dependence: $\Gamma_{1,1,2}(t)=\text{Const}_1+t\times\text{Const}_2$. Thus physically the coherence length scale is essentially constant for a short time then tends to grow as $\sqrt{t}$ at large t, i.e., the same statistical behavior as a diffusion system.

While the results derived in this section appear to be making remarkably broad statements about very general classes of nonlinear PDEs, it must be reemphasized that the findings are only valid for the cases where it can be shown that the conditions for the existence of the correlation moments as given in Sec. 2 are satisfied.

7 Conclusions

The work presented introduces the correlation moment technique as a method for gaining insight into the time evolution of the width of the auto-correlation function of the solutions of certain PDEs. This approach does not require the explicit solution of the equation system. For purely deterministic cases this provides information on the time dependence of fundamental length scales and for stochastic initial conditions it characterizes the evolution of statistical dependence within the solution.

The expressions for the growth of deterministic or statistical dependence can be obtained in terms of the conserved quantities and hence the initial conditions. This is achieved by exploiting the various properties of correlation moments that have been determined here.

In addition, a variety of results have been developed which show how various classes of terms in PDEs affect the structure of a sequence of correlation moment equations. This allows results to be obtained about the behavior of the PDE solution, in particular how the presence of certain types of terms affect integral measures of the solution. It is also shown how correlation moments provide a very simple, natural approach, to determining certain subsets of conserved quantities associated with the PDEs.

Acknowledgment

The author would like to acknowledge the comprehensive critique of this document performed by George Weiss, currently at Imperial College London.

References

[1] Budd, C. J., and Peletier, M. A., 2000, "Approximate Self-Similarity in Models of Geological Folding," SIAM J. Appl. Math., **60**(3), pp. 990–1016.
[2] Morris, E. L., Zienkiewicz, H. K., and Belmont, M. R., 1998, "Short Term Forecasting of the Sea Surface Shape," Int. J. Shpbld. Prog., **45**(444), pp. 383–400.
[3] Edgar, D. R., Horwood, J. M. K., Thurley, R., and Belmont, M. R., 2000, "The Effects of Parameters on the Maximum Prediction Time Possible in Short Term Forecasting of the Sea Surface Shape," Int. J. Shpbld. Prog., **47**(451), pp. 287–302.
[4] Belmont, M. R., Bogsjo, K., Horwood, J. M., and Thurley, R. W. F., 2004, "The Effect of Statistically Dependent Phases on Short Term Prediction of the Sea," Int. J. Shipbiuld. Prog., **51**(4), pp. 314–339.
[5] Îto, K., 1944, "Stochastic Integral," Proc. Imp. Acad. Tokyo, **20**, pp. 519–524.
[6] Îto, K., 1951, "On Stochastic Differential Equations," Mem. Am. Math. Soc., **4**.
[7] Doob, J. L., 1953, *Stochastic Processes*, Academic, New York.
[8] Adomain, G., 1976, "Nonlinear Stochastic Differential Equations," J. Math. Anal. Appl., **55**(2), pp. 441–452.
[9] Adomain, G., 1983, *Stochastic Systems*, Academic, New York.
[10] Lee, Y. W., 1967, *Statistical Theory of Communication*, Wiley, New York.
[11] Wiener, N., 1959, *The Fourier Integral and Certain of its Applications*, Dover, New York.
[12] Townsend, A. A., 1976, *The Structure of Turbulent Shear Flow*, Cambridge University Press, Cambridge, UK.
[13] Vom, Scheidt, Starkloff, H.-J., and Wunderlich, R., 2000, "Asymptotic Expansions of Integral Functionals of Weakly Correlated Random Processes," Z. Anal. ihre Anwend., **19**(1), pp. 255–268.
[14] Besicovitch, A. S., 1932, *Almost Periodic Functions*, Cambridge University Press, Cambridge, UK.
[15] Corduneau, C., 1968, *Almost Periodic Functions*, Wiley Interscience, New York.
[16] Crank, J., 1975, *The Mathematics of Diffusion*, Oxford University Press, Oxford, UK.
[17] Muira, R. M., Garderner, C. S., and Kruskal, M. D., 1968, "Kortweg-de Vries Equation and Generalisations, II. Existence of Conservation Laws and Constants of Motion," J. Math. Phys., **9**, pp. 1204–1209.
[18] Drazin, P. G., and Johnson, R. S., 1989, *Solitons: An Introduction*, Cambridge University Press, Cambridge.
[19] Johnson, R. S., 1980, "Water Waves and the Kortweg-deVries Equations," J. Fluid Mech., **97**, pp. 701–19.
[20] Whitham, G. B., 1974, *Linear and Nonlinear Waves*, Wiley, New York.
[21] Gardener, C. S., Greene, J. M., Kruskal, M. D., and Muira, R. M., 1967, "Method for Solving the Kortweg-de Vries Equation," Phys. Rev. Lett., **19**, pp. 1095–1097.
[22] Gardener, C. S., Greene, J. M., Kruskal, M. D., and Muira.

Interaction of a Screw Dislocation With an Arbitrary Shaped Elastic Inhomogeneity

Xu Wang

L. J. Sudak

Department of Mechanical and Manufacturing Engineering,
University of Calgary,
Calgary, Alberta,
Canada T2N-1N4

In this paper, the interaction between a screw dislocation and an arbitrary shaped elastic inhomogeneity with different material properties than the surrounding matrix is investigated. The exact solution to this problem is derived by means of complex variable methods and Faber series expansion. Specifically, the conformal mapping function maps the matrix region surrounding the inhomogeneity onto the outside of a unit circle in the image plane, while the analytic function defined in the elastic inhomogeneity is expressed in terms of a Faber series expansion. Once the series form solution is obtained, the stress fields due to the screw dislocation can be obtained. Also the image force on the screw dislocation due to its interaction with the elastic inhomogeneity is derived. Three examples of a screw dislocation interacting with (1) an equilateral triangular inhomogeneity, (2) a square inhomogeneity, and (3) a five-pointed star-shaped inhomogeneity are presented to illustrate how the stiffness of the triangular, square or five-pointed star-shaped inhomogeneity can influence the mobility of the screw dislocation.
[DOI: 10.1115/1.2073307]

1 Introduction

The interaction of dislocations with elastic inhomogeneities is an important topic in studying the strengthening and hardening mechanisms of materials. To simplify the analysis, most of the researchers assumed that the inhomoneneity is of circular shape (see, for example, [1,2]), is of elliptical shape (see [3–5] for details), or the inhomogeneity is rigid (or a cavity) with its shear modulus infinite (or zero) (see, for example, [6,7]).

Despite extensive study of inclusions with simple shapes, little effort has been devoted to inclusions of arbitrary shape. For example, Tsukrov and Novak [8] used a computational procedure to calculate the contribution of arbitrary shaped inclusions to the effective moduli of two-dimensional elastic solids. In [9,10], Ru developed a method for evaluating the stress fields of an arbitrary shaped inclusion embedded in full and half planes of isotropic and anisotropic elasticity, respectively. The key limitation of Ru's method is based on the assumption that elastic mismatch between dissimilar materials is negligible. Recently, the Faber series method [11] has been employed to study the problem of an arbitrary shaped inclusion perfectly bonded to the surrounding matrix. In particular, Gao and Noda [12] use the Faber series method to investigate the anti-plane problem of an arbitrary shaped piezoelectric inclusion embedded in an infinite piezoelectric medium.

The focus of the current paper is to investigate, in detail, the interaction problem of a screw dislocation with an arbitrary shaped elastic inhomogeneity. The main feature of this work is that the material properties of the inclusion and the surrounding matrix are different. The solution, in series form, is obtained by means of complex variable methods and Faber series expansion. A rigid inclusion or a cavity can be treated as a special case by letting the shear modulus of the inhomogeneity become infinite or zero, respectively. Once the solution is obtained, the stress fields in the inhomogeneity and in the matrix can be derived. In addition, the image force on the screw dislocation due to its interaction with the elastic inhomogeneity is also derived. In fact, we calculate the image force on the screw dislocation interacting with (1) an equilateral triangular inhomogeneity, (2) a square inhomogeneity, and (3) a five-pointed star-shaped inhomogeneity. It is found that the stiffness of the inhomogeneity has a significant influence on the nature of the image force (either attractive or repulsive) and also on the magnitude of the image force.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received January 27, 2005; final manuscript received May 17, 2005. Review conducted by Z. Sou. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Basic Formulations

As shown in Fig. 1, we consider a domain in R^2, infinite in extent, containing an arbitrary shaped elastic inhomogeneity with elastic properties different from those of the surrounding matrix. The linearly elastic materials occupying the inhomogeneity and matrix are assumed to be homogeneous and isotropic with shear moduli μ_1 and μ_2, respectively. We represent the matrix by the domain S_2 and assume that the inhomogeneity occupies the region S_1. The interface L separating the inhomogeneity and the surrounding matrix is assumed to be perfect (i.e., both the displacement and traction vectors are continuous across L). In addition, a screw dislocation with Burgers vector b is located at the point $z=z_0$ in the matrix.

For the anti-plane problem discussed in this paper, the displacement u_3, the stresses σ_{31}, σ_{32}, and resultant the force R_3 along any arc can be expressed in terms of a single analytic function $f(z)$ as

$$u_3 = \mathrm{Im}\{f(z)\},$$

$$\sigma_{32} + \mathrm{i}\sigma_{31} = \mu f'(z), \tag{1}$$

$$R_3 = -\mu \mathrm{Re}\{f(z)\}.$$

We consider the following conformal mapping function

$$z = m(\zeta) = R\left(\zeta + \sum_{n=1}^{+\infty} m_n \zeta^{-n}\right), \tag{2}$$

which maps the region occupied by the matrix to $|\zeta|>1$ in the mapped ζ plane.

Let the analytic functions in the inhomogeneity and the matrix be denoted by $f_1(z)$ and $f_2(z)$, respectively. For convenience, we write $f_i(z)=f_i(m(\zeta))=f_i(\zeta)$.

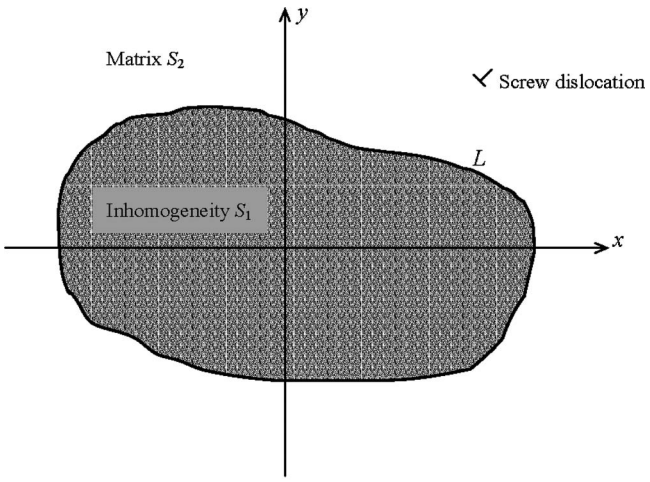


Fig. 1 A screw dislocation interaction with an arbitrarily shaped inhomogeneity

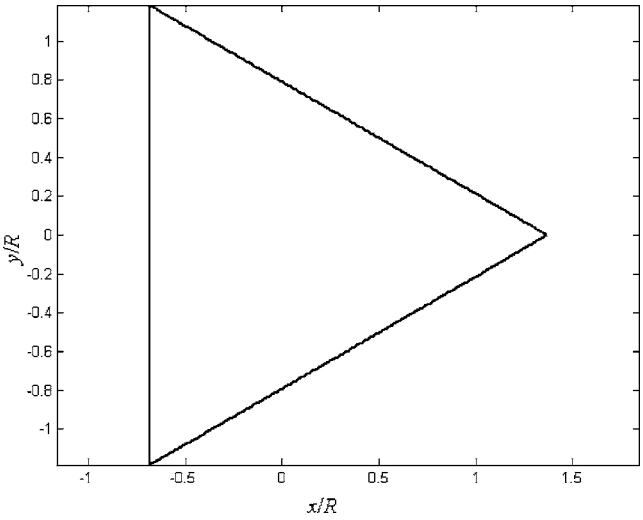


Fig. 2 The equilateral triangle described by Eq. (18)

3 Determination of $f_1(z)$ and $f_2(\zeta)$

The analytic function $f_1(z)$, defined within the inhomogeneity, can be expanded as a Faber series and expressed as follows:

$$f_1(z) = \sum_{n=1}^{+\infty} a_n P_n(z), \tag{3}$$

where $a_n (n=1,2,3,\cdots,+\infty)$ are unknown complex constants to be determined, and $P_k(z)$ is the kth degree Faber polynomial which can be explicitly expressed as

$$P_k(z) = P_k(m(\zeta)) = \zeta^k + \sum_{n=1}^{+\infty} \beta_{k,n}\zeta^{-n} (k = 1,2,3, \ldots, +\infty), \tag{4}$$

where the coefficients $\beta_{k,n}$ are determined by the following recurrence relations [12]

$$\begin{aligned} \beta_{1,n} &= m_n, \\ \beta_{k+1,n} &= m_{k+n} + \beta_{k,n+1} + \sum_{i=1}^{n} m_{n-i}\beta_{k,i}, \\ &\quad - \sum_{i=1}^{k} m_{k-i}\beta_{i,n} \quad (k,n = 1,2,3, \ldots, +\infty). \end{aligned} \tag{5}$$

Hence, $f_1(\zeta)$ can be expressed as follows:

$$f_1(\zeta) = \sum_{n=1}^{+\infty} \left[a_n \zeta^n + \left(\sum_{k=1}^{+\infty} a_k \beta_{k,n} \right) \zeta^{-n} \right]. \tag{6}$$

The continuity condition of displacement and traction across the interface $|\zeta|=1$ can be expressed as

$$f_2^-(\zeta) - \bar{f}_2^+(1/\zeta) = f_1^+(\zeta) - \bar{f}_1^-(1/\zeta),$$

$$f_2^-(\zeta) + \bar{f}_2^+(1/\zeta) = \Gamma[f_1^+(\zeta) + \bar{f}_1^-(1/\zeta)], (|\zeta| = 1), \tag{7}$$

where $\Gamma = \mu_1/\mu_2$.

Inserting (6) into (7) yields the following:

$$\begin{aligned} f_2^-(\zeta) - \bar{f}_2^+(1/\zeta) &= \sum_{n=1}^{+\infty} \left[\left(a_n - \sum_{k=1}^{+\infty} \bar{a}_k \bar{\beta}_{k,n} \right) \zeta^n \right. \\ &\quad \left. - \left(\bar{a}_n - \sum_{k=1}^{+\infty} a_k \beta_{k,n} \right) \zeta^{-n} \right], \\ f_2^-(\zeta) + \bar{f}_2^+(1/\zeta) &= \Gamma \sum_{n=1}^{+\infty} \left[\left(a_n + \sum_{k=1}^{+\infty} \bar{a}_k \bar{\beta}_{k,n} \right) \zeta^n \right. \\ &\quad \left. + \left(\bar{a}_n + \sum_{k=1}^{+\infty} a_k \beta_{k,n} \right) \zeta^{-n} \right], \quad (|\zeta| = 1). \end{aligned} \tag{8}$$

Applying Liouville's theorem, we obtain two expressions for $f_2(\zeta)$

$$\begin{aligned} f_2(\zeta) &= -\sum_{n=1}^{+\infty} \left(\bar{a}_n - \sum_{k=1}^{+\infty} a_k \beta_{k,n} \right) \zeta^{-n} + \frac{b}{2\pi} \ln \frac{\zeta - 1/\bar{\zeta}_0}{\zeta} + \frac{b}{2\pi} \ln(\zeta - \zeta_0), \\ f_2(\zeta) &= \Gamma \sum_{n=1}^{+\infty} \left(\bar{a}_n + \sum_{k=1}^{+\infty} a_k \beta_{k,n} \right) \zeta^{-n} - \frac{b}{2\pi} \ln \frac{\zeta - 1/\bar{\zeta}_0}{\zeta} + \frac{b}{2\pi} \ln(\zeta \\ &\quad - \zeta_0), \quad (|\zeta| > 1), \end{aligned} \tag{9}$$

where $\zeta = \zeta_0 = m^{-1}(z_0) (|\zeta_0| > 1)$ is the location of the screw dislocation.

In view of the fact that the two expressions for $f_2(\zeta)$ must be compatible with each other, then we arrive at the following set of algebraic equations:

$$(\Gamma + 1)a_n + (\Gamma - 1) \sum_{k=1}^{+\infty} \bar{\beta}_{k,n} \bar{a}_k = -\frac{b\zeta_0^{-n}}{n\pi} \quad (n = 1,2,3,\ldots,+\infty). \tag{10}$$

Truncating the above infinite system of linear algebraic equations at a sufficiently large integer N, we obtain

$$\mathbf{A}\mathbf{x} + \mathbf{B}\bar{\mathbf{x}} = \mathbf{f}, \tag{11}$$

where

$$\mathbf{A} = (\Gamma + 1)\mathrm{diag}[1 \quad 1 \quad \cdots \quad 1], \tag{12a}$$

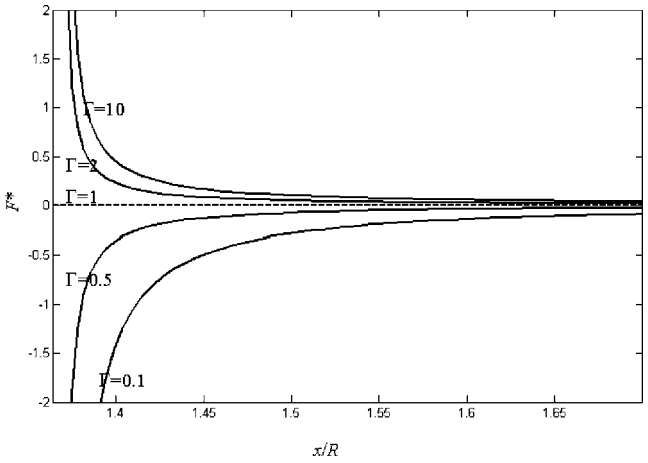


Fig. 3 The normalized image force on a screw dislocation located on the positive *x* axis interacting with the equilateral triangular inhomogeneity

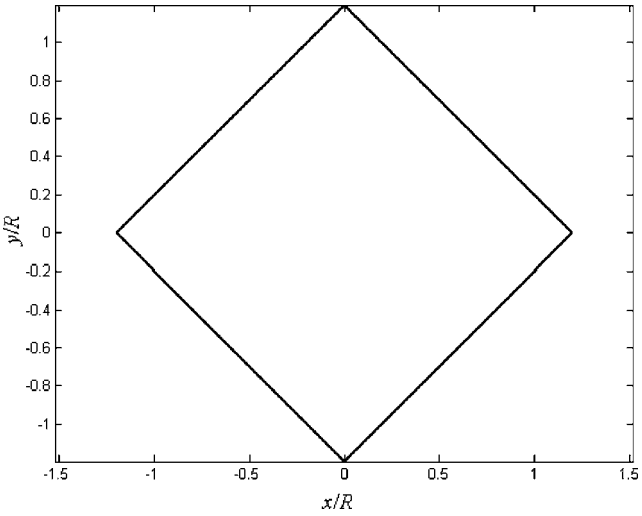


Fig. 4 The square described by Eq. (20)

$$\mathbf{B} = (\Gamma - 1)\begin{bmatrix} \bar{\beta}_{1,1} & \bar{\beta}_{2,1} & \cdots & \bar{\beta}_{N,1} \\ \bar{\beta}_{1,2} & \bar{\beta}_{2,2} & \cdots & \bar{\beta}_{N,2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\beta}_{1,N} & \bar{\beta}_{2,N} & \cdots & \bar{\beta}_{N,N} \end{bmatrix},$$

$$\mathbf{x} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}, \quad \mathbf{f} = -\frac{b}{\pi}\begin{bmatrix} \dfrac{1}{\zeta_0} \\ \dfrac{1}{2\zeta_0^2} \\ \vdots \\ \dfrac{1}{N\zeta_0^N} \end{bmatrix}. \tag{12b}$$

The above set of algebraic equations can be resolved to give

$$\begin{bmatrix} \mathbf{x} \\ \bar{\mathbf{x}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \bar{\mathbf{B}} & \mathbf{A} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{f} \\ \bar{\mathbf{f}} \end{bmatrix}. \tag{13}$$

Remark: The two analytic functions $f_1(z)$ and $f_2(\zeta)$ have now been uniquely determined.

4 Stress Field

Once the analytic function $f_1(z)$ for the inhomogeneity is obtained, the stress field within the arbitrary shaped inhomogeneity can be expressed as

$$\sigma_{32} + \mathrm{i}\sigma_{31} = \mu_1 \sum_{n=1}^{+\infty} a_n P_n'(z). \tag{14}$$

Particularly, the stresses within the inhomogeneity are distributed along the interface L as follows:

$$\sigma_{32} + \mathrm{i}\sigma_{31} = \mu_1 \frac{\displaystyle\sum_{n=1}^{+\infty} n\left[a_n\zeta^{n-1} - \left(\sum_{k=1}^{+\infty} a_k\beta_{k,n}\right)\zeta^{-n-1}\right]}{R\left(1 - \displaystyle\sum_{n=1}^{+\infty} m_n n\zeta^{-n-1}\right)} \quad (|\zeta| = 1). \tag{15}$$

Similarly, once the analytic function $f_2(\zeta)$ for the matrix is obtained, the stresses in the matrix are calculated to be

$$\sigma_{32} + \mathrm{i}\sigma_{31} = \mu_2 \frac{\displaystyle\sum_{n=1}^{+\infty} n\left(\bar{a}_n - \sum_{k=1}^{+\infty} a_k\beta_{k,n}\right)\zeta^{-n-1} + (b/2\pi)([1/\zeta(\bar{\zeta}_0\zeta - 1)] + [1/(\zeta - \zeta_0)])}{R\left(1 - \displaystyle\sum_{n=1}^{+\infty} m_n n\zeta^{-n-1}\right)}, (|\zeta| > 1). \tag{16}$$

5 Image Force on the Screw Dislocation

The image force acting on the screw dislocation due to its interaction with the arbitrary shaped elastic inhomogeneity is derived and takes the following form:

$$F_x - \mathrm{i}F_y = \frac{\mu_2 b\left[\displaystyle\sum_{n=1}^{+\infty} n\left(\bar{a}_n - \sum_{k=1}^{+\infty} a_k\beta_{k,n}\right)\zeta_0^{-n-1} - \{(b/4\pi)[\sum_{n=1}^{+\infty} m_n n(n+1)\zeta_0^{-n-2}/(1 - \sum_{n=1}^{+\infty} m_n n\zeta_0^{-n-1})]\} + \{(b/2\pi)[1/\zeta_0(|\zeta_0|^2 - 1)]\}\right]}{R\left(1 - \displaystyle\sum_{n=1}^{+\infty} m_n n\zeta_0^{-n-1}\right)} (|\zeta_0| > 1), \tag{17}$$

where F_x and F_y are respectively the x and y components of the material force.

6 Examples

6.1 An Equilateral Triangular Inhomogeneity. First, let us consider a screw dislocation interacting with an equilateral triangular inhomogeneity. The specific conformal mapping function is taken to be the following (see, for example, [13,14]):

$$z = m(\zeta) = R\left(\zeta + \frac{1}{3}\zeta^{-2} + \frac{1}{45}\zeta^{-5} + \cdots + \frac{(-1)^k}{(1-3k)}C_{2/3}{}^k\zeta^{1-3k} + \cdots\right) \quad (R>0). \tag{18}$$

In this paper, the above mapping function is truncated at $k=50$ and we take $N=150$ to ensure that the obtained results are sufficiently accurate. The equilateral triangle described by the above mapping function is illustrated in Fig. 2. The calculated coefficients $\beta_{k,n}$ $(k,n=1,2,3,\ldots,10)$ are given by

$$[\beta] = \begin{bmatrix} 0 & 0.3333 & 0 & 0 & 0.0222 & 0 & 0 & 0.0062 & 0 & 0 \\ 0.6667 & 0 & 0 & 0.1556 & 0 & 0 & 0.0272 & 0 & 0 & 0.0098 \\ 0 & 0 & 0.4000 & 0 & 0 & 0.1000 & 0 & 0 & 0.0291 & 0 \\ 0 & 0.3111 & 0 & 0 & 0.2321 & 0 & 0 & 0.0719 & 0 & 0 \\ 0.1111 & 0 & 0 & 0.2901 & 0 & 0 & 0.1497 & 0 & 0 & 0.0567 \\ 0 & 0 & 0.2000 & 0 & 0 & 0.2182 & 0 & 0 & 0.1055 & 0 \\ 0 & 0.0951 & 0 & 0 & 0.2096 & 0 & 0 & 0.1599 & 0 & 0 \\ 0.0494 & 0 & 0 & 0.1437 & 0 & 0 & 0.1827 & 0 & 0 & 0.1199 \\ 0 & 0 & 0.0873 & 0 & 0 & 0.1582 & 0 & 0 & 0.1493 & 0 \\ 0 & 0.0492 & 0 & 0 & 0.1135 & 0 & 0 & 0.1499 & 0 & 0 \end{bmatrix}. \tag{19}$$

Figure 3 demonstrates the normalized image force, $F^* = RF_x/\mu_2 b^2$, on the screw dislocation which is located on the positive x axis in the matrix. It is observed that the screw dislocation will be attracted to the triangular inhomogeneity (i.e. $F_x<0$) when the inhomogeneity is softer than the surrounding matrix ($\Gamma<1$) and the magnitude of the attractive force will be higher when Γ ($\Gamma<1$) becomes smaller and the dislocation is closer to the tip of the inhomogeneity. On the other hand, the triangular inhomogeneity will repel the screw dislocation (i.e. $F_x>0$) when the inhomogeneity is stiffer than the surrounding matrix ($\Gamma>1$) and the magnitude of the repulsive force will be higher when Γ($\Gamma>1$) becomes larger and the dislocation is closer to the tip of the inhomogeneity.

6.2 A Square Inhomogeneity. Next, we address a screw dislocation interacting with a square inhomogeneity. The specific conformal mapping function is taken to be the following (see, for example, [13,14])

$$z = m(\zeta) = R\left[\zeta + \frac{1}{6}\zeta^{-3} + \frac{1}{57}\zeta^{-7} + \cdots + \frac{(-1)^k}{(1-4k)}C_{1/2}{}^k\zeta^{1-4k} + \cdots\right] \quad (R>0). \tag{20}$$

The above mapping function is truncated at $k=50$ and we take $N=200$. The square described by the above mapping function is illustrated in Fig. 4. The calculated coefficients $\beta_{k,n}$ $(k,n=1,2,3,\ldots,10)$ are given by

$$[\beta] = \begin{bmatrix} 0 & 0 & 0.1667 & 0 & 0 & 0 & 0.0179 & 0 & 0 & 0 \\ 0 & 0.3333 & 0 & 0 & 0 & 0.0635 & 0 & 0 & 0 & 0.0173 \\ 0.5000 & 0 & 0 & 0 & 0.1369 & 0 & 0 & 0 & 0.0395 & 0 \\ 0 & 0 & 0 & 0.2381 & 0 & 0 & 0 & 0.0770 & 0 & 0 \\ 0 & 0 & 0.2282 & 0 & 0 & 0 & 0.1193 & 0 & 0 & 0 \\ 0 & 0.1905 & 0 & 0 & 0 & 0.1525 & 0 & 0 & 0 & 0.0728 \\ 0.1250 & 0 & 0 & 0 & 0.1671 & 0 & 0 & 0 & 0.0977 & 0 \\ 0 & 0 & 0 & 0.1539 & 0 & 0 & 0 & 0.1175 & 0 & 0 \\ 0 & 0 & 0.1186 & 0 & 0 & 0 & 0.1256 & 0 & 0 & 0 \\ 0 & 0.0866 & 0 & 0 & 0 & 0.1214 & 0 & 0 & 0 & 0.0943 \end{bmatrix}. \tag{21}$$

Figure 5 demonstrates the normalized image force, $F^* = RF_x/\mu_2 b^2$, on the screw dislocation located on the positive x axis in the matrix. The phenomenon observed for the square inhomogeneity is identical to that for an equilateral triangular inhomogeneity.

6.3 A Five-Pointed Star-Shaped Inhomogeneity. Finally, we investigate a screw dislocation interacting with a five-pointed star-shaped inhomogeneity. The specific conformal mapping function is taken to be the following [13]:

$$z = m(\zeta) = R\left[\zeta + \sum_{p=1}^{+\infty}\left(\sum_{k=0}^{p}(-1)^k C_{4/5}{}^k C_{-2/5}{}^{p-k}\right)\frac{1}{1-5p}\zeta^{1-5p}\right] \quad (R>0). \tag{22}$$

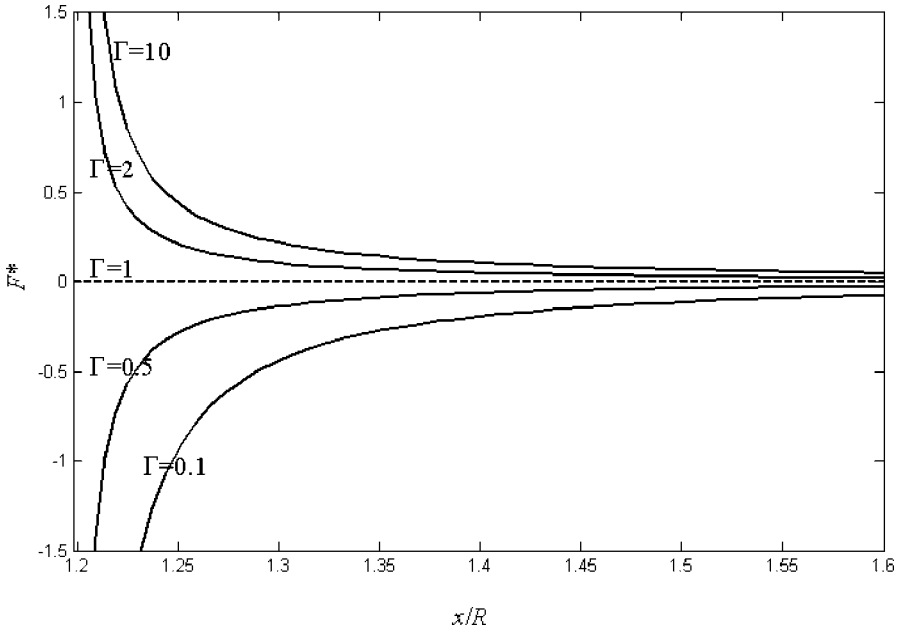


Fig. 5 The normalized image force on a screw dislocation located on the positive *x* axis interacting with the square inhomogeneity

Notice that there is a misprint in Eq. (27) of Ref. [13]. "$2/n$" should read "$-2/n$." The above mapping function is truncated at $p=500$ and we take $N=250$. The five-pointed star described by the mapping function Eq. (22) is illustrated in Fig. 6. The calculated coefficients $\beta_{k,n}$ $(k,n=1,2,3,\ldots,10)$ are given by

$$[\beta]=\begin{bmatrix} 0 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0 & -0.0578 & 0 \\ 0 & 0 & 0.6 & 0 & 0 & 0 & 0 & -0.0256 & 0 & 0 \\ 0 & 0.9 & 0 & 0 & 0 & 0 & 0.0967 & 0 & 0 & 0 \\ 1.2 & 0 & 0 & 0 & 0 & 0.3089 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6111 & 0 & 0 & 0 & 0 & 0.0833 \\ 0 & 0 & 0 & 0.4633 & 0 & 0 & 0 & 0 & 0.3160 & 0 \\ 0 & 0 & 0.2256 & 0 & 0 & 0 & 0 & 0.4947 & 0 & 0 \\ 0 & -0.1022 & 0 & 0 & 0 & 0 & 0.5653 & 0 & 0 & 0 \\ -0.52 & 0 & 0 & 0 & 0 & 0.4740 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1667 & 0 & 0 & 0 & 0 & 0.4130 \end{bmatrix}. \tag{23}$$

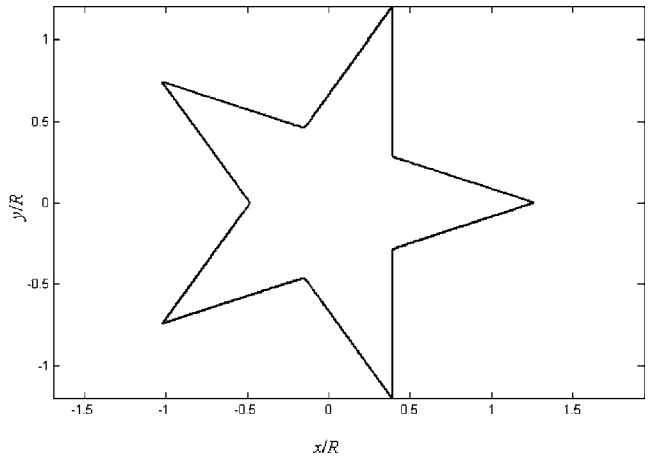


Fig. 6 The five-pointed star described by Eq. (22)

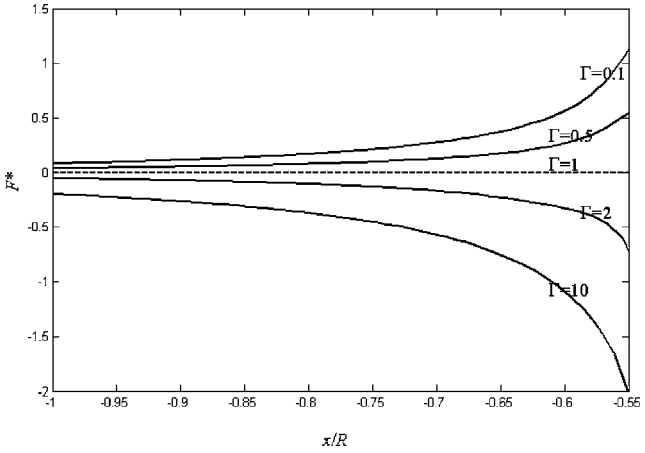


Fig. 7 The normalized image force on a screw dislocation located on the *negative x* axis interacting with the five-pointed star-shaped inhomogeneity

Figure 7 demonstrates the normalized image force, $F^* = RF_x/\mu_2 b^2$, on the screw dislocation located on the *negative x* axis in the matrix. By noting that the dislocation will be attracted to the inhomogeneity when $F_x > 0$, and the dislocation will be repelled from the inhomogeneity when $F_x < 0$, then the phenomenon observed here is similar to that observed for an equilateral triangular or a square inhomogeneity, respectively.

7 Conclusions

In this paper, we present the problem of a screw dislocation interacting with an arbitrary shaped elastic inhomogeneity. The key feature of this work is that the material properties of the inhomogeneity and the surrounding matrix are different. Through the introduction of a conformal mapping function, the region occupied by the matrix can be mapped to the outside of a unit circle in the ζ-plane. In addition, the analytic function defined in the elastic inhomogeneity is expanded into a Faber series. Once the series form solution is obtained, the stress fields due to the screw dislocation can be obtained. Also the image force on the screw dislocation due to its interaction with the elastic inhomogeneity is derived. Several examples of practical importance are presented to demonstrate the feasibility of the obtained solution and to illustrate the influence of the stiffness of the elastic inhomogeneity on the mobility of the screw dislocation. The case where the screw dislocation lies in the arbitrary shaped elastic inhomogeneity can also be similarly addressed. As well, the obtained solution can be easily applied to investigate a matrix crack interacting with an arbitrary shaped elastic inhomogeneity.

Acknowledgment

This work is supported by the Natural Sciences and Engineering Research Council of Canada through grand NSERC No. 249516.

References

[1] Dundurs, J., and Mura, T., 1964, "Interaction Between an Edge Dislocation and a Circular Inclusion," J. Mech. Phys. Solids **12**, pp. 177–189.
[2] Dundurs, J., 1967, "On the Interaction of a Screw Dislocation With Inhomogeneities," Recent Adv. Eng. Sci., **2**, pp. 223–233.
[3] Sendeckyj, G. P., 1970, "Fundamental Aspects of Dislocation Theory". J. A. Summons, Wit R. De, and R. Bulouch, eds., National Bureau of Standards (U.S.), Special Publication 317, Vol. I, p. 57.
[4] Stagni, L., and Lizzio, R., 1983, "Shape Effects in the Interaction Between an Edge Dislocation and an Elliptical Inhomogeneity," Appl. Phys. A: Solids Surf. **A30**, pp. 217–221.
[5] Chen, D. H., 1994, "Interference Between an Elliptical Inclusion and Point Force or Dislocation," J. Jpn. Soci. Mech. Engi., **60**, pp. 2796–2801.
[6] Wu, K. C., 1992, "Interaction of a Dislocation With an Elliptic Hole or Rigid Inclusion in an Anisotropic Material," J. Appl. Phys. **72**, pp. 2156–2163.
[7] Santare, M. H., and Keer, L. M., 1986, "Interaction Between an Edge Dislocation and a Rigid Elliptic Inclusion," ASME J. Appl. Mech. **53**, pp. 382–385.
[8] Tsukrov, I., and Novak, J., 2004, "Effective Elastic Properties of Solids With Two-dimensional Inclusions of Irregular Shapes," Int. J. Solids Struct. **41**, pp. 6905–6924.
[9] Ru, C. Q., 1999, "Analytic Solution for Eshelby's Problem of an Inclusion of Arbitrary Shape in a Plane or Half-plane," ASME J. Appl. Mech. **66**, pp. 315–322.
[10] Ru, C. Q., 2003, "Eshelby Inclusion of Arbitrary Shape in an Anisotropic Plane or Half-plane," Acta Mech. **160**, pp. 219–234.
[11] Curtiss, J. H., 1971, "Faber Polynomials and Faber Series," Am. Math. Monthly **78**, pp. 577–596.
[12] Gao, C. F., and Noda, N., 2004, "Faber Series Method for Two-dimensional Problems of Arbitrarily Shaped Inclusion in Piezoelectric Materials," Acta Mech. **171**, pp. 1–13.
[13] Chen, T., and Chiang, S. C., 1997, "Electroelastic Fields and Effective Moduli of a Medium Containing Cavities or Rigid Inclusions of Arbitrary Shape Under Anti-plane Mechanical and In-plane Electric Fields," Acta Mech. **121**, pp. 79–96.
[14] Thorpe, M. F., 1992, "The Conductivity of a Sheet Containing a few Polygonal Holes or Superconducting Inclusions," Proc. R. Soc. London, Ser. A **437**, pp. 215–227.

Energetics of Epitaxial Island Arrangements on Substrate Mesas

N. D. Machtay
e-mail: noah.machtay@stonybrook.edu

R. V. Kukta

Department of Mechanical Engineering,
State University of New York,
Stony Brook, NY 11794-2300

Self-assembly of strained epitaxial deposits (islands) grown on a substrate is a promising route to fabricate nanostructures of significance for electronic and optoelectronic devices. The challenge is to achieve specific island arrangements that are required for device functionality and high performance. This article investigates growth on a topographically patterned substrate as a means to control the arrangement of islands. By taking free energy to consist of elastic energy and surface energy, minimum energy configurations are calculated for islands on a raised substrate mesa. Configurations of one, two, and three islands at different positions on the mesa are considered to determine their relative energies as a function of mesa size, island size, mismatch strain between the island and substrate materials, surface energy, and elastic moduli. Insight is offered on the mechanisms responsible for certain physical observations such as a transition from the formation of multiple islands to a single island as mesa size is reduced.
[DOI: 10.1115/1.2073327]

1 Introduction

Spontaneous self-assembly of nanostructures in lattice-mismatched epitaxial systems has been a problem of long-standing interest for fabricating advanced electronic and optical devices. The idea is to deposit a material onto a substrate surface and allow inherent physical processes to control formation of the deposited material into nanostructures, such as mounds, lines, or other simple units that are the building-blocks of a particular device. The spontaneous formation of nanostructures in these systems is a well-known phenomenon that occurs during deposition, however the challenge is to produce units of a particular size that are organized in a predefined configuration. For example, a device might require mounds of uniform size arranged in a regular two-dimensional array or in one-dimensional lines or corrals. A method under current investigation for guiding the organization of epitaxial nanostructures is growth on topographically patterned substrates. Relatively large substrate features formed, for example, by lithography and ion-etching act as templates for positioning smaller nanoscale structures.

Experiments have demonstrated that patterned substrates offer some control over the sites where islands (mounds) tend to form. Kitajima et al. [1] deposited Ge on Si(001) substrates patterned with square mesa-like structures. Ge is lattice-mismatched with respect to Si, which under the constraint of epitaxy produces a residual misfit strain in the system. Kitajima et al. [1] found that for relatively wide mesas, islands tend to form at random locations with a notable concentration towards the edges of the mesa. For smaller mesas, islands formed only at the corners and for yet smaller mesas a single island was observed at the center. Jin et al. [2] also observed island formation at the corners of a square mesa in the same system, however with continued growth they found that another island would form at the center of the mesa. Similar observations were reported by Lee et al. [3], however their mesas were circular in shape rather than square. Islands of fairly uniform size and spacing where found to form around the periphery of the mesa and in some cases islands were also found to form in the center. These experiments demonstrate that a relatively large scale topography can guide the organization of smaller scale features, however they offer limited insight into the mechanisms that govern this spontaneous organization. Conventional wisdom is that islands will tend to form at sites of low free energy, as islands near the edge of a mesa tend to be most favorable because the relatively high compliance at the boundary can act to relax misfit strain and thereby reduce the strain energy of the system. However, recent experiments by Zhong et al. [4,5] suggest that energetics alone does not control organization. Kinetic phenomena might have a large influence on organization, as different growth rates often produce different configurations in the same system.

The current article investigates island formation atop mesa structures to determine if observed organized configurations [1–3] are likely energetic or kinetic in origin. Free energies are calculated to assess the relative favorability of different morphologies. For analytical convenience, the system is treated as two-dimensional (plane strain) where a morphology is characterized by the number of islands atop the mesa, island positions, and the shape of individual islands. Island shapes are restricted to circular arcs and characterized by volume and height-to-width aspect ratio. A similar assumption was made by Johnson and Freund [6] to study the energetics of islands on a flat substrate. The specific question is whether islands are most likely to form (a) at the edges of the mesa, (b) in the center, or (c) as a combination with islands at the both the edges and at the center. Kitajima et al. [1] found that each of these configurations can occur depending on the diameter of the mesa and Jin et al. [2] found that either (a) or (c) can occur depending on the amount of material deposited. Free energies of the different configurations with a given volume of deposited material are compared to find the configuration of minimum free energy and determine if it changes as mesa width is decreased or as the amount of material deposited is increased.

The analysis is based on the assumption that the system free energy consists of surface free energy and strain energy, which arises from a lattice mismatch between the island and substrate materials. The island-substrate interface is commensurate and free of defects; therefore its free energy is considered to be negligible. The same assumptions were used in previous analyses of island formation on a flat substrate [6–8].

It is useful to review the case of islands on a flat substrate, as similar mechanisms guide island positioning on patterned substrates. For an island of fixed volume, the total strain energy tends

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received February 2, 2005; final manuscript received May 14, 2005. Review conducted by 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.

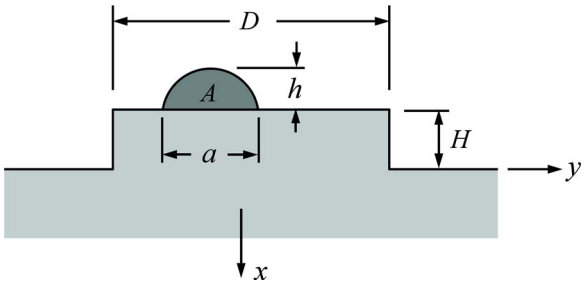


Fig. 1 Geometry of the system under consideration

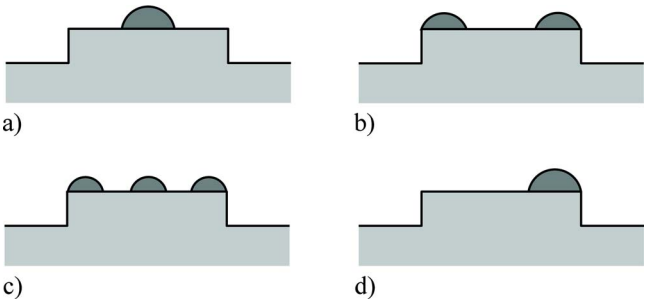


Fig. 2 The four island configurations considered for the energetic analysis. (*a*) A single centered island, (*b*) dual edge mounted islands, (*c*) three islands, one at each edge and one in the center, and (*d*) an asymmetric single edge mounted island.

to be lower for a high aspect ratio island than a low aspect ratio one. This is because a higher aspect ratio is more effective at relaxing mismatch strain in the island and thereby decreases strain energy in the system. At the same time, surface area and therefore surface energy increases as an island increases in aspect ratio. Strain energy and surface energy compete to determine the equilibrium shape of an island. Surface energy is decreased when two islands coalesce to form a single large island. It is found that the total free energy is reduced when two islands on a flat substrate coalesce [8]; however, there is evidence that elastic interactions between islands might impede the kinetic process of coalescence [9–11].

The nature of the competition between strain energy and surface energy implies how substrate patterning might affect island morphology. For example, consider two configurations: (a) two small islands at the edge of a mesa versus (b) one large island at the center of the mesa. A tendency to minimize strain energy would favor configuration (a), as islands at the edges would be more effective in relaxing mismatch strain—the edges are more compliant than the centers. On the other hand, a tendency to minimize surface energy would favor configuration (b), as surface energy is reduced when two islands merge to form a larger island. Consequently, in cases where mismatch strain is large, multiple islands clustered at the mesa edge are expected to be favored, while in cases where surface energy is large a single large island is expected to be favored. This transition is observed in the present investigation. However, more interestingly, a second transition is observed such that as mismatch strain is increased or as surface energy is decreased the favored configuration transitions from a single island (b) to two islands (a) and back to a single island (b). Similar transitions are found to occur with decreasing mesa size and with an increasing amount of island material deposited on the substrate. These observations are consistent with published experiments [1,2].

2 Model

The system consists of islands deposited atop a raised substrate mesa as illustrated in Fig. 1. The islands and substrate are modeled as two-dimensional isotropic linear elastic materials with similar elastic constants. Plane strain conditions are assumed. The topology is periodic in the y direction with one mesa per period. The wavelength is taken to be large enough that the elastic interaction is negligible between adjacent mesas and islands on adjacent mesas. Furthermore, substrate thickness in the x direction is taken to be sufficiently large to approximate the case of an isolated mesa on a half-plane. Island morphologies are modeled as circular arcs characterized by a base dimension a and an aspect ratio h/a. The mesa has a rectangular shape with width D and height H. The substrate and island surfaces are traction-free, and tractions are balanced across the interface between the islands and the substrate. Each island is lattice mismatched with respect to the substrate and the shared interface is coherent such that the island is strained by an amount $\epsilon_{yy}=\epsilon_m$ relative to the substrate, where ϵ_m is the mismatch strain. It is noted that while a two-dimensional geometry misses certain inherently three-dimensional effects, it provides qualitative insight to the mechanisms that control self-assembly in three-dimensional systems.

Free energy consists of strain energy stored in each island and the substrate and free energy of the island and substrate surfaces. Surface energy is taken be isotropic and is given by the same value γ on both the islands and the substrate. Particular configurations are considered (Fig. 2) to determine which has the lowest free energy. These are based on observations in published experiments [1–3]. For each configuration, island positions are fixed but the island aspect ratio is variable so that a given configuration actually represents a series of shapes. The minimum energy configuration is determined amongst all configurations and all island aspect ratios. For a configuration with two or more islands, all islands are assumed to have the same aspect ratio.

The free energy of a given configuration depends on a number of variables. They are surface energy γ, mismatch strain ϵ_m, biaxial film modulus $M=2\mu(1+\nu)/(1-\nu)$, where μ and ν are shear modulus and Poisson's ratio, respectively, the mesa dimensions D and H, and the cross-sectional area A and aspect ratio h/a of the islands. Under the restriction that the islands are circular arcs, A, h, and a are not independent—free energy depends only on two of these variables. Here the aspect ratio h/a is chosen rather than h or a because it characterizes island shape. Area A characterizes the island size and is given by

$$A = a^2 f_A(\alpha) \tag{2.1}$$

where

$$\alpha = h/a \tag{2.2}$$

is the height-to-width aspect ratio and

$$f_A(\alpha) = \left(\frac{1}{8\alpha} + \frac{\alpha}{2}\right)^2 \arctan\frac{4\alpha}{|1-4\alpha^2|} - \frac{|1-4\alpha^2|}{16\alpha} \tag{2.3}$$

is a nondimensional function of aspect ratio. Some configurations have multiple islands and to avoid confusion with the area of a single island, the total cross-sectional area of material in all islands is defined as

$$A_T = nA \tag{2.4}$$

where n is the number of islands. Area A_T corresponds to the amount of island material deposited on the substrate.

The system free energy is given by

$$E = U + \Gamma \tag{2.5}$$

where U is strain energy and Γ is the surface energy measured relative to the case of no islands on the substrate. All energies are measured per unit depth into the plane of Fig. 2. Surface energy is calculated as

$$\Gamma = an\gamma[f_L(\alpha) - 1] \tag{2.6}$$

where

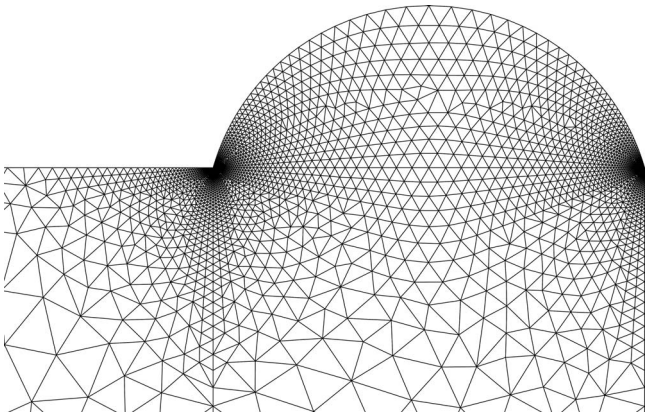


Fig. 3 A typical finite element mesh in the vicinity of an island at the edge of a mesa

$$f_L(\alpha) = \left(\frac{1}{4\alpha} + \alpha\right)\arccos\frac{1-4\alpha^2}{1+4\alpha^2} \tag{2.7}$$

is a dimensionless function of island aspect ratio. On dimensional grounds, the total strain energy of a given configuration is of the form

$$U = MA_T G(\epsilon_m, \nu, \alpha, \bar{D}, \bar{H}) \tag{2.8}$$

where G is a dimensionless function and

$$\bar{D} = \frac{D}{\sqrt{A_T}} \quad \text{and} \quad \bar{H} = \frac{H}{\sqrt{A_T}}. \tag{2.9}$$

Following the observation of Johnson and Freund [6] that the strain energy scales as ϵ_m^2, (2.8) can be rewritten as

$$U = M\epsilon_m^2 A_T F_c(\nu, \alpha, \bar{D}, \bar{H}) \tag{2.10}$$

where F_c is a dimensionless function—subscript "c" acts as a reminder that F_c depends on the particular configuration of Fig. 2. Free energy is nondimensionalized as

$$\bar{E} = \frac{E}{M\epsilon_m^2 A_T} = F_c(\nu, \alpha, \bar{D}, \bar{H}) + \bar{L}_*\left(\frac{f_L(\alpha) - 1}{\pi\sqrt{f_A(\alpha)/n}}\right) \tag{2.11}$$

where

$$L_* = \frac{\pi\gamma}{M\epsilon_m^2} \quad \text{and} \quad \bar{L}_* = \frac{L_*}{\sqrt{A_T}}. \tag{2.12}$$

Length L_* is characteristic of the epitaxial system [12].

Evaluation of (2.11) requires the function $F_c(\nu, \alpha, \bar{D}, \bar{H})$ be calculated, which is defined by (2.10). It is calculated by the finite element method using the commercial software package Abaqus. The epitaxial mismatch between the island and substrate materials was imposed by applying a transformation strain in the islands. Substantial mesh refinement is required in the vicinity of island edges due to large stress concentrations at those regions. Mesh refinement at the corners of the substrate mesa is typically needed only when an island is nearby. A convergence analysis was done to ensure accurate evaluation of the function F_c. Figure 3 illustrates a typical mesh in the vicinity of an island at the edge of a mesa.

3 Results and Discussion

Given a particular epitaxial system (i.e., the film and substrate materials) and given the substrate topology, the system free energy depends on the amount of film material deposited A_T, the aspect ratio of the islands α, and the configuration of the islands (Fig. 2) as described by (2.11). In this section the combination of these variables that minimizes free energy is determined.

3.1 Islands on a Flat Substrate. To illustrate the analysis in a simple context and to review some basic results, the case of an *isolated* island on a *flat* substrate surface is considered first. In this case, the free energy depends only on island volume (per unit length) A_T and island aspect ratio α. It is found that for each volume there is a particular aspect ratio that characterizes the minimum energy island morphology. In general, island aspect ratio increases as volume increases. It is also demonstrated that there is a tendency for two islands to coalesce and form a single larger island.

Equilibrium Shape of an Isolated Island. Parameters $\bar{D}$ and $\bar{H}$ that characterize substrate topology (normalized by island volume) drop out of the expressions (2.10) and (2.11). Furthermore, the subscript can also be removed from F_c as there is no longer a need to distinguish between configurations, and (2.10) and (2.11) can be rewritten as

$$U = M\epsilon_m^2 A_T F(\nu, \alpha) \tag{3.1}$$

and

$$\bar{E} = F(\nu, \alpha) + \bar{L}_*\left(\frac{f_L(\alpha) - 1}{\pi\sqrt{f_A(\alpha)/n}}\right) \tag{3.2}$$

respectively. Equation (3.2) is the nondimensional total energy of n non-interacting (isolated) islands on a flat substrate. A variable number of islands n are retained in this expression for later reference. Here it is taken to be unity. Given the function $F(\nu, \alpha)$, calculated for an isolated island by finite element analysis, and the functions $f_A(\alpha)$ and $f_L(\alpha)$, given by (2.3) and (2.7), curves of free energy versus α are obtained for fixed values of $\bar{L}_*$, as shown by the solid lines in Fig. 4(*a*).

Each of these curves has a single minimum. Noting from (2.12) that $\bar{L}_*$ depends on the inverse of island size $\sqrt{A_T}$, each minimum represents the most energetically favorable configuration for an island of a given size. The locus of minima is traced as a dashed line in Fig. 4(*a*). Figure 4(*b*) plots the minimum energy configurations as α versus $\bar{L}_*$ and shows that aspect ratio increases with decreasing $\bar{L}_*$ or equivalently with increasing island volume. These curves are the same as those obtained previously by Johnson and Freund [6].

This exercise illustrates the tendency of surface energy to flatten the morphology—island aspect ratio decreases as surface energy is increased. At the same time it illustrates that a high mismatch strain tends to produce high aspect ratio islands. As island aspect ratio is increased, the constraint of the substrate on the island is reduced and allows a larger amount of the mismatch strain in the island to be relaxed. A similar phenomenon should occur for islands near the edge of a substrate mesa. For an island at a mesa edge, enhanced compliance at the edge allows a larger amount of strain relaxation than an island at the center of a mesa. Hence mismatch strain should drive islands to form at the edges.

Island Coarsening. The same data show that two islands will tend to combine and form a single larger island. Consider (3.2), which is the total energy of n non-interacting (isolated) islands. Compare, for example, the energy of a single island ($n=1$) versus two smaller islands of the same total volume ($n=2$). If the aspect ratios α are equivalent and $\bar{L}_*$ is nonzero, the energy of the two-island configuration is always larger than that of the single larger island configuration. In this case coalescence is driven by only the second term in (3.2), which implies that it is surface energy and not mismatch strain that drives coalescence. When the islands are allowed to relax their aspect ratios to achieve their minimum energy shapes, it is found that mismatch strain also drives coalescence. This is because the function $F(\nu, \alpha)$ decreases monotonically with increasing aspect ratio and, according to Fig. 4(*b*), island aspect ratio increases with increasing island size (decreas-

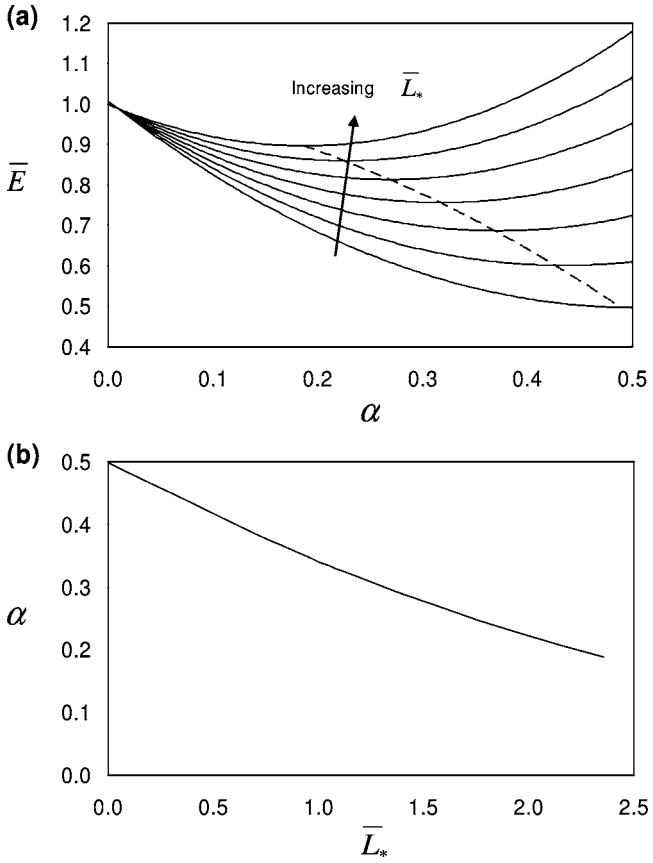


Fig. 4 (*a*) Plot of nondimensional free energy versus aspect ratio for an island on a flat substrate. Solid lines plot curves of fixed $\bar{L}_*=0$, $\bar{L}_*=0.125\pi$, $\bar{L}_*=0.25\pi$, $\bar{L}_*=0.375\pi$, $\bar{L}_*=0.5\pi$, $\bar{L}_*=0.625\pi$, $\bar{L}_*=0.75\pi$. The dashed line traces energy minima for different values of $\bar{L}_*$. (*b*) Aspect ratio of the minimum free energy island plotted versus $\bar{L}_*$ obtained from the dashed line in (*a*). Poisson ratio is $\nu=0.25$.

ing $\bar{L}_*$). Increasing aspect ratio also tends to increase surface energy, which acts against coalescence. The combination of these effects gives rise to the behavior shown by Fig. 5. Figure 5 plots the energy density of two islands relative to a single larger island versus $\bar{L}_*$. The relative energy density is representative of the driving force for coalescence of the two islands. It is found that

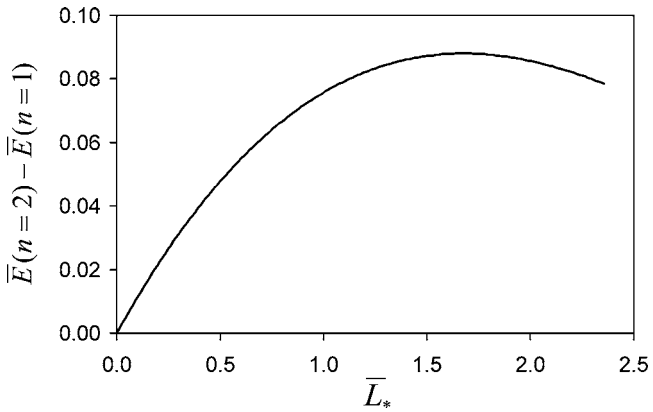


Fig. 5 Dimensionless free energy of two non-interacting islands measured relative to that of a single island of the same total volume plotted versus the normalized characteristic length $\bar{L}_*$. Poisson ratio is $\nu=0.25$.

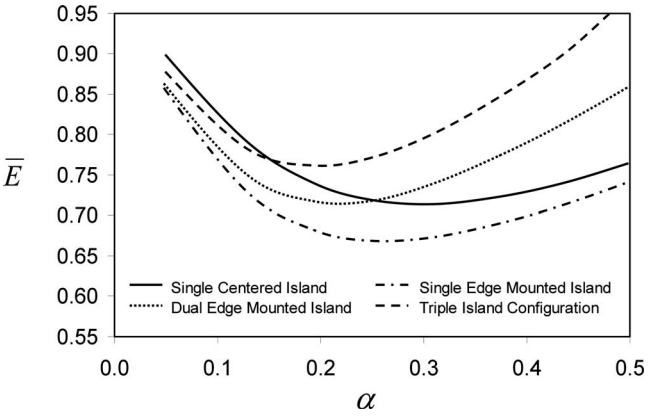


Fig. 6 Plots of dimensionless free energy versus aspect ratio for islands on a substrate mesa in the case $\bar{L}_*=1$, $\bar{D}=23.7$, $\bar{H}=8.4$, and $\nu=0.25$. Curves correspond to the configurations shown in Fig. 2.

the tendency for islands to coalesce grows, reaches a maximum, and then begins to decrease as $\bar{L}_*$ is increased. As $\bar{L}_*\to\infty$ (as mismatch strain vanishes), the minimum energy configuration becomes a flat film of uniform thickness. Since the single and double island configurations are indistinguishable in this case, the driving force is expected to vanish for large values of $\bar{L}_*$. The key feature to note for future reference is that while there is a driving force for coalescence in all cases, it may be large or small depending on the system parameters, which include island size, mismatch strain, and surface energy as characterized by $\bar{L}_*$ (2.12).

3.2 Islands on a Substrate Mesa. For the case of islands on a substrate mesa, a similar analysis is done, however the minimum energy configuration now depends on the substrate patterning ($\bar{D}$ and $\bar{H}$) and the island sizes and positions relative to the pattern (Fig. 2). The analysis is done on each of the four configurations in Fig. 2. For a given total volume of material within the islands and a given epitaxial system (fixed L_*), the aspect ratio of the islands is determined that renders free energy a minimum. Within each test configuration, all island aspect ratios are taken to be equivalent.

Figure 6 is a sample calculation of free energy versus island aspect ratio for each of the four configurations in the case of $\bar{L}_*=1$, $\bar{D}=23.7$, and $\bar{H}=8.4$. Considering, for example, a single island with the shape of a semi-circular cap ($\alpha=0.5$) and diameter a, the ratios of mesa width and height to island size are $D/a\approx 15$ and $H/a\approx 5$.

The plot determines the island aspect ratio and the corresponding free energy of the minimum energy state for each configuration. In this case the ground state is the single asymmetric edge mounted island with an aspect ratio of $\alpha\approx 0.25$. Two metastable states are observed to have roughly the same free energy. They are the single centered island with $\alpha\approx 0.3$ and the dual edge mounted islands with $\alpha\approx 0.22$. The triple island configuration with $\alpha\approx 0.2$ is a metastable state with the highest free energy amongst the configurations considered. One would expect configurations with four or more islands to have yet higher free energies.

Similar plots are obtained for a range of $\bar{L}_*$ to obtain a plot of the energy minima versus $\bar{L}_*$ for each configuration, which is shown in Fig. 7. This plot can be interpreted as a phase diagram that displays the equilibrium and metastable configurations for fixed geometric parameters, $\bar{D}=23.7$ and $\bar{H}=8.4$, and a range of epitaxial systems as characterized by a surface energy γ, a mismatch strain ϵ_m, and a film modulus M per (2.12)—γ increases and ϵ_m and M decrease as $\bar{L}_*$ is increased. Note that according to

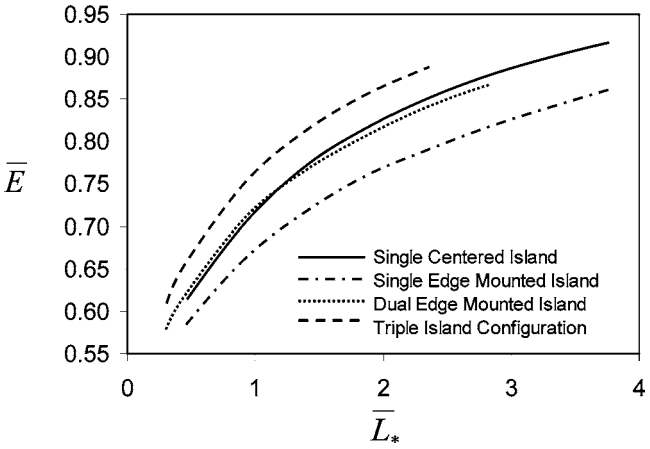


Fig. 7 Minimum free energy state for each of the configurations in Fig. 2 plotted versus the normalized characteristic length. The geometric parameters are $\bar{D}$=23.7 and $\bar{H}$=8.4 and Poisson ratio is ν=0.25. These curves plot the minima of the curves in Fig. 6 and similar plots that are not shown.

the definitions of $\bar{D}$ and $\bar{H}$, (2.9), the ratio of the mesa size to the amount of film material on the mesa is fixed. Therefore this plot does not address how the configuration might evolve during deposition. This issue is addressed later.

Figure 8 plots the corresponding island aspect ratios of the minimum energy configurations in Fig. 7 versus $\bar{L}_*$. The data along with (2.3) and (2.9) can be used to determine the configuration geometries. Island aspect ratios range from about α=0.1 to 0.4. In the case of a single island the ratio of mesa width to island width ranges from about D/a=6 to 13. For double and triple island configurations the values of D/a are respectively double and triple of these values.

Small Islands and the Early Stage of Growth. As the islands are fairly small compared to the mesa size and islands are well separated in the double and triple island configurations, the trend exhibited in Fig. 7 is not expected to change as island size is decreased further. Hence Fig. 7 characterizes the behavior of the initial formation of islands. The figure shows that the single asymmetric edge-mounted island is always the minimum energy configuration for the systems considered, which implies that it is the most favorable configuration for the initial stage of film growth. While this configuration has been observed to form [1], there are many examples of other configurations [1–3]. Whether a single asymmetric island initially forms is a matter of kinetics. For example, high deposition rates can cause many islands to form atop a mesa [11]. Even when the deposition rate is slow, in which case growth tends to follow energetic predictions, it is likely that islands will nucleate at both edges as in Fig. 2(*b*) rather than asymmetrically on one edge because the edges are similar and both are equally favorable sites. Supposing that the double island configuration forms initially, the question of whether it can spontaneously transform into the lower energy configuration of Fig. 2(*d*) remains. This is addressed later. For now it is sufficient to note that similar double island configurations on a different patterned substrate have been found to be stable against such a transformation and may in fact form instead of a lower energy single island configuration [11]. The nature of the metastable configurations in Fig. 7 therefore warrants discussion.

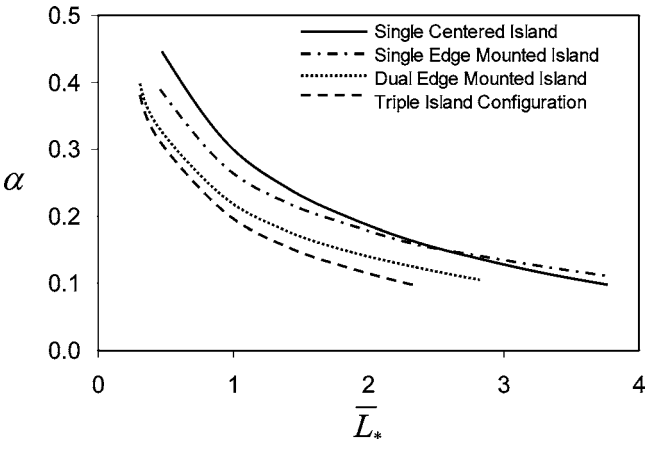


Fig. 8 Island aspect ratio α of the minimum energy configurations in Fig. 7 plotted versus normalized characteristic length

An interesting feature of the metastable configurations in Fig. 7 is that as $\bar{L}_*$ increases, there is a transition from the single centered island to the dual edge mount islands being the lower energy state. This is somewhat unexpected. As discussed previously, large values of ϵ_m are expected drive island formation towards the mesa edges because the edges admit greater strain relaxation in the islands. In this case large ϵ_m (small $\bar{L}_*$) favors the single centered island over islands at the edges. One possible explanation for this is as follows. There is always a driving force for two islands to coalesce and form a single larger island as shown in Fig. 5. A single island will be favored when the tendency for islands to coalesce is stronger than the tendency for islands to form at the edges. The driving force for coalescence decreases when $\bar{L}_*$ becomes sufficiently large. While the tendency for islands to form at the edges of the mesa also decreases as $\bar{L}_*$ increases (mismatch strain decreases), the driving force for islands to form at the edges becomes stronger than that of coalescence. An additional driving force associated with the elastic interaction between islands may also play a role in this transition. However, in the system of Fig. 7 the effect of elastic interactions is likely small. This effect becomes important as islands become larger and more closely spaced as discussed later.

As mismatch strain vanishes, the free energy of the system is determined by surface energy alone, which is represented by the second term on the right-hand side of (2.11). It is then clear that a system with n islands will have a greater free energy than a system with just a single island by a factor of $\sqrt{n}$. This implies that for very large $\bar{L}_*$ a second transition will exist favoring the single centered island for very low levels of mismatch strain and the dual edge islands for higher strain.

Large Islands and Later Stages of Growth. Figure 7 and two similar plots shown in Fig. 9 correspond to various stages of film growth. From Fig. 7 to Fig. 9(*a*) and to Fig. 9(*b*) the amount of film material on the substrate mesa is progressively increased, as demonstrated shortly. In each case, the single asymmetric island has the lowest free energy for all values of $\bar{L}_*$ considered. Figure 9(*a*) exhibits the same transition between the double and single island metastable states as in Figure 7, however the transition point is shifted to a slightly smaller value of $\bar{L}_*$. As the amount of film material is increased further, as represented by Fig. 9(*b*), the transition disappears and the single centered island becomes favored over the dual edge-mounted islands for all values of $\bar{L}_*$ considered.

Elastic interactions are likely responsible for the disappearance of the transition in Fig. 9(*b*). Elastic interaction energy tends to increase the system free energy and gives islands the tendency to repel [13]. On a finite mesa, the repulsion can cause islands to separate only as far the edges permit. Two edge-mounted islands cannot separate further to reduce free energy. The elastic interaction energy increases as the islands increase in size while their separation distance is held fixed. In the case of Fig. 9(*b*), the interaction energy raises the free energy of the dual edge mounted

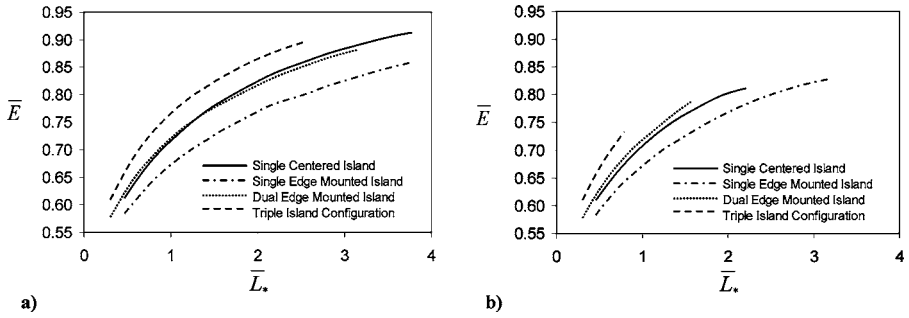


Fig. 9 Minimum free energy of each configuration in Fig. 2 plotted versus the normalized characteristic length for the geometric parameters (*a*) $\bar{D}=13.7$ and $\bar{H}=4.8$ and (*b*) $\bar{D}=4.7$ and $\bar{H}=1.6$. These plots are similar to Fig. 7. The Poisson ratio is $\nu=0.25$ in both.

island configuration above that of the single centered island for all cases considered. This illustrates an important distinction in the behavior of islands on flat versus patterned substrates. In both cases island interactions have a role in the development of morphology. However, a substrate pattern can constrain islands such that they are forced to interact at close range. This enhances the role of interactions over the case of a flat substrate where islands are unconstrained except by adjacent islands.

In Figs. 7 and 9, the ratio of mesa width D to mesa height H is approximately 2.9. To give a physical example of these systems and to illustrate how Figs. 7 and 9 are interpreted in the context of film growth, consider $\gamma=2$ J/m^2, $\epsilon_m=4\%$, and $M=140$ GPa, which is representative of Ge islands on Si. The characteristic length is calculated as $L_*=28$ nm. Take the mesa width to be $D=400$ nm. For the case of Fig. 7 the amount of film material on the mesa is $A=285$ nm^2—this corresponds to a base width of 40 nm for the case of a single island with aspect ratio $\alpha=0.25$. The normalized characteristic length is $\bar{L}_*=1.66$, which lies slightly to the right of the double to single island transition in Fig. 7. Applying the same mesa width $D=400$ nm to Fig. 9(*a*), the amount of film material becomes $A=852$ nm^2. In the case of a single island with $\alpha=0.25$ the base width is 70 nm. Note that the island is larger than in the example of Fig. 7 and hence Fig. 9(*a*) corresponds to a later stage of growth. The normalized characteristic length is $\bar{L}_*=0.96$, which lies slightly to the left of the double to single island transition in Fig. 9(*a*). In this example, the case of the smaller amount of material (Fig. 7 with $\bar{L}_*=1.66$) favored the double island over the single centered island, while the single centered island becomes favored as material is added on the mesa [Fig. 9(*a*) with $\bar{L}_*=0.96$]. Applying the same system to Fig. 9(*b*), the corresponding amount of film material is $A=7243$ nm^2, in which case a single $\alpha=0.25$ island has a base width of 204 nm. The normalized characteristic length is $\bar{L}_*=0.33$. This point lies slightly outside of the systems illustrated in Fig. 9(*b*), but the trend suggests that the single asymmetric island is most favored and the single centered island configuration has the next lowest free energy.

As discussed previously, while a single asymmetric island is the most energetically favorable at the early growth stage, it is likely that kinetic mechanisms will cause islands to form simultaneously at both edges. Symmetric island configurations at the edges of a mesa are often observed in physical experiments [1–3]. Supposing that dual edge-mounted islands form initially, the question arises of whether the system can transform into a state of lower free energy, for example into an asymmetric edge mounted island which is always lower in energy. Furthermore, if the dual island configuration remains after a sufficient amount of deposition, the case of a single centered island becomes a lower free energy configuration and a transformation could occur towards that state. While energetics does not provide a complete picture, it offers insight to why certain configurations arise. Kitajima et al. [1] observed island configurations on square (three-dimensional) mesas. Islands were found on the mesa corners in cases where the deposited volume was small compared to the mesa size. In the case of a larger deposited volume a single centered island was observed. The present investigation suggests that this transition is energetically favorable. By analogy with the two-dimensional analysis, it is expected that islands on the corners will be favored at small deposited volumes and a single centered island will become favored once the volume exceeds a critical value.

Effect of Mesa Size. Figures 7 and 9 also apply when the mesa size is reduced while the amount of deposited material remains fixed. In this case, the mesa size is successively reduced from Fig. 7 to Fig. 9(*a*) and to Fig. 9(*b*). The plots depend only on the ratio $D/\sqrt{A_T}$ and therefore increasing the amount of material deposited for a fixed mesa size is equivalent to reducing the mesa size for a fixed amount of material deposited. The latter corresponds to the investigation of Kitajima et al. [1]. They observed a transition from islands on the corners of a square mesa to a single centered island as the mesa size was reduced. The same insight offered for the case of increasing the amount of material deposited applies here.

Coarsening and Transition Pathways Between Configurations. It is unlikely that only one or two islands will initially form on the mesa. Instead, many islands will nucleate randomly over the surface with the highest probability of islands forming at low energy sites such as the mesa edges. The activation energy for nucleation will be lowest at an edge because an island's free energy is lowest there. Which configuration of Fig. 2 ultimately appears depends on how the initial distribution coarsens. Coarsening takes a system towards a state of lower free energy as material diffuses from sites of high chemical potential to sites of low chemical potential. The chemical potential of an island is defined as the increase in free energy as an infinitesimal amount of material is added to the island. Coarsening occurs as material deposited on the substrate preferentially diffuses towards islands of low chemical potential or as material detaches from certain islands and diffuses to ones of lower chemical potential. Islands of low chemical potential grow at the expense of high chemical potential islands or, in the case that deposition is slow compared to surface diffusion, islands of high chemical potential may disappear as their material is redistributed amongst islands of lower chemical potential.

In the case of a flat substrate a large island typically has a lower chemical potential than a smaller one, which drives coarsening. The chemical potential of an island is largely independent of position, except when islands are interacting at close range the chemical potential might depend on their relative positions. A position dependence is introduced in the case of a patterned substrate. Even when island sizes are equal, those at energetically favored sites will have relatively low chemical potentials and

therefore grow at the expense of other islands.

Suppose, for example, that only three islands nucleate during the initial growth stage and are positioned as shown in Fig. 2(*c*). To understand how the system might coarsen, the chemical potential of each island must be calculated. The chemical potentials can be obtained from the data of Fig. 7, provided it can be applied to a range of island sizes holding all other parameters fixed. In particular, the data must be applicable to smaller island sizes than used to obtain Fig. 7 in order to address the early stages of growth. This is true only if island interactions are negligible and islands are small compared to the mesa dimensions such that a further size reduction has a negligible effect. Island interactions were found to be negligible by calculating the elastic energy of the dual edge mounted configuration from that of the single asymmetric island, which neglects the interaction energy. The result was identical to the corresponding curve in Fig. 7 within numerical error. A similar calculation confirmed that interactions were negligible in the three-island configuration. Island size relative to mesa size affects only the elastic energy of the system, which is represented by the first term F_c in (2.11). The functions F_c obtained for the systems of Figs. 7 and 9(*a*) were found to be equivalent within numerical error. Considering that island size is reduced by a factor of 3 from the system of Fig. 9(*a*) to that of Fig. 7, a further size reduction is expected to change F_c by a negligible amount. For the case of an island on one of the mesa edges the function is well represented by

$$F_{\text{edge}}(\nu=0.25,\alpha,\cdot,\cdot)\approx -7.431\alpha^3+8.142\alpha^2-3.375\alpha+1.000 \tag{3.3}$$

for aspect ratios $\alpha<0.4$. For the case of an island at the center it is well represented by

$$F_{\text{center}}(\nu=0.25,\alpha,\cdot,\cdot)\approx -2.467\alpha^3+4.097\alpha^2-2.473\alpha+1.000 \tag{3.4}$$

for aspect ratios $\alpha<0.45$. These results along with (2.11) can be used to evaluate the chemical potentials during the earliest stages of growth until island interactions become significant or island sizes become large compared to the mesa dimensions. It is noted, however, that island sizes must be large enough for the continuum framework to apply.

The equilibrium aspect ratio of an edge or centered island is readily evaluated by using (3.3) or (3.4) in (2.11) and determining α such that free energy is minimized. For systems having sufficiently large values of $\bar{D}$ and $\bar{H}$, the aspect ratio is a function only of $\bar{L}_*$. The results are in agreement with Fig. 8. Using $\alpha=\alpha(\bar{L}_*)$ and $\bar{L}_*=L_*/\sqrt{A}$ in (2.11), the chemical potential χ of an island is evaluated as

$$\frac{\chi}{M\epsilon_m^2}=\frac{D(\bar{E}A)}{DA}=F_c(\nu,\alpha,\cdot,\cdot)-\frac{\bar{L}_*}{2}\frac{\partial F_c(\nu,\alpha,\cdot,\cdot)}{\partial\alpha}\frac{d\alpha}{d\bar{L}_*}+\frac{\bar{L}_*(f_L(\alpha)-1)}{2\pi\sqrt{f_A(\alpha)}}+\frac{\bar{L}_*^2}{4\pi f_A(\alpha)^{3/2}}\frac{d\alpha}{d\bar{L}_*}\left([f_L(\alpha)-1]\frac{df_A(\alpha)}{d\alpha}-2f_A(\alpha)\frac{df_L(\alpha)}{d\alpha}\right). \tag{3.5}$$

Results for the edge-mounted and centered islands are plotted in Fig. 10.

It is found that an edge-mounted island generally has a lower chemical potential than a centered island of the same size (same value of $\bar{L}_*$). Hence, if the configuration of Fig. 2(*c*) forms during the initial growth stage, the tendency will be for material to diffuse from the centered island to the edge mounted islands. If deposition is fast compared to surface diffusion, the configuration of Fig. 2(*c*) might persist past the transition in Fig. 7 where a single centered island is favored. However, according to Fig. 10,

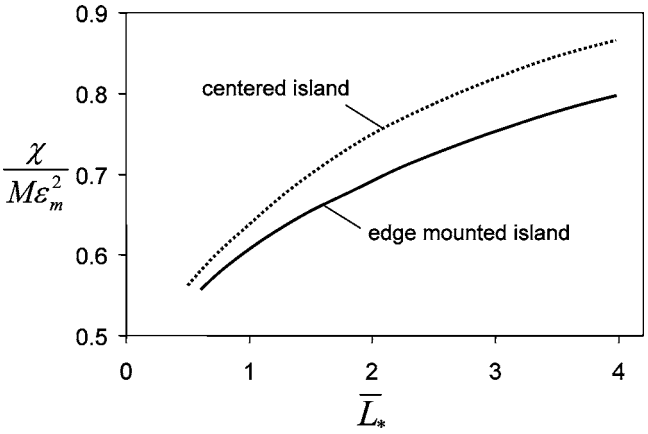


Fig. 10 Plots of chemical potential versus normalized characteristic length for small edge mounted and centered islands. The results apply to systems of sufficiently large $\bar{D}$ and $\bar{H}$. Poisson ratio is $\nu=0.25$.

the edge-mounted island will still have a lower chemical potential and mass transport will occur such that the centered island shrinks and the edge islands grow. While the centered island is favored, there is an activation barrier that must be overcome for it to form from the configuration of Fig. 2(*c*). As shown in Fig. 10, the centered island must be sufficiently larger than an edge island for it to have a lower chemical potential. For example, considering an edge-mounted island with $\bar{L}_*=2$, the centered islands must have an $\bar{L}_*$ less than about 1.4 for it chemical potential to be smaller. This implies that the centered island must be about twice as large in volume. This ratio decreases as $\bar{L}_*$ decreases or equivalently as island sizes increase. For example, if $\bar{L}_*=1$ for an edge-mount island, the centered island must be about 1.6 times larger. This is consistent with the observation from Fig. 7 that a centered island becomes more favorable as the amount of deposition increases. It is unlikely that a centered island will grow to be substantially larger than islands at the edges, as the tendency is for the edge-mounted islands to grow faster. However, if many islands nucleate on the mesa and deposition is fast enough, two islands near the center might merge (grow into each other) to form a larger island with a chemical potential lower than the edge islands. A single centered island might then form. It is noted that the observation of edge islands generally having a lower chemical potential than a centered island of the same size is limited to cases where elastic interactions are small. As the islands grow, elastic interactions will begin to affect their chemical potentials, which might cause the centered island to achieve a lower chemical potential than the edge mounted islands at some stage in the growth process. This has yet to be investigated and is left for future work. It is concluded that a single centered island is likely to form only if the deposition rate is fast compared to mass transport along the surface. Islands must nucleate at multiple sites across the mesa and grow fast enough for islands near the center to remain through to a sufficient amount of growth where island interactions become significant or islands begin to merge.

If a single centered island forms, free energy can be reduced further if it transforms into a single asymmetric edge-mounted island. A transformation cannot occur with a spontaneous flow of mass from the centered island to an incipient island at the edge. This is because a very small edge island typically has a larger chemical potential than a large centered island. A transformation might occur as the entire island translates towards the edge by surface diffusion. This, however, would occur rather slowly if at all. Large islands appear to be immobile, either because they move very slowly or because their motion requires a substantial activation energy [11].

In the case of a dual edge-mounted island configuration, free energy can be reduced if it transforms into an asymmetric edge mounted island. The transformation is unlikely to occur by the translation of one island towards the other because the islands tend to repel each other and are attracted to the edge. It might occur with an incremental transfer of mass such that one island grows while the other shrinks, provided free energy decreases throughout the entire process. Suppose that both islands are initially of the same size A—the free energy is $2E(A)$. If a small amount of mass ΔA is transferred between the islands, the free energy becomes $E(A-\Delta A)+E(A+\Delta A)$. For small values of ΔA the change in energy is

$$\Delta E = \frac{D^2E}{DA^2}(\Delta A)^2 + O(\Delta A)^3 = \frac{D\chi}{DA}(\Delta A)^2 + O(\Delta A)^3$$

$$= -\frac{D\chi}{D\bar{L}_*}\frac{\bar{L}_*}{2A} + O(\Delta A)^3. \tag{3.6}$$

It is observed that an incremental mass transfer will reduce the free energy if $D\chi/D\bar{L}_* > 0$, which is shown to be the case in Fig. 10. This implies that small and well-separated edge-mounted islands are unstable. Mass will spontaneously diffuse from one island to the other and form a single edge mounted island. As Fig. 10 does not account for island interactions, the same conclusion cannot be made for islands that are large and closely spaced. Island interaction might play an important role in stabilizing the dual edge-mounted island configuration as observed in a similar two-island configuration on a different patterned substrate [11]. More investigation is needed to determine if island interactions have a stabilizing effect on the dual edge-mounted island configuration addressed here.

4 Conclusions

The energetics of strained epitaxial islands on a substrate mesa were studied to understand the mechanisms that drive the formation of different island configurations during growth. The goal was to determine if configurations that arise in physical experiments correspond to minimum energy states or if they are kinetically driven. The free energy of the system was taken to consist of the elastic energy induced by a lattice mismatch between the substrate and island materials and a surface energy. Systems with a single island at either the center of the mesa, or asymmetrically mounted at a single edge of the substrate mesa were considered, along with a dual island configuration with one island at each mesa edge, and the three-island configuration with one island at each edge and the third mounted at the center of the mesa top. These represent configurations reported in the experimental literature.

It was found that a single asymmetrically mounted island located at one mesa edge is generally the minimum energy configuration for this type of system. This is attributed to the reduction in strain energy associated with an enhanced compliance near the mesa edge and the general tendency of material to agglomerate into a single island to reduce both strain energy and surface energy. Possible metastable states consisting of two edge mounted islands and a single centered island were found to show a transition in their relative favorability that depends on the volume of material deposited and mesa size. A single island is energetically favored when the deposited volume is large and when the mesa size is relatively small. Two islands are favored for small deposited volumes and when mesa size is relatively large. This is attributed to a competition between the energetic forces driving coalescence and those tending to position islands at the mesa edges. This transition is lost for sufficiently large islands or small mesas where the interaction energy between islands becomes significant and makes multiple island configurations increasingly unfavorable. These findings provide insight into the observation of Kitajima et al. [1] that mesa size plays an important role in determining the configuration that develops.

Island growth was considered by comparing the chemical potential of islands at different sites and sizes on the mesa. Multiple islands are likely to nucleate during the earliest stage of growth, and as the ultimate configuration arises some islands disappear while others grow. Islands of low chemical potential will grow faster than others by preferential deposition or as material diffuses from islands of higher chemical potential. It was found that an island at the edge of the mesa will have the lowest chemical potential unless another island is substantially larger in size. Hence, islands at the edges are likely to grow at the expense of other islands. This is true even in cases where a single centered island is energetically favored over two edge-mounted islands of the same total volume. It is concluded that the formation of a centered island is unlikely except in systems where deposition rate is much faster than mass transfer by surface diffusion. In such cases two or more islands near the center could grow together to form a single large island with a lower chemical potential than smaller islands at the edge. For the systems considered, an island at the center must be about 1.5 to 2 times larger than an island at the edge in order for its chemical potential to be lower. These observations assume that the elastic interaction between islands is negligible. Further investigation is needed to determine the implication of strong interactions that occur when islands are large and spaced closely together. It was found that when island interactions are negligible the dual edge-mounted island configuration is unstable and can spontaneously transform into a single asymmetric edge-mounted island via mass transfer from one island to the other. Recent work by Kukta and Kouris [11] suggests that island interactions might stabilize such a configuration. Further investigation is needed to determine if the dual edge-mounted island configuration is stabilized in the current system.

Acknowledgment

The support of the National Science Foundation under Grant No. CMS-0134123 is gratefully acknowledged.

References

[1] Kitajima, T., Liu, B., and Leone, S. R., 2002, "Two-Dimensional Periodic Alignment of Self-Assembled Ge Islands on Patterned Si(001) Surfaces," Appl. Phys. Lett., **80**(3), pp. 497–499.
[2] Jin, G., Liu, J. L., Thomas, S. G., Luo, Y. H., Wang, K. L., and Nguyen, B. Y., 1995, "Controlled Arrangement of Self-Organized Ge Islands on Patterned Si (001) Substrates," Appl. Phys. Lett., **75**(18), pp. 2752–2754.
[3] Lee, H. M., Yang, T. H., Luo, G., and Chang, E. Y., 2003, "Flower-Like Distributed Self-Organized Ge Dots on Patterned Si(001) Substrates," Jpn. J. Appl. Phys., Part 1, **42**, pp. L718–L720.
[4] Zhong, Z., Halilovic, A., Muhlberger, M., Schaffler, F., and Bauer, G., 2003, "Positioning of Self-Assembled Ge Islands on Stripe-Patterned Si(001) Substrates," J. Appl. Phys., **93**, pp. 6258–6264.
[5] Zhong, Z., Halilovic, A., Muhlberger, M., Schaffler, F., and Bauer, G., 2003, "Ge Island Formation on Stripe-Patterned Si(001) Substrates," Appl. Phys. Lett., **82**, pp. 445–447.
[6] Johnson, H. T., and Freund, L. B., 1997, "Mechanics of Coherent and Dislocated Island Morphologies in Strained Epitaxial Material Systems," J. Appl. Phys., **81**, pp. 6081–6090.
[7] Spencer, B. J., and Tersoff, J., 1996, "Equilibrium Shapes of Small Strain Islands," *Evolution of Epitaxial Structure and Morpholgy*, Mater. Res. Soc. Proc., Pittsburgh, PA, Vol. 399, pp. 283–288.
[8] Kukta, R. V., and Freund, L. B., 1997, "Minimum Energy Configuration of Epitaxial Material Clusters on a Lattice-Mismatched Substrate," J. Mech. Phys. Solids, **11/12**, pp. 1835–1860.
[9] Floro, J. A., Sinclair, M. B., Chason, E., Freund, L. B., Twesten, R. D., Hwang, R. Q., and Lucadamo, G. A., 2000, "Novel Sige Island Coarsening Kinetics: Ostwald Ripening and Elastic Interactions," Phys. Rev. Lett., **84**, pp. 701–704.
[10] Liu, P., Zhang, Y., and Lu, C., 2002, "Coarsening Kinetics of Heteroepitaxial Islands in Nucleationless Stranski-Krastanov Growth," Phys. Rev. B, **68**, pp. 035402–1—035402–8.
[11] Kukta, R. V., and Kouris, D., 2005, "On the Mechanisms of Epitaxial Island Alignment on Patterned Substrates," J. Appl. Phys., **97**, pp. 033527–1—033527–1.
[12] Srolovitz, D. J., 1989, "On the Stability of Surfaces of Stress Solids," Acta Metall., **37**, pp. 621–625.
[13] Rickman, J. M., and Srolovitz, D. J., 1993, "Defect Interactions on Solid-Surfaces," Surf. Sci., **284**(1-2), pp. 211–221.

The Mode III Interface Crack in Piezo-Electro-Magneto-Elastic Dissimilar Bimaterials

R. Li
Post-Doctoral Fellow

G. A. Kardomateas
Professor
Fellow ASME

School of Aerospace Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332-0150

The mode III interface crack problem is investigated for dissimilar piezo-electro-magneto-elastic bimaterial media, taking the electro-magnetic field inside the crack into account. Closed form solutions are derived for impermeable and permeable cracks. The conventional singularity of $r^{-1/2}$ is found for the fields at the distance r ahead of the interface crack tip. Expressions for extended crack tip stress fields and crack opening displacements (ECODs) are derived explicitly, and so are some fracture parameters, such as extended stress intensity factors (ESIFs) and energy release rate (G) for dissimilar bimaterials. An approach called the "energy method," finding the stationary point of the saddle surface of energy release rate with respect to the electro-magnetic field inside the crack, is proposed. By this method, the components of the induced electro-magnetic field inside the crack are determined, and the results are in exact agreement with those in the literature if the two constituents of the bimaterial media are identical. The effects from mechanical and electro-magnetic property mismatches, such as differences in the stiffness, electric permittivity and magnetic permeability, between the two constituents of the bimedia on the mode III interface crack propagation are illustrated by numerical simulations. The results show that the applied electric and magnetic loading usually retard the growth of the interface crack and the directions of the combined mechanical, electric, and magnetic loading have a significant influence on the mode III interface crack propagation. [DOI: 10.1115/1.2073328]

1 Introduction

One class of contemporary materials, widely used in engineering in devices (in sensor, transducer, actuator components, etc.), are the piezoelectric and piezomagnetic composite materials. Due to their exceptional functions, such as flat frequency response [1–4] and transformation of energy from one form to the other (mechanical, electric, and magnetic energy, or thermal energy) [5,6], this type of composite exhibiting piezoelectric and piezomagnetic properties has found increasing applications in microwave electronic, optoelectronic, and electronic instruments. Like in conventional composites, defects or flaws may usually be introduced during the manufacturing process or during service by impact loading. These defects would often deteriorate the performance of the devices being made of piezo-electro-magneto-elastic media.

Recently, more and more attention has been directed towards the problems of cracks in the electro-magneto-elastic solids [7–12]. Dissimilar bimaterials or layered composites are often incorporated into a variety of components, such as smart structure sensors, actuators, and broadband magnetic probes. Having been recognized as one of the common failure modes of general dissimilar bimaterial media, the interface cracks could also be developed in the piezo-electro-magneto-elastic structures and thus affect the features of the electro-magneto-elastic apparatus. Though these interface cracks may severely diminish the performance of this type of structure, one may see that little attention has been given to the study of the magneto-electric coupling effects on the interface crack propagation behavior in piezo-electro-magneto-elastic bimaterials.

The magneto-electric coupling effect of piezoelectric and piezomagnetic fields usually has a significant influence on the behavior of piezo-electro-magneto-elastic bimaterials or layered structures [3,4,13]. This coupling among the magnetic, electric, and elastic fields is also expected to have an influence on the propagation behavior of interface crack/delaminations when piezoelectric, piezomagnetic, and magneto-electric effects, or any two of these effects, are present simultaneously [1]. These coupling effects usually complicate this interface crack problem. In order to get insight into the interface crack problems of dissimilar piezo-electro-magneto-elastic bimaterial composites, the mode III interface crack is investigated in this paper by using Stroh's formulism [14] and the complex variable method. Two types of mode III interface cracks are analyzed. One is called permeable interface crack for which the magneto-electric field inside the interface crack is considered. The other type is called impermeable.

This paper is organized as follows: (1) In Sec. 2 is a summary of some basic equations for piezo-electro-magneto-elasticity in Strohs formalism. (2) A compact form solution to the interface crack is formulated in Sec. 3. The expressions for the ECOD, ESIF, and the energy release rate are derived in closed form. The "energy method" is also proposed in this section and used to obtain the solution to the magneto-elastic field inside the interface crack. One may interestedly find that this method could be extended to more complicated problems in piezo-electro-magnetic elastic solids. (3) The numerical results in Sec. 4 show the influence of the property mismatches between the two constituents on the interface crack propagation. An interesting result one may find is that the applied external electric-magnetic field may slow the growth of mode III interface cracks in piezo-electro-magneto-elastic bimaterial solids. Since all the formulas in this paper are obtained in explicit expressions, and are thus easily trackable, this study may serve as a benchmark for further investigations in piezo-electro-magneto-elastic media

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received February 19, 2005; final manuscript received June 1, 2005. Review conducted by 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.

2 Basic Equations

In a fixed Cartesian coordinate system (x_1, x_2, x_3), the generalized Hookes law for an elastic material with both piezoelectric and piezomagnetic fields is of the following form [1]:

$$\sigma_{ij} = c_{ijkl}u_{k,l} + e_{lij}\varphi^E_{,l} + \varrho_{lij}\varphi^H_{,l},$$

$$D_i = e_{ikl}u_{k,l} - \varepsilon_{il}\varphi^E_{,l} - \alpha_{il}\varphi^H_{,l}, \tag{1}$$

$$B_i = \varrho_{ikl}u_{k,l} - \alpha_{li}\varphi^E_{,l} - \mu_{il}\varphi^H_{,l}$$

where i,j,k,l range in {1, 2, 3} and the repeated indices imply summation, the comma stands for differentiation with respect to corresponding coordinate variables; σ_{ij} is the elastic stress, u_k is the elastic displacement, and c_{ijkl} is the elastic moduli tensor; D_i is the electric displacement, φ^E is the electrostatic potential, and ε_{il} is the electric permittivity; B_i is the magnetic induction (magnetic fluxes), φ^H is the magnetic scalar potential, and μ_{il} the magnetic permeability; e_{ikl}, ϱ_{ikl}, and α_{li} are the piezoelectric, piezomagnetic, and magnetoelectric coefficients, respectively. For the material constants, the following relationships hold:

$$c_{ijkl} = c_{jikl} = c_{ijlk} = c_{klij}, \quad e_{ikl} = e_{ilk}, \quad \varrho_{ikl} = \varrho_{ilk},$$
$$\alpha_{il} = \alpha_{li}, \quad \varepsilon_{il} = \varepsilon_{li}, \quad \mu_{il} = \mu_{li} \tag{2}$$

The equilibrium equations in the absence of body forces read

$$\sigma_{ij,j} = 0, \quad D_{i,i} = 0, \quad B_{i,i} = 0 \tag{3}$$

For two-dimensional antiplane deformation of a transversely isotropic solid, we have

$$u_1 = 0, \quad u_2 = 0, \quad u_3 = u_3(x_1, x_2),$$
$$\varphi^E = \varphi^E(x_1, x_2), \quad \varphi^H = \varphi^H(x_1, x_2) \tag{4}$$

One may define the extended displacement as

$$\mathbf{u} = [u_3, \varphi^E, \varphi^H]^T \tag{5}$$

For a plane system, a nontrivial solution to Eq. (3) may then take the following form:

$$\mathbf{u} = A\, f(z_\alpha) + \bar{A}\bar{f}(\bar{z}_\alpha), \quad \psi = B\, f(z_\alpha) + \bar{B}\bar{f}(\bar{z}_\alpha), \quad z_\alpha = x_1 + p_\alpha x_2 \tag{6}$$

where ψ is the stress function vector and $f(z_\alpha)$ are functions to be determined by boundary conditions.

If one defines the extended stress fields as

$$\mathbf{t} = [\sigma_{32}, D_2, B_2]^T, \quad \mathbf{s} = [\sigma_{31}, D_1, B_1]^T \tag{7}$$

then these stresses can be written in terms of the stress functions as

$$\mathbf{s} = \left(-\frac{\partial \psi_i}{\partial x_2}\right)^T, \quad \mathbf{t} = \left(\frac{\partial \psi_i}{\partial x_1}\right)^T = \psi' \tag{8}$$

Substituting Eq. (6) back into the equation (3), one readily obtains

$$A = I = \mathrm{diag}[1,1,1], \quad B = i\begin{pmatrix} c_{44} & e_{15} & \varrho_{15} \\ e_{15} & -\varepsilon_{11} & -\alpha_{11} \\ \varrho_{15} & -\alpha_{11} & -\mu_{11} \end{pmatrix}, \quad p_\alpha = i \tag{9}$$

where $i^2 = -1$.

If we define a matrix M as

$$M = iAB^{-1}, \tag{10}$$

then

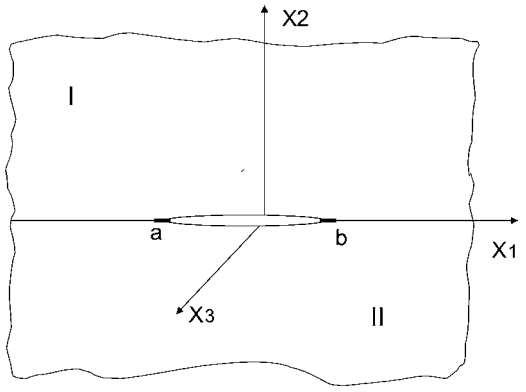


Fig. 1 An interface crack between dissimilar anisotropic bimedia

$$M = \begin{pmatrix} \varepsilon_{11}\mu_{11} - \alpha_{11}^2 & e_{15}\mu_{11} - \alpha_{11}\varrho_{15} & \varepsilon_{11}\varrho_{15} - \alpha_{11}e_{15} \\ e_{15}\mu_{11} - \alpha_{11}\varrho_{15} & -\varrho_{15}^2 - c_{44}\mu_{11} & c_{44}\alpha_{11} + e_{15}\varrho_{15} \\ \varepsilon_{11}\varrho_{15} - \alpha_{11}e_{15} & c_{44}\alpha_{11} + e_{15}\varrho_{15} & -e_{15}^2 - c_{44}\alpha_{11}^2 \end{pmatrix} \Bigg/ \Delta \tag{11}$$

where

$$\Delta = c_{44}\varepsilon_{11}\mu_{11} + e_{15}^2\mu_{11} + \varepsilon_{11}\varrho_{15}^2 - 2\alpha_{11}e_{15}\varrho_{15} - c_{44}\alpha_{11}^2 \tag{12}$$

The matrix M is real and symmetric.

3 A Solution to Mode III Interface Crack

Let the medium "*I*" occupy the upper half-space (donated by *L*) and medium "*II*" be in the lower half-space (donated by *R*) (Fig. 1). Then from Eqs. (6) and (9), one has the following expression for this bimedia:

$$\mathbf{u}^I = \phi_I(z) + \bar{\phi}_I(\bar{z}), \quad \psi^I = B_I\phi_I(z) + \bar{B}_I\bar{\phi}_I(\bar{z}) \tag{13}$$

where, $\mathbf{u}^I, \psi^I$ are displacement and stress functions for $z \in L$, and

$$\mathbf{u}^{II} = \phi_{II}(z) + \bar{\phi}_{II}(\bar{z}), \quad \psi^{II} = B_{II}\phi_{II}(z) + \bar{B}_{II}\bar{\phi}_{II}(\bar{z}) \tag{14}$$

where $\mathbf{u}^{II}, \psi^{II}$ are displacement and stress functions for $z \in R$.

For convenience, the symbols "*I*" and "*II*," denoting the quantities for medium "*L*" and "*R*," respectively, may be put as subscripts or subscripts.

Let the interface crack be located in the region $a < x_1 < b, -\infty < x_3 < \infty$ of the plane $x_2 = 0$. The $p_0^\infty = [\sigma_{i2}^\infty]^T = [\sigma_{32}^\infty, D_2^\infty, B_2^\infty]^T$ is applied at infinity (Fig. 1). Inside the crack often is air or vacuum, and the electro-magnetic field usually is considered constant under uniform remote applied load [11,12, etc.]. These unknown components for the electro-magnetic field are denoted as D_α^0, B_α^0, E_α^0, and H_α^0, which, respectively, observe the relationships

$$E_\alpha^0 = \frac{D_\alpha^0}{\varepsilon_0}, \quad H_\alpha^0 = \frac{B_\alpha^0}{\varepsilon_0}, \quad \alpha = \{1,2\} \tag{15}$$

Employing the superposition principle leads the original boundary value problem to an equivalent problem with the loading

$$p_0 = [\sigma_{32}^{\infty}, \Delta D_2^0, \Delta B_2^0]^T \tag{16}$$

being applied on the two surfaces of the interface crack, where, in Eq. (16),

$$\Delta D_2^0 = D_2^{\infty} - D_2^0, \quad \Delta B_2^0 = B_2^{\infty} - B_2^0. \tag{17}$$

The displacement continuity along the bonded interface gives

$$\phi_{I+}(x_1) + \bar{\phi}_{I-}(x_1) = \phi_{II-}(x_1) + \bar{\phi}_{II+}(x_1)$$

or

$$\phi_{I+}(x_1) - \bar{\phi}_{II+}(x_1) = \phi_{II-}(x_1) - \bar{\phi}_{I-}(x_1) \tag{18}$$

A function can be defined being analytical on the whole plane, except the cut along the interface crack, as follows:

$$\Phi(z) = \begin{cases} \phi_I(z) - \bar{\phi}_{II}(z), & z \in L \\ \phi_{II}(z) - \bar{\phi}_I(z), & z \in R \end{cases} \tag{19}$$

Then, this function automatically satisfies the condition (18). Here, a convention

$$\phi(z) = \phi_{\pm}(x_1), \quad x_2 \to 0^{\pm} \tag{20}$$

is employed and will be used in the following sections.

Differentiation of Eq. (19) with respect to z yields

$$\Phi'(z) = \begin{cases} \phi_I'(z) - \bar{\phi}_{II}'(z), & z \in L \\ \phi_{II}'(z) - \bar{\phi}_I'(z), & z \in R \end{cases} \tag{21}$$

The stress continuity on the bonded interface leads to

$$B_I\phi_{I+}'(x_1) + \bar{B}_I\bar{\phi}_{I-}'(x_1) = B_{II}\phi_{II-}'(x_1) + \bar{B}_{II}\bar{\phi}_{II+}'(x_1) \tag{22}$$

Similarly, we can define a function, which automatically satisfies the condition (22) and is analytical on the whole plane except the cut along the interface crack, as

$$\omega(z) = \begin{cases} B_I\phi_I'(z) - \bar{B}_{II}\bar{\phi}_{II}'(z) & z \in L \\ B_{II}\phi_{II}'(z) - \bar{B}_I\bar{\phi}_I'(z), & z \in R \end{cases} \tag{23}$$

From Eqs. (21) and (23), we obtain

$$B_I\phi_I'(z) = N[\mathrm{i}\ \Phi'(z) + \bar{M}_{II}\omega(z)], \tag{24a}$$

$$\bar{B}_{II}\bar{\phi}_{II}'(z) = B_I\phi_I'(z) - \omega(z), \quad z \in L; \tag{24b}$$

and

$$B_{II}\phi_{II}'(z) = \bar{N}[\mathrm{i}\ \Phi'(z) + \bar{M}_I\omega(z)], \tag{25a}$$

$$\bar{B}_I\bar{\phi}_I'(z) = B_{II}\phi_{II}'(z) - \omega(z) \quad z \in R \tag{25b}$$

In the above equations, the following matrix was used:

$$N^{-1} = M_I + \bar{M}_{II} = M_I + M_{II} \tag{26}$$

Since M_I and M_{II} are real symmetric, so is N. Furthermore, define

$$H = M_I + M_{II} \tag{27}$$

Therefore, the boundary traction conditions along the interface crack surface give

$$B_I\phi_{I+}'(x_1) + B_{II}\phi_{II-}'(x_1) - \omega_-(x_1) = -p_0(x_1), \tag{28a}$$

$$B_{II}\phi_{II-}'(x_1) + B_I\phi_{I+}'(x_1) - \omega_+(x_1) = -p_0(x_1) \tag{28b}$$

Subtraction of Eq. (28b) from (28a)) yields

$$\omega_+(x_1) - \omega_-(x_1) = 0 \tag{29}$$

which implies that the $\omega(z)$ is continuous on the whole interface.

By the analytical continuation principle [15], the function $\omega(z)$ is analytical on the whole plane. But according to Liouville's theorem [15], this $\omega(z)$ must be a constant function in the whole domain. However, the condition that this function should vanish at infinity means this constant function must be identically zero in the whole plane, i.e.,

$$\omega(\mathbf{z}) = 0, \quad \text{for all } \mathbf{z} \tag{30}$$

Either Eq. (28a) or (28b) leads to a general Hilbert equation in matrix notation:

$$\Phi_+'(x_1) + \Phi_-'(x_1) = i\ Hp_0(x_1), \quad a < x_1 < b \tag{31}$$

The homogenous equation corresponding to the general Hilbert equation (31) can be written as

$$X_+(x_1) + X_-(x_1) = 0, \quad a < x_1 < b \tag{32}$$

where

$$X(z) = \frac{1}{\sqrt{(z-a)(z-b)}}\mathrm{diag}[1,1,1] \tag{33}$$

A solution which vanishes at infinite could be [16]

$$\Phi'(z) = \frac{1}{2\pi i}X(z)\int_{\overline{ab}} \frac{[X(x_1)]_+^{-1}H[\mathrm{i}\ p_0(x_1)]dx_1}{x_1 - z} \tag{34}$$

Specifically, for constant applied loading, one has (see the Appendix)

$$\Phi'(z) = \mathrm{diag}\left(1 - \frac{z-(a+b)/2}{\sqrt{(z-a)(z-b)}}\right)\frac{H}{2}(ip_0) \tag{35}$$

Integrating Eq. (35) results in

$$\Phi(z) = \mathrm{diag}[z - \sqrt{(z-a)(z-b)}]\frac{H}{2}(ip_0) \tag{36}$$

where the constant contributing rigid body motion is omitted.

Next, let us consider some fracture characterizing parameters such as the crack tip field intensity factors, extended displacement discontinuities near the crack tips, and the energy release rate.

From the equations (24a) and (25b), the extended traction along the interface could be expressed as

$$\mathbf{t}(x_1) = N\ i\Phi_+(x_1) + \bar{N}\ i\Phi_-(x_1) = H^{-1}[i\ \Phi_+(x_1) + i\ \Phi_-(x_1)] \tag{37}$$

We shall show that the right-hand side of Eq. (37) is real, as required.

Substituting the stress function (34) to traction expression (37) leads to

$$\mathbf{t}(x_1) = -p + [X_+(x_1) + X_-(x_1)]\left(x_1 - \frac{a+b}{2}\right)p_0/2 \tag{38}$$

When Eq. (32) is employed, the traction along the interface reads:

$$\mathbf{t}(x_1) = \begin{cases} -p_0 + [(x_1-a)(x_1-b)]^{-1/2}\mathrm{diag}\left(x_1 - \dfrac{a+b}{2}\right)p_0, & x_1 < a \text{ and } b < x_1 \\ -p_0, & a < x_1 < b \end{cases} \tag{39}$$

which is a real vector as expected.

Then the extended tractions at a distance "r" ahead of the crack tip such as "b" (Fig. 1) can be expressed in the form of

$$\mathbf{t}(r) = (2\pi\, r)^{-1/2}\sqrt{\frac{\pi(b-a)}{2}}p_0 = (2\pi\, r)^{-1/2}[K_\sigma, K_D, K_B]^T \tag{40}$$

where K's are real numbers and defined as

$$K_\sigma = \sqrt{\frac{\pi(b-a)}{2}}\sigma_{32}^{\infty}, \quad K_D = \sqrt{\frac{\pi(b-a)}{2}}\Delta D_2^0,$$

$$K_B = \sqrt{\frac{\pi(b-a)}{2}}\Delta B_2^0 \tag{41}$$

These K's may be called the extended stress intensity factors (ESIFs). If we let

$$K = [K_\sigma, K_D, K_B]^T \tag{42}$$

then the expression (42) becomes

$$K = \sqrt{\frac{\pi(b-a)}{2}}p_0 \tag{43}$$

with p_0 defined in (16).

One may also extend the conventional crack open displacement (COD) to piezo-magneto-electric materials. From Eqs. (13), (14), and (19), this extended crack open displacement (ECOD) may readily be evaluated by

$$\Delta\mathbf{u}(x_1) = \mathbf{u}_+^I(x_1) - \mathbf{u}_-^{II}(x_1) = \Phi_+(x_1) - \Phi_-(x_1) = \begin{cases} [(x_1-a)(b-x_1)]^{1/2}H\, p_0, & a < x_1 < b \\ 0, \quad x_1 < a \quad \text{or} \quad b < x_1; \end{cases} \tag{44}$$

Then the ECOD at a small distance "r" behind the tip of the interface crack may read

$$\Delta\mathbf{u}(r) = \sqrt{\frac{r}{2\pi}}H(2K) \tag{45}$$

also an expected real vector.

Now, the energy release rate, G, can be computed and it reads

$$G = \frac{1}{2}\lim_{\delta\to 0^+}\frac{1}{\delta}\int_0^{\delta}\mathbf{t}(r)^T\Delta\mathbf{u}(\delta - r)\mathrm{d}r = \frac{1}{2}K^THK \tag{46}$$

One may realize that all the expressions derived so far include the unknown components D_2^0 and B_2^0 of the electro-magnetic field inside the crack. There are two approaches to determine these unknowns. The first method views the crack as a degenerated hole, using the continuous conditions on the hole surface to determine the electric-magnetic fields. This method may work well for monolithic material as shown in literature such as in [12], because of the convenient affine mapping function. But it is hard to extend this method derived for monolithic materials to the bimaterial media because of the differences in the material properties between the two constituents of a bimaterial system. To offset this difficulty, here, another approach, called the "energy method," is proposed. As one may know, when a remote load starts to apply, an electric-magnetic field begins to build up inside the interface crack. This newly built field causes reactions to fields induced by the applied loading inside the whole material system. One may see that the energy release rate, G, is a saddle surface with respect to variables, D_2^0 and B_2^0, the electric-magnetic field inside the interface crack. This means for each value of G, there exist many corresponding sets of D_2^0 and B_2^0 except at the stationary point, in which only a unique D_2^0 and B_2^0 corresponds to a unique value of G.

Therefore, the value of D_2^0 and B_2^0 at the stationary point could be the final competition result of the above-mentioned interaction. Then one would have following equations:

$$\frac{\partial G}{\partial D_2^0} = H_{12}\sigma_{32}^{\infty} + H_{22}\Delta D_2^0 + H_{32}\Delta B_2^0 = 0, \tag{47a}$$

$$\frac{\partial G}{\partial B_2^0} = H_{13}\sigma_{32}^{\infty} + H_{23}\Delta D_2^0 + H_{33}\Delta B_2^0 = 0 \tag{47b}$$

which leads to

$$D_2^0 = D_2^{\infty} - \Delta D_2^0 = D_2^{\infty} - (H_{23}H_{31} - H_{21}H_{33})/(H_{22}H_{33} - H_{23}^2)\sigma_{32}^{\infty},$$

$$B_2^0 = B_2^{\infty} - \Delta B_2^0 = B_2^{\infty} - (H_{21}H_{32} - H_{22}H_{31})/(H_{22}H_{33} - H_{23}^2)\sigma_{32}^{\infty} \tag{48}$$

where H_{ij} $(i,j=1,2,3)$ are elements of the bimaterial matrix H defined in (27). The result of (48) can be shown the same as those in literature if the two media of this current bimaterial are identical. This agreement justifies the above energy method. From the result in (48), one may see that the electric-magnetic field inside the interface crack is a function of the bimaterial property under given remote applied loading.

One may also observe from (47a) that if one wants $D_2^0 \to 0$ without magnetic field, then H_{22} needs to approach a very big value. This is called electrically impermeable. The parameter λ_e, introduced by McMeeking [17], is used to characterize the electric permeability. A similar parameter, λ_m, could be defined from the observation made on (47b), in which if $B_2^0 \to 0$ without electric field, then H_{33} has to approach a very big value, a phenomenon called magnetically impermeable. These two parameters λ_e and λ_m have the relationship of $\lambda_m/\lambda_e = (\varepsilon_0/\mu_0)/(H_{22}/H_{33})$.

Therefore, for an impermeable interface crack, $D_2^0 = B_2^0 = 0$ and the ESIF can be expressed as

$$K = \sqrt{\frac{\pi(b-a)}{2}}[\sigma_{32}^{\infty}, D_2^{\infty}, B_2^{\infty}]^T \tag{49}$$

The energy release rate for this interface crack reads

$$G_{imp} = \frac{1}{4}K^THK = \frac{\pi(b-a)}{8}[H_{11}(\sigma_{32}^{\infty})^2 + H_{22}(D_2^{\infty})^2 + H_{33}(B_2^{\infty})^2 + 2H_{21}\sigma_{32}^{\infty}D_2^{\infty} + 2H_{31}\sigma_{32}^{\infty}B_2^{\infty} + 2H_{32}D_2^{\infty}B_2^{\infty}] \tag{50}$$

For a permeable interface crack, $\lambda_e = \lambda_m = 0$, the D_2^0 and B_2^0 are given by Eq. (48), and the ESIF can be expressed as

$$K = \sqrt{\frac{\pi(b-a)}{2}}[\sigma_{32}^{\infty}, \Delta D_2^0, \Delta B_2^0]^T \tag{51}$$

The corresponding energy release rate reads

$$G_{perm} = \frac{\pi(b-a)}{8}\frac{\det(H)}{\det(\hat{H})}(\sigma_{32}^{\infty})^2 \tag{52}$$

where the matrix $\hat{H}$, a principal submatrix of H, is

$$\hat{H} = \begin{pmatrix} H_{22} & H_{23} \\ H_{23} & H_{33} \end{pmatrix} \tag{53}$$

and det( ) is the determinant of a square matrix.

One interesting observation from Eq. (52) is that, though the energy release rate, G, is independent of the applied electric-magnetic load, it is affected by electric-magnetic properties of the two constituents of the bimaterial media.

4 Numerical Results

In this section, the influence of the material property mismatches between the two constituents of the bimedia and the effects from magneto-electric coupling on the interface crack

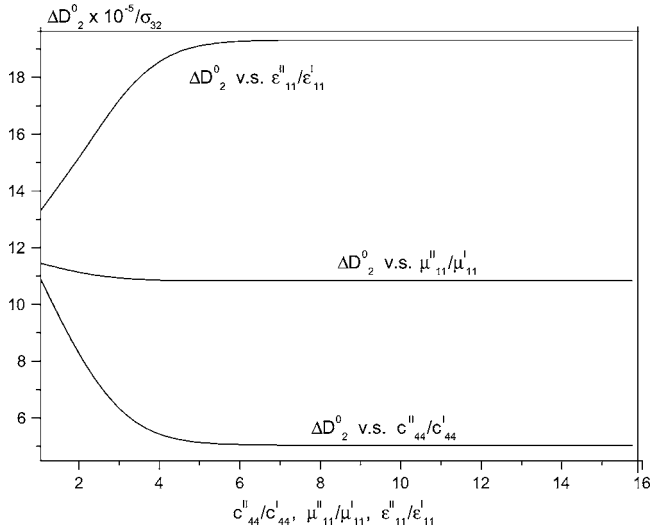


Fig. 2 ΔD_2^0 versus the bimaterial properties

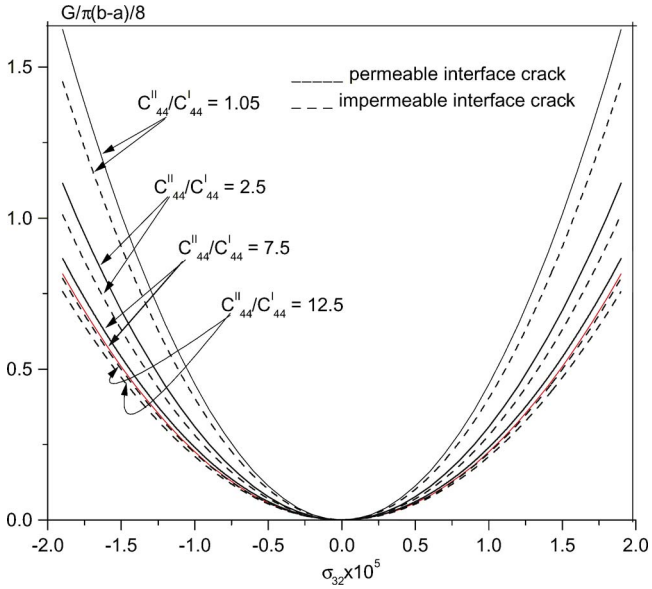


Fig. 4 Energy release rate, G, versus the stiffness ratio c_{44}^{II}/c_{44}^{I}, for an interface crack under pure mechanical loading

growth behavior will be demonstrated by some numerical results. The basic data for the material properties selected here are similar to those in [6]. These constants read as c_{44}^{I}=43.7 GPa; e_{15}^{I} =8.12 C/m^2; $\varepsilon_{11}^{I}=7.86\times10^{-9}$ C/Vm; α_{11}^{I}=0.0; ϱ_{15}^{I} =165.0 N/Am; $\mu_{11}^{I}=180.5\times10^{-6}$ Ns^2/C^2, for the upper medium (medium "*I*"); and c_{44}^{II}=44.6 GPa; e_{15}^{II}=3.48 C/m^2; ε_{11}^{II}=3.42 $\times10^{-9}$ C/Vm; α_{11}^{II}=0.0; ϱ_{15}^{II}=385.0 N/Am; μ_{11}^{II}=414.5 $\times10^{-6}$ Ns^2/C^2, for the lower medium (medium "II").

Figures 2 and 3 present the influences of the bimaterial property mismatches c_{44}^{II}/c_{44}^{I}, $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$, and $\mu_{11}^{II}/\mu_{11}^{I}$ on ΔD_2^0 and ΔB_2^0, which relate to the magneto-electric field, D_2^0 and B_2^0, inside the interface crack by Eq. (17). One may easily see from Fig. 2 that the electric displacement ΔD_2^0 decreases as the degree of anisotropy of these two constituents of the bimedia, defined by c_{44}^{II}/c_{44}^{I}, increases, while it increases as the electric permittivity ratio, $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$, increases. But it practically does not change as the magnetic permeability ratio, $\mu_{11}^{II}/\mu_{11}^{I}$, increases. The magnetic induction field ΔB_2^0 decreases as c^{II}/c^{I} and $\mu_{11}^{II}/\mu_{11}^{I}$ increase, while it increase as $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$ increases, as shown in Fig. 3. One can also see the ΔD_2^0 and ΔB_2^0 do not vary with the increase of c_{44}^{II}/c_{44}^{I}, $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$, and $\mu_{11}^{II}/\mu_{11}^{I}$ after they reach some value.

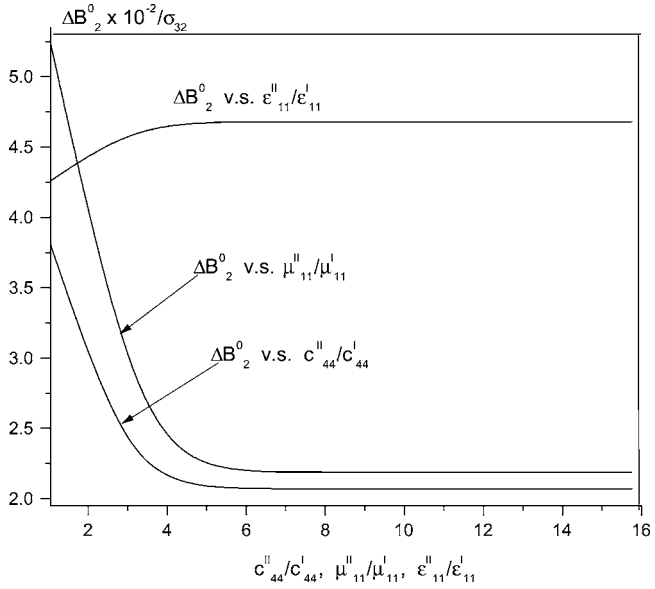


Fig. 3 ΔB_2^0 versus the bimaterial properties

Figure 4 shows the influence on the energy release rate, G, of the mismatch of the degree of anisotropy for the bimaterial media under pure mechanical tension. The G decreases as the c_{44}^{II}/c_{44}^{I} increase, both for permeable and impermeable interface cracks. It can also be seen that when c_{44}^{II}/c_{44}^{I} reaches some value (around 12.5 for this bimedia), the G almost does not vary with the increase in the mismatch on c_{44} between the two constituents of the bimaterial media. Another interesting result observed from this figure is that for a given σ_{23}, G_{perm} is larger than G_{imp}. This observation shows that the electric-magnetic field inside the interface crack may have an interaction with the stress field inside the bimaterial system, thus it has an influence on the propagation behavior of the interface crack. This observation may also suggest that the design of a piezo-electro-magneto-elastic bimaterial system based on a permeable assumption is more conservative than based on impermeable assumption.

Figures 5–7 show the influences on G from the directions of applied D_2^{∞} and B_2^{∞}, respectively. Figure 5 shows the results for loading D_2^{∞} and σ_{32}, Fig. 6 for B_2^{∞} and σ_{32}, while Fig. 7 is for combined loading D_2^{∞}, B_2^{∞}, and σ_{32}. In these figures, a negative G can be observed under certain mechanically applied load, namely σ_{32}^{rtd}, for a given D_2^{∞} and/or B_2^{∞}. These negative values on G may suggest that the applied electric-magnetic loading would retard the propagation of an interface crack in piezo-electro-magnetic bimaterials, a result which was also found in Ref. [12] for cracks in monolithic piezo-electro-magnetic materials. The σ_{32}^{rtd} varies as the direction of D_2^{∞} or B_2^{∞} revises. One can also observe that there exists a direction in which the combined loading applied would make σ_{32}^{rtd} reach its maximum and minimum value.

Figures 8 and 9 more clearly show the retarding effects, respectively, of $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$ and $\mu_{11}^{II}/\mu_{11}^{I}$ on the energy release rate G under pure loading D_2^{∞} or B_2^{∞}. In these two pictures, the value of G is always negative since the applied mechanical loading σ_{32} is zero. The G increases as $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$ or $\mu_{11}^{II}/\mu_{11}^{I}$ increases, a result consist with the observation in Figs. 11 and 12.

Plotted in Figs. 10–12 are, correspondingly, the influences of c_{44}^{II}/c_{44}^{I}, $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$, and $\mu_{11}^{II}/\mu_{11}^{I}$ on the energy release rate, G, under combined electric, magnetic, and mechanical loading for an impermeable interface crack. The plotting in solid line is for the

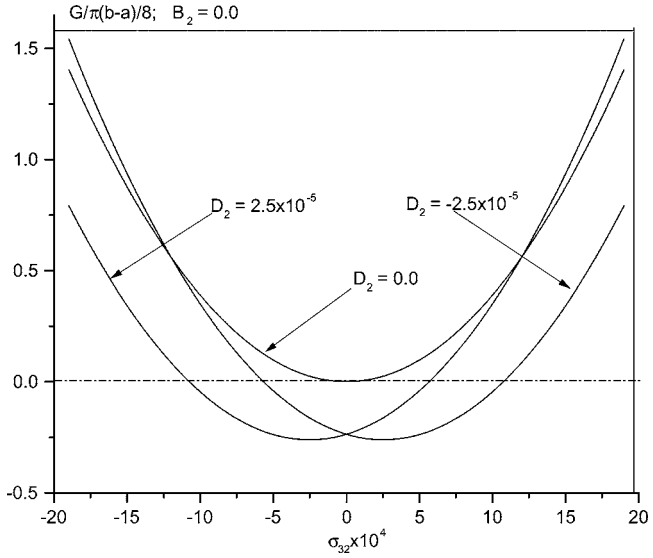


Fig. 5 The effect of the direction of the applied D_2 on the energy release rate, G, for an impermeable interface crack

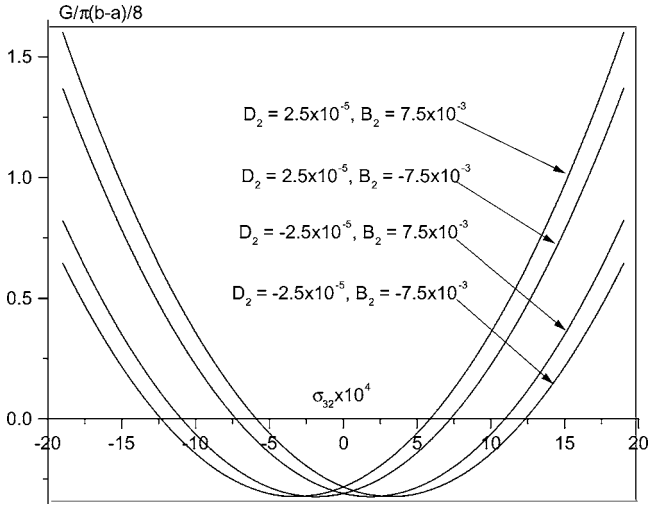


Fig. 7 The effect of the directions of the combined applied B_2 and D_2 on the energy release rate, G, for an impermeable interface crack

positive direction in D_2^∞ and B_2^∞, the dashed line is for revised direction in D_2^∞ and B_2^∞. The G decreases as the c_{44}^{II}/c_{44}^{I} increases and keeps almost unchanged when c_{44}^{II}/c_{44}^{I} reaches a certain value for both applied loading directions, as shown in Fig. 10. On the contrary, the G increases as the $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$ and $\mu_{11}^{II}/\mu_{11}^{I}$ increase, respectively. The observations in these figures may suggest that a reasonable selection in the mechanical and electric-magnetic properties for the two constituents of a bimaterial media may lower the energy release rate, making this bimedia much safer with regard to propagation of cracks.

Finally, it should be mentioned that the important contribution of our paper is the novel procedure, which has been developed to solve for the electric-magnetic fields inside an interface crack in a general bimaterial. The exact agreement of the results from this method with the results from the mapping method for the special case of homogeneous material (i.e., no bimaterial) in the literature, which, again, is the only case solved in the literature, provides validity for our "energy method" approach. It should be noted at this point that the contribution of the electric-magnetic fields inside a crack is very important for the devices being made of piezo-magneto-electro-elastic materials, since these fields may interfere with the desirable signals of electric-magnetic fields, like in broadband detecting devices. The results of our study could offer tentative guidelines for the damage-tolerant design of the devices.

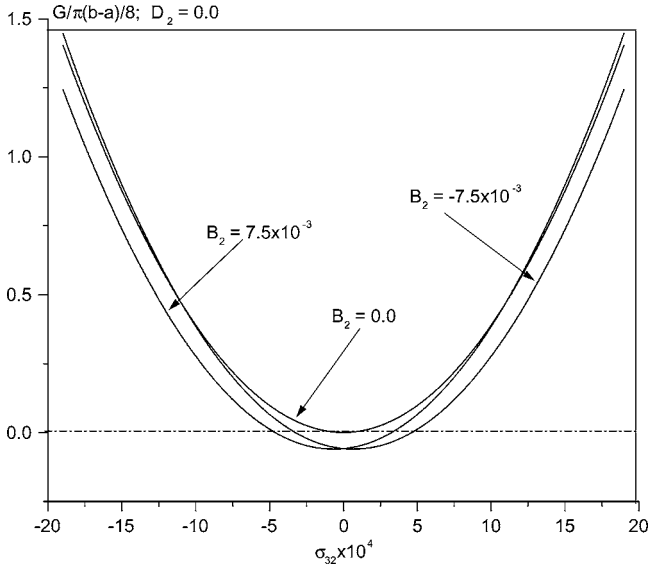


Fig. 6 The effect of the direction of the applied B_2 on the energy release rate, G, for an impermeable interface crack

5 Conclusions

In the present paper, the mode III interface crack in dissimilar piezo-magneto-electro-elastic bimaterial media is investigated in Stroh's formulism. In this study, the electric-magnetic field inside the interface crack is also considered and an "energy method" is proposed for obtaining the solution to this electric-magnetic field. Two types of interface cracks, namely permeable and impermeable cracks, are addressed. All the solutions are derived in closed form. The following conclusions can be reached from the results in this study:

1. The "energy method" is a very effective way to derive a solution to the electric-magnetic field inside a crack, thus

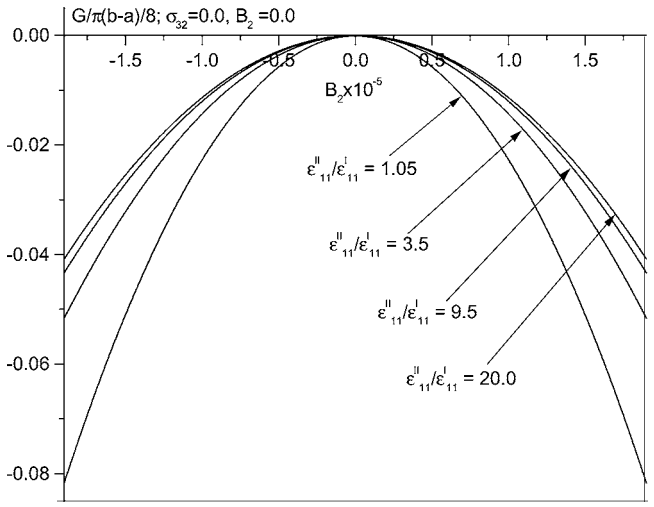


Fig. 8 Energy release rate, G, versus the electric permittivity ratio, $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$, for an impermeable interface crack under pure D_0^∞ loading

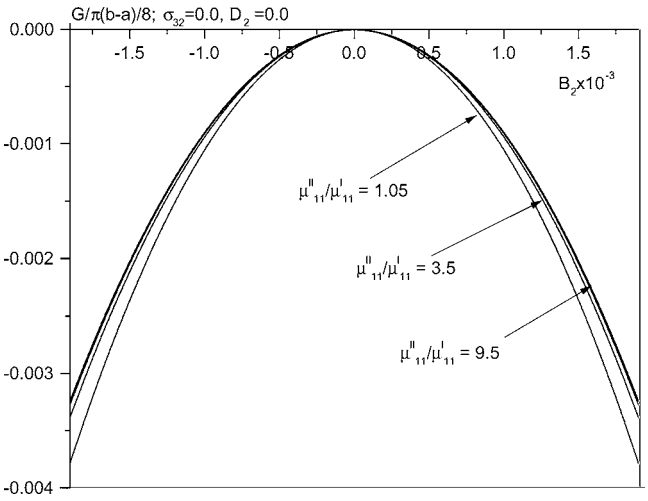


Fig. 9 Energy release rate, *G*, versus the magnetic permeability ratio, $\mu_{11}^{II}/\mu_{11}^{I}$, for an impermeable interface crack under pure B_0^{∞} loading

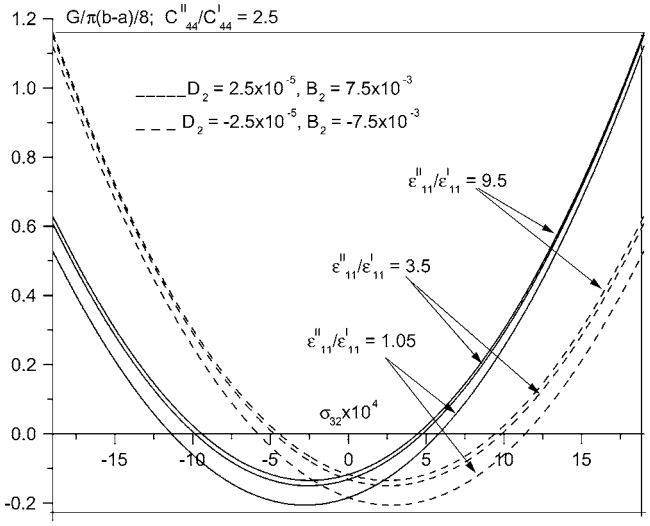


Fig. 11 Energy release rate, *G*, versus the electric permittivity ratio, $\varepsilon_{11}^{II}/\varepsilon_{11}^{I}$, for an impermeable interface crack under combined applied loading

solving the whole interface crack problem when the electric-magnetic field inside a crack is taken into account.

2. The mismatches of c_{44}, ε_{11}, and μ_{11} between the two constituents of a bimaterial media have strong effects on the potential propagation of a mode III interface crack. There exists an optimal selection on c_{44}, ε_{11}, and μ_{11} that would minimize the energy release rate for this mode III interface crack.
3. The directions of the applied loading D_2^{∞} and B_2^{∞} also have an effect on the possible growth of the interface crack in a piezo-electro-magneto-elastic bimaterial media.
4. The applied electric and/or magnetic loading D_2^{∞} and B_2^{∞} usually retard the propagation of the mode III interface crack.

Acknowledgment

The financial support of the Office of Naval Research, Grant No. N00014-90-J-1995, and the interest and encouragement of the Grant Monitor, Dr. Y.D.S. Rajapakse, are both gratefully acknowledged.

Appendix: Contour Integral for $\Phi(z)'$

The method used here can be viewed as the generalization of the technique in [16, 110, and 70] which is for a single equation. Let γ be a contour which includes the arc $\overline{ab}$, and let this contour shrink into the arc $\overline{ab}$. Then for $q(x_1)$ constant

$$\int_{\gamma} \frac{[X(\xi)]^{-1}N^{-1}}{\xi - z}d\xi = \int_{\overline{ab}} \frac{[X_+(x_1)]^{-1}N^{-1}}{x_1 - z}dt + \int_{\overline{ba}} \frac{[X_-(x_1)]^{-1}\bar{N}}{x_1 - z}dx_1$$

$$= \int_{\overline{ab}} \frac{[X_+(x_1)]^{-1}N^{-1}}{x_1 - z}dx_1 - \int_{\overline{ab}} \frac{[X_-(x_1)]^{-1}N^{-1}}{x_1 - z}dx_1 \tag{A1}$$

From Eq. (32), one could have

$$X_-(x_1) = -\bar{N}^{-1}NX_+(x_1), \quad a < x_1 < b \tag{A2}$$

Substituting Eq. (A2) into (A1) leads

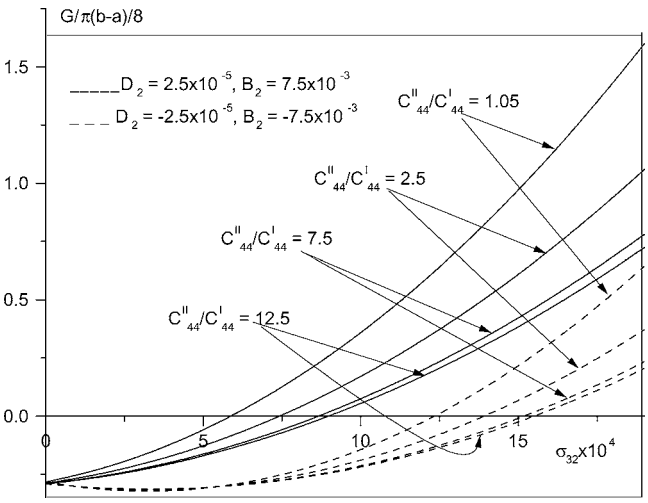


Fig. 10 Energy release rate, *G*, versus the stiffness ratio, c_{44}^{II}/c_{44}^{I}, for an impermeable interface crack under combined applied loading

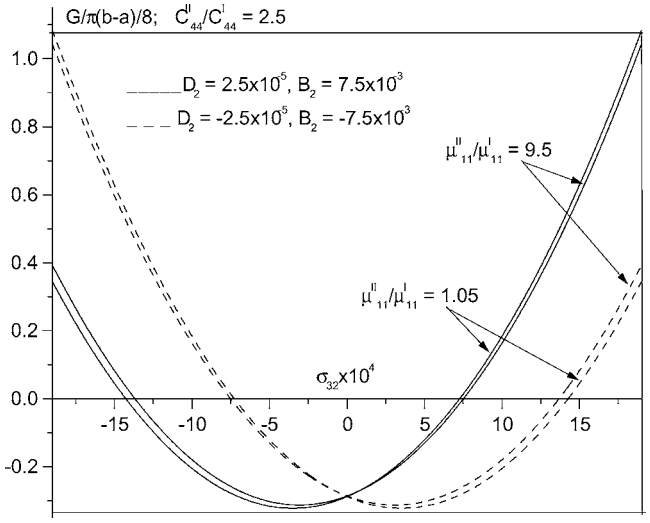


Fig. 12 Energy release rate, *G*, versus the magnetic permeability ratio, $\mu_{11}^{II}/\mu_{11}^{I}$, for an impermeable interface crack under combined applied loading

$$\int_{\gamma} \frac{[X(\xi)]^{-1}N^{-1}}{\xi - z} d\xi = \int_{\overline{ab}} \frac{[X_+(x_1)]^{-1}N^{-1}[I + \bar{N}N^{-1}]}{x_1 - z} dx_1 \quad \text{(A3)}$$

Then,

$$\int_{\overline{ab}} \frac{[X_+(x_1)]^{-1}N^{-1}}{x_1 - z} dx_1 = \int_{\gamma} \frac{[X(\xi)]^{-1}N^{-1}[I + \bar{N}N^{-1}]^{-1}}{\xi - z} d\xi$$

$$= \int_{\gamma} \frac{[X(\xi)]^{-1}[N + \bar{N}]^{-1}}{\xi - z} d\xi \quad \text{(A4)}$$

Since

$$N = \bar{N} = H^{-1} \quad \text{(A5)}$$

then,

$$\Phi'(z) = \frac{1}{2\pi i} X(z) \int_{\overline{ab}} \frac{[X_+(x_1)]^{-1}N^{-1}[ip]}{x_1 - z} dx_1$$

$$= \text{diag}\left(1 - \frac{z - (a+b)/2}{\sqrt{(z-a)(z-b)}}\right) \frac{H}{2}(ip) \quad \text{(A6)}$$

References

[1] Alshits, V. I., Darinskii, A. N., and Lothe, J., 1992, "On the Existence of Surface Waves in Half-Infinite Anisotropic Elastic Media with Piezoelectric and Piezomagnetic Properties," Wave Motion, **16**, pp. 265–283.

[2] Benveniste, Y., 1995, "Magnetoelectric Effect in Fibrous Composites with Piezoelectric and Piezomagnetic Phases," Phys. Rev. B, **51**(22), pp. 16424–16427.

[3] Bichurinm, M. I., Kornev, V. M., Petrov, V. M., Tatarenko, A. S., Kiliba, Y. V., and Srinivasan, G., 2001, "Theory of Magnetoelectric/Magnetostrictive Multilayer Composites," Phys. Rev. B, **64**, pp. 094409-1–094409-6.

[4] Bichurin, M. I., Filippov, D. A., Petrov, V. M., Laletsin, V. M., and Paddubnaya, N., 2003, "Resonance Magnetoelectric Effect in Layered Magnetostrictive-Piezoelectric Composites," Phys. Rev. B, **68**, pp. 132408-1–132408-4.

[5] Pan, E. 2002, "Three-Dimensional Green's Functions in Anisotropic Magneto-Electro-Elastic Bimaterials," Math. Phys., **53**, pp. 815–838.

[6] Sih, G. C., and Song, Z. F., 2003, "Magnetic and Electric Poling Effects Associated with Crack Growth in $BaTio_3-CoFe_2O_4$ Composites," Theor. Appl. Fract. Mech., **39**, pp. 209–227.

[7] Chung, M. Y., and Ting, T. C. T., 1995, "The Green Function for a Piezoelectric Piezomagnetic Anisotropic Elastic Medium with an Elliptic Hole or Rigid Inclusion," Philos. Mag. Lett., **72**(6), pp. 405–410.

[8] Gao, C. F., Kessler, H., and Balke, H., 2003, "Crack Problems in Magneto-electroelastic Solids. Part I: Exact Solution of a Crack," Int. J. Eng. Sci., **41**, pp. 969–981.

[9] Gao, C. F., Kessler, H., and Balke, H., 2003, "Crack Problems in Magneto-electroelastic Solids. Part I: General Solution of Collinear Cracks," Int. J. Eng. Sci., **41**, pp. 969–981.

[10] Song, Z. F., and Sih, G. C., 2003, "Crack Initiation Behavior in Magnetoelectroelasstic Composite Under In-Plane Deformation," Theor. Appl. Fract. Mech. **39**, pp. 189–207.

[11] Gao, C. F., Tong, P., and Zhang, T. Y., 2003, "Interfacial Crack Problems in Magneto-Electroelastic Solids," Int. J. Solids Struct., **41**, pp. 2105–2121.

[12] Gao, C. F., Tong, P., and Zhang, T. Y., 2004, "Fracture Mechanics for a Mode III Crack in a Magnetoelectroelastic Solid," Int. J. Solids Struct., **41**, pp. 6613–6629.

[13] Nan, C. W., 1994, "Magnetoelectric Effect in Composites of Piezoelectric and Piezomagnetic Phases," Phys. Rev. B, **50**(9), pp. 6082–6088.

[14] Stroh, A. N., 1958, "Dislocations and Cracks in Anisotropic Elasticity," Philos. Mag., **8**(3), pp. 625–646.

[15] Rudin, W., 1987, *Real and Complex Analysis*, McGraw-Hill, New York.

[16] Muskhelishvili, N. I., 1953, *Some Basic Problems of the Mathematical Theory of Elasticity*, P. Noordoff and Company, New York.

[17] McMeeking, R. M., 1989, "Electrostrictive Forces Near Crack Like Flaws," J. Appl. Math. Phys., **40**, pp. 615–627.

Hygrothermal Stresses in Unsymmetric Laminates Disturbed by Elliptical Holes

M. C. Hsieh

Chyanbin Hwu
e-mail: chwu@mail.ncku.edu.tw

Institute of Aeronautics and Astronautics,
National Cheng Kung University,
Tainan, Taiwan, R.O.C

Since the composite laminates are now in widespread use, many coupling phenomena within the composite laminates are widely focused in the world. One of those is the problem of stretching-bending coupling under hygrothermal environment. Because of the hygrothermal stress concentration, the understanding of hygrothermal effects on holes in laminates becomes important for the practical engineering design. However, due to mathematical infeasibility, most of the analytical solutions presented in literature are for two-dimensional problems or just for mechanical loading conditions, not for general composite laminates under hygrothermal environment. In order to establish a systematic analytical approach, in this study by extending the Stroh formalism for two-dimensional linear anisotropic elasticity and the Stroh-like formalism for coupled stretching-bending analysis of composite laminates, we develop an extended Stroh-like formalism for the coupled stretching-bending analysis under hygrothermal environment. By using the extended Stroh-like formalism developed in this paper, the explicit solutions for an unbounded laminate, symmetric or unsymmetric, disturbed by an elliptical hole subjected to uniform heat flow and moisture transfer in the x_1-x_2 plane or x_3 direction are now obtained. With these solutions three typical numerical examples are illustrated and compared by ANSYS finite element software package. [DOI: 10.1115/1.2074708]

1 Introduction

Composite laminates are increasingly being used not only in traditional areas like aerospace, but also in many engineering applications. Some of these applications are the structures under hygrothermal environment. Although the hygrothermal effects on holes in laminates have been widely discussed, due to mathematical infeasibility most of the analytical solutions found in the literature are for two-dimensional problems or for mechanical loading conditions, for isotropic materials, or for special laminates [1–3], not for the general composite laminates under hygrothermal environment. For a unidirectional lamina the coefficients of thermal and moisture expansion, like its other properties, change with direction. Thus, the hygrothermal changes result in unequal strains in the longitudinal and transverse directions. Hygrothermal strains do not produce a resultant force or moment when the body is completely free to expand, bend, and twist. However, for a composite laminate each individual lamina is not completely free to deform. The lamina stresses are therefore induced by the constraints placed on its deformation by adjacent lamina [4,5]. Like the cases of mechanical loading, the existence of holes in laminates will cause high stress concentration around holes under hygrothermal environment. Moreover, the unsymmetry of laminates will cause coupling between stretching and bending, which may complicate the analysis. Due to the designable characteristics of composite laminates, sometimes the engineering designers want to utilize the coupling effects to do something that cannot be achieved by using metallic or symmetric laminates. Thus, the study of hygrothermal stress analysis in unsymmertic laminates becomes important for practical engineering design.

Recently, we developed a Stroh-like formalism for coupled stretching-bending analysis of composite laminates [6] by extending the Stroh formalism for two-dimensional linear anisotropic elasticity [7,8] and successfully solved the hole problems in general composite laminates [9]. It is expected that with the help of our Stroh-like formalism, we may easily solve the corresponding hygrothermal problems of holes in general composite laminates. Like the extension of Stroh formalism to anisotropic thermoelasticity [3], we extend our Stroh-like formalism to the hygrothermal stress analysis of laminates in this paper. By using this formalism, the general solutions for hygrothermal stresses in unsymmetric laminates disturbed by an elliptical hole subjected to uniform heat flow and moisture transfer in the x_1-x_2 plane and x_3 direction are now obtained analytically. To illustrate our exact solutions, three numerical examples discussing the above hygrothermal effects of the holes in unsymmetric laminates are presented in this paper. The ANSYS finite element software package is also used to compare both numerical examples, which shows our present solutions are simple and correct.

2 Basic Equations

In a fixed rectangular coordinate system x_i, $i=1,2,3$, let U_i, σ_{ij}, e_{ij}, T, H, q_i, and m_i be, respectively, displacement, stress, strain, change in temperature, change in moisture content, heat flux, and moisture transfer. If the coupling terms between the elastic deformation, heat conduction, and moisture transport are neglected, the heat conduction, the moisture diffusion, the strain-displacement relation, the constitutive law, the force, heat and moisture equilibrium equations for linear anisotropic elastic materials under static loading, and small deformation conditions can be written as [10]

$$q_i = -k^t_{ij}T_{,j}, \quad m_i = -k^h_{ij}H_{,j}, \quad e_{ij} = \frac{1}{2}(U_{i,j} + U_{j,i}),$$

$$\sigma_{ij} = C_{ijks}e_{ks} - C_{ijks}\alpha^t_{ks}T - C_{ijks}\alpha^h_{ks}H,$$

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received March 23, 2005; final manuscript received June 28, 2005. Review conducted by 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.

$$\sigma_{ij,j}=0,\quad q_{i,i}=0,\quad m_{i,i}=0,\quad i,j,k,s=1,2,3, \tag{2.1}$$

where repeated indices imply summation, a subscript comma stands for differentiation, and C_{ijks}, k_{ij}^t, k_{ij}^h and $\alpha_{ij}^t, \alpha_{ij}^h$ are, respectively, the elastic constants, heat conduction coefficients, moisture diffusion coefficients, and the coefficients of thermal and moisture expansion. C_{ijks} are assumed to be fully symmetric, i.e., $C_{ijks}=C_{jiks}=C_{ijsk}=C_{ksij}$ and are required be positive definite due to the positiveness of strain energy. k_{ij}^t, k_{ij}^h, α_{ij}^t, and α_{ij}^h are also assumed to be symmetric, i.e., $k_{ij}^t=k_{ji}^t$, $k_{ij}^h=k_{ji}^h$, $\alpha_{ij}^t=\alpha_{ji}^t$, and $\alpha_{ij}^h=\alpha_{ji}^h$.

Equation (2.1) constitutes 23 partial differential equations in terms of three coordinate variables x_i , $i=1,2,3$. If the deformations are considered to be dependent upon two coordinate variables x_1 and x_2 only, a general solution satisfying these 23 equations can be found by following Stroh formalism for two-dimensional linear anisotropic thermoelasticity [3,8].

In this paper we consider a composite laminate composed of layers of various materials. Each layer is assumed to be made of anisotropic materials. If the laminate thickness is smaller than its other dimensions, according to Kirchhoff's assumptions the displacement, temperature, and moisture content may be assumed to vary linearly through laminate thickness as

$$U_i(x_1,x_2,x_3)=u_i(x_1,x_2)+x_3\beta_i(x_1,x_2),\quad i=1,2,$$

$$U_3(x_1,x_2,x_3)=w(x_1,x_2),$$

$$T(x_1,x_2,x_3)=T^0(x_1,x_2)+x_3T^*(x_1,x_2),$$

$$H(x_1,x_2,x_3)=H^0(x_1,x_2)+x_3H^*(x_1,x_2), \tag{2.2a}$$

where

$$\beta_1=-w_{,1},\quad \beta_2=-w_{,2}. \tag{2.2b}$$

(u_1,u_2,w), T^0 and H^0 are the middle surface displacements, temperature, and moisture content, and β_i, $i=1,2$, are the negative of the slope of the middle surface in the x_1 and x_2 directions. T^* and H^* are the rates of changes in temperature and moisture content.

Based upon the assumptions given in Eqs. (2.2) and the 23 basic Eqs. (2.1) for anisotropic materials under hygrothermal condition, we may now write down the kinematic relations, the constitutive laws, and the equilibrium equations for hygrothermal stress analysis of composite laminates as follows:

$$\breve{q}_i=-K_{ij}^tT_{,j}^0-K_{ij}^{*t}T_{,j}^*-K_{i3}^tT^*,\quad \breve{m}_i=-K_{ij}^hH_{,j}^0-K_{ij}^{*h}H_{,j}^*-K_{i3}^hH^*,$$

$$\varepsilon_{ij}=\frac{1}{2}(u_{i,j}+u_{j,i}),\quad \kappa_{ij}=\frac{1}{2}(\beta_{i,j}+\beta_{j,i}),$$

$$N_{ij}=A_{ijkl}\varepsilon_{kl}+B_{ijkl}\kappa_{kl}-A_{ij}^tT^0-A_{ij}^hH^0-B_{ij}^tT^*-B_{ij}^hH^*,$$

$$M_{ij}=B_{ijkl}\varepsilon_{kl}+D_{ijkl}\kappa_{kl}-B_{ij}^tT^0-B_{ij}^hH^0-D_{ij}^tT^*-D_{ij}^hH^*,$$

$$N_{ij,j}=0,\quad M_{ij,ij}+p=0,\quad Q_i=M_{ij,j},\quad \breve{q}_{i,i}+q=0,$$

$$\breve{m}_{i,i}+m=0,\quad i,j,k,l=1,2, \tag{2.3}$$

where ε_{ij} and κ_{ij} denote the midplane strain and plate curvature; N_{ij}, M_{ij}, and Q_i denote the stress resultants, bending moments and shear forces; $\breve{q}_i$ and $\breve{m}_i$ denote the heat flux resultant and moisture transfer resultant, A_{ijkl}, B_{ijkl}, and D_{ijkl} are, respectively, the extensional, coupling and bending stiffness tensors; $A_{ij}^t, B_{ij}^t, D_{ij}^t$ and $A_{ij}^h, B_{ij}^h, D_{ij}^h$ are the corresponding tensors for the thermal and moisture expansion coefficients; K_{ij}^t, K_{ij}^h and K_{ij}^{*t}, K_{ij}^{*h} are the coefficients related to the heat conduction and moisture diffusion coefficients; p, q, and m are the lateral distributed load, heat flux, and moisture concentration transfer applied on the laminates. Their definitions are

$$N_{ij}=\int_{-h/2}^{h/2}\sigma_{ij}dx_3,\quad M_{ij}=\int_{-h/2}^{h/2}\sigma_{ij}x_3dx_3,\quad Q_i=\int_{-h/2}^{h/2}\sigma_{i3}dx_3,$$

$$\breve{q}_i=\int_{-h/2}^{h/2}q_idx_3,\quad \breve{m}_i=\int_{-h/2}^{h/2}m_idx_3,$$

$$A_{ijks}=\int_{-h/2}^{h/2}C_{ijks}dx_3,\quad B_{ijks}=\int_{-h/2}^{h/2}C_{ijks}x_3dx_3,$$

$$D_{ijks}=\int_{-h/2}^{h/2}C_{ijks}x_3^2dx_3,$$

$$A_{ij}^t=\int_{-h/2}^{h/2}C_{ijks}\alpha_{ks}^tdx_3,\quad B_{ij}^t=\int_{-h/2}^{h/2}C_{ijks}\alpha_{ks}^tx_3dx_3,$$

$$D_{ij}^t=\int_{-h/2}^{h/2}C_{ijks}\alpha_{ks}^tx_3^2dx_3,$$

$$A_{ij}^h=\int_{-h/2}^{h/2}C_{ijks}\alpha_{ks}^hdx_3,\quad B_{ij}^h=\int_{-h/2}^{h/2}C_{ijks}\alpha_{ks}^hx_3dx_3,$$

$$D_{ij}^h=\int_{-h/2}^{h/2}C_{ijks}\alpha_{ks}^hx_3^2dx_3,$$

$$K_{ij}^t=\int_{-h/2}^{h/2}k_{ij}^tdx_3,\quad K_{ij}^{*t}=\int_{-h/2}^{h/2}k_{ij}^tx_3dx_3,\quad K_{ij}^h=\int_{-h/2}^{h/2}k_{ij}^hdx_3,$$

$$K_{ij}^{*h}=\int_{-h/2}^{h/2}k_{ij}^hx_3dx_3, \tag{2.4}$$

in which h is laminate thickness. Note that like C_{ijks}, k_{ij}^t, k_{ij}^h, α_{ij}^t, and α_{ij}^h, according to the definitions given in Eq. (2.4) A_{ijkl}, B_{ijkl}, D_{ijkl}, A_{ij}^t, B_{ij}^t, D_{ij}^t, A_{ij}^h, B_{ij}^h, D_{ij}^h, K_{ij}^t, K_{ij}^h, K_{ij}^{*t}, and K_{ij}^{*h} still preserve the symmetry property.

3 Extended Stroh-Like Formalism

Since the basic equations stated in Eqs. (2.2) and (2.3) are quite general, it is not easy to find a solution satisfying all these basic equations. In the following, we consider two special cases that occur frequently in engineering applications. One is the case that temperature and moisture distributions depend on x_1 and x_2 only, i.e., $T^*=H^*=0$, and the other is the case that temperature and moisture distributions depend on x_3 only, i.e., $T=T^0+x_3T^*$ and $H=H^0+x_3H^*$ in which T^0, T^*, H^0, and H^* are constants independent of x_1 and x_2.

3.1 Case 1: Temperature and Moisture Content Depend on x_1 and x_2 Only. If the temperature and moisture content are assumed to depend on x_1 and x_2 only and the lateral distributed load, heat flux, and moisture concentration transfer applied on the laminates are neglected, i.e., $T^*=H^*=p=q=m=0$, the basic equations stated in Eq. (2.3) can be simplified as

$$\breve{q}_{i,i}=-K_{ij}^tT_{,ij}=0,\quad \breve{m}_{i,i}=-K_{ij}^hH_{,ij}=0,$$

$$N_{ij,j}=A_{ijkl}u_{k,lj}+B_{ijkl}\beta_{k,lj}-A_{ij}^tT_{,j}-A_{ij}^hH_{,j}=0,$$

$$M_{ij,ij}=B_{ijkl}u_{k,lij}+D_{ijkl}\beta_{k,lij}-B_{ij}^tT_{,ij}-B_{ij}^hH_{,ij}=0,\quad i,j,k,l=1,2. \tag{3.1}$$

By following the steps described in [3] for Stroh formalism of two-dimensional thermoelasticity and in [6] for Stroh-like formal-

ism of coupled stretching-bending analysis of composite laminates, we can find a general solution satisfying the basic Eqs. (3.1) and call it the extended Stroh-like formalism. The solution is

$$T = 2\,\mathrm{Re}\{g_t'(z_t)\}, \quad H = 2\,\mathrm{Re}\{g_h'(z_h)\},$$

$$\breve{q}_i = -2\,\mathrm{Re}\{(K_{i1}^t + \tau_t K_{i2}^t) g_t''(z_t)\}, \quad \breve{m}_i = -2\,\mathrm{Re}\{(K_{i1}^h + \tau_h K_{i2}^h) g_h''(z_h)\},$$

$$\mathbf{u}_d = 2\,\mathrm{Re}\{\mathbf{A}\mathbf{f}(z) + \mathbf{c}_t g_t(z_t) + \mathbf{c}_h g_h(z_h)\},$$

$$\boldsymbol{\phi}_d = 2\,\mathrm{Re}\{\mathbf{B}\mathbf{f}(z) + \mathbf{d}_t g_t(z_t) + \mathbf{d}_h g_h(z_h)\}, \tag{3.2a}$$

where

$$\mathbf{u}_d = \begin{Bmatrix} \mathbf{u} \\ \boldsymbol{\beta} \end{Bmatrix}, \quad \boldsymbol{\phi}_d = \begin{Bmatrix} \boldsymbol{\phi} \\ \boldsymbol{\psi} \end{Bmatrix}, \quad \mathbf{u} = \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix}, \quad \boldsymbol{\beta} = \begin{Bmatrix} \beta_1 \\ \beta_2 \end{Bmatrix},$$

$$\boldsymbol{\phi} = \begin{Bmatrix} \phi_1 \\ \phi_2 \end{Bmatrix}, \quad \boldsymbol{\psi} = \begin{Bmatrix} \psi_1 \\ \psi_2 \end{Bmatrix}, \tag{3.2b}$$

$$\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3\ \mathbf{a}_4], \quad \mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2\ \mathbf{b}_3\ \mathbf{b}_4], \tag{3.2c}$$

$$\mathbf{f}(z) = \begin{Bmatrix} f_1(z_1) \\ f_2(z_2) \\ f_3(z_3) \\ f_4(z_4) \end{Bmatrix}, \quad z_k = x_1 + \mu_k x_2, \quad k = 1,2,3,4, \tag{3.2d}$$

$$z_t = x_1 + \tau_t x_2, \quad z_h = x_1 + \tau_h x_2. \tag{3.2e}$$

In the above, Re stands for the real part of a complex number and the prime ($\bullet'$) denotes differentiation with respective to its argument. $\mathbf{u}_d$ and ϕ_d are the generalized displacement and stress function vectors. ϕ_1, ϕ_2 and ψ_1, ψ_2 are the stress functions related to the stress resultants N_{ij}, shear forces Q_i, effective shear forces V_i and bending moments M_{ij} by

$$N_{i1} = -\phi_{i,2}, \quad N_{i2} = \phi_{i,1},$$

$$M_{i1} = -\psi_{i,2} - \lambda_{i1}\eta, \quad M_{i2} = \psi_{i,1} - \lambda_{i2}\eta,$$

$$Q_1 = -\eta_{,2}, \quad Q_2 = \eta_{,1}, \quad V_1 = -\psi_{2,22}, \quad V_2 = \psi_{1,11}, \tag{3.3a}$$

where

$$\eta = \frac{1}{2}\psi_{k,k} = \frac{1}{2}(\psi_{1,1} + \psi_{2,2}), \tag{3.3b}$$

and λ_{ij} is the permutation tensor defined as

$$\lambda_{11} = \lambda_{22} = 0, \quad \lambda_{12} = -\lambda_{21} = 1. \tag{3.3c}$$

$f_k(z_k)$, $k=1,2,3,4$, $g_t(z_t)$, and $g_h(z_h)$ are six holomorphic functions of complex variables z_k, z_t, and z_h, which will be determined by the boundary conditions set for each particular problem. μ_k, τ_t, τ_h and $(\mathbf{a}_k, \mathbf{b}_k)$, $(\mathbf{c}_t, \mathbf{d}_t)$, $(\mathbf{c}_h, \mathbf{d}_h)$ are, respectively, the *material eigenvalues* and *eigenvectors*, which can be determined by the following eigenrelation

$$K_{11}^t + 2\tau_t K_{12}^t + \tau_t^2 K_{22}^t = 0, \quad K_{11}^h + 2\tau_h K_{12}^h + \tau_h^2 K_{22}^h = 0, \tag{3.4a}$$

$$\mathbf{N}\xi = \mu\xi, \quad \mathbf{N}\boldsymbol{\eta}_t = \tau_t\boldsymbol{\eta}_t + \boldsymbol{\gamma}_t, \quad \mathbf{N}\boldsymbol{\eta}_h = \tau_h\boldsymbol{\eta}_h + \boldsymbol{\gamma}_h, \tag{3.4b}$$

where $\mathbf{N}$ is a 8×8 real matrix which is the *fundamental elasticity matrix* for coupled stretching-bending analysis; ξ, $\boldsymbol{\eta}_t$, and $\boldsymbol{\eta}_h$ are three 8×1 complex vectors which are composed of the material eigenvectors. These matrices and vectors are composed of some well-defined submatrices and vectors and are defined by

$$\mathbf{N} = \begin{bmatrix} \mathbf{N}_1 & \mathbf{N}_2 \\ \mathbf{N}_3 & \mathbf{N}_1^T \end{bmatrix}, \quad \xi = \begin{Bmatrix} \mathbf{a} \\ \mathbf{b} \end{Bmatrix}, \quad \boldsymbol{\eta}_t = \begin{Bmatrix} \mathbf{c}_t \\ \mathbf{d}_t \end{Bmatrix}, \quad \boldsymbol{\eta}_h = \begin{Bmatrix} \mathbf{c}_h \\ \mathbf{d}_h \end{Bmatrix}, \tag{3.5}$$

in which the superscript T denotes the transpose of a matrix. Detailed definitions of the submatrices $\mathbf{N}_1$, $\mathbf{N}_2$, and $\mathbf{N}_3$ have been given in [6] and are listed in the Appendix A for the readers' convenience. $\boldsymbol{\gamma}_t$ and $\boldsymbol{\gamma}_h$ are two 8×1 complex vectors related to the elastic constants and the coefficients of thermal and moisture expansion, whose detailed expressions can be obtained by following either the displacement formalism or mixed formalism described in [6]. They are

$$\boldsymbol{\gamma}_t = \left(\mathbf{L}_2 + \frac{1}{2}\mathbf{J}_2\right)^{-1} \begin{Bmatrix} \boldsymbol{\alpha}_1^t \\ \boldsymbol{\alpha}_2^t \end{Bmatrix} = -\mathbf{I}_t \begin{bmatrix} \mathbf{0} & (\mathbf{N}_m)_2 \\ \mathbf{I} & (\mathbf{N}_m)_1^T \end{bmatrix} \begin{Bmatrix} \tilde{\boldsymbol{\alpha}}_1^t \\ \tilde{\boldsymbol{\alpha}}_2^t \end{Bmatrix}, \tag{3.6a}$$

where

$$\boldsymbol{\alpha}_1^t = \begin{Bmatrix} \boldsymbol{\alpha}_{A1}^t \\ \boldsymbol{\alpha}_{B1}^t \end{Bmatrix}, \quad \boldsymbol{\alpha}_2^t = \begin{Bmatrix} \boldsymbol{\alpha}_{A2}^t \\ \boldsymbol{\alpha}_{B2}^t \end{Bmatrix}, \quad \tilde{\boldsymbol{\alpha}}_1^t = \begin{Bmatrix} \tilde{\boldsymbol{\alpha}}_{A1}^t \\ \tilde{\boldsymbol{\alpha}}_{B2}^t \end{Bmatrix}, \quad \tilde{\boldsymbol{\alpha}}_2^t = \begin{Bmatrix} \tilde{\boldsymbol{\alpha}}_{A2}^t \\ -\tilde{\boldsymbol{\alpha}}_{B1}^t \end{Bmatrix}, \tag{3.6b}$$

and

$$\boldsymbol{\alpha}_{Ai}^t = \begin{Bmatrix} A_{1i}^t \\ A_{2i}^t \end{Bmatrix}, \quad \boldsymbol{\alpha}_{Bi}^t = \begin{Bmatrix} B_{1i}^t \\ B_{2i}^t \end{Bmatrix}, \quad \tilde{\boldsymbol{\alpha}}_{Ai}^t = \begin{Bmatrix} \tilde{A}_{1i}^t \\ \tilde{A}_{2i}^t \end{Bmatrix},$$

$$\tilde{\boldsymbol{\alpha}}_{Bi}^t = \begin{Bmatrix} \tilde{B}_{1i}^t \\ \tilde{B}_{2i}^t \end{Bmatrix}, \quad i = 1,2, \tag{3.6c}$$

$$\tilde{A}_{ij}^t = A_{ij}^t - \tilde{B}_{ijkl} B_{kl}^t, \quad \tilde{B}_{ij}^t = \tilde{D}_{ijkl} B_{kl}^t, \tag{3.6d}$$

in which $\tilde{B}_{ijkl}$ and $\tilde{D}_{ijkl}$ are the tensor notations of $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{D}}$ defined in Eq. (A3). Same expressions as Eq. (3.6) are defined for γ_h only by replacing the subscript or superscript from t to h. Detailed definitions of $\mathbf{L}_2$, $\mathbf{J}_2$, $\mathbf{I}_t$, $(\mathbf{N}_m)_1$, and $(\mathbf{N}_m)_2$ have also been given in [6] and are listed in Appendix A. The second equality of Eq. (3.6*a*) comes from the equivalence between the displacement formalism and mixed formalism discussed in [6].

By using the relations given in Eq. (3.3), the stress resultants N_n, N_s, N_{ns}, bending moments M_n, M_s, M_{ns}, shear forces Q_n, Q_s and effective shear forces V_n, V_s in the tangential-normal (s-n) coordinate system, can be obtained directly from the stress functions and their relations as [9]

$$N_n = \mathbf{n}^T\boldsymbol{\phi}_{,}, \quad N_{ns} = \mathbf{s}^T\boldsymbol{\phi}_{,s} = -\mathbf{n}^T\boldsymbol{\phi}_{,n}, \quad N_s = -\mathbf{s}^T\boldsymbol{\phi}_{,n},$$

$$M_n = \mathbf{n}^T\boldsymbol{\psi}_{,}, \quad M_{ns} = \mathbf{s}^T\boldsymbol{\psi}_{,s} - \eta = -\mathbf{n}^T\boldsymbol{\psi}_{,n} + \eta, \quad M_s = -\mathbf{s}^T\boldsymbol{\psi}_{,n},$$

$$Q_n = \eta_{,s}, \quad Q_s = -\eta_{,n}, \quad V_n = (\mathbf{s}^T\boldsymbol{\psi}_{,s})_{,s}, \quad V_s = -(\mathbf{n}^T\boldsymbol{\psi}_{,n})_{,n}, \tag{3.7a}$$

where

$$\eta = \frac{1}{2}(\mathbf{s}^T\boldsymbol{\psi}_{,s} + \mathbf{n}^T\boldsymbol{\psi}_{,n}), \tag{3.7b}$$

and

$$\mathbf{s}^T = (\cos\theta, \sin\theta), \quad \mathbf{n}^T = (-\sin\theta, \cos\theta). \tag{3.7c}$$

3.2 Case 2: Temperature and Moisture Content Depend on x_3 Only. If the temperature and moisture content depend on x_3 only, the distribution assumed in the third and fourth equations of Eq. (2.2*a*) can be rewritten as

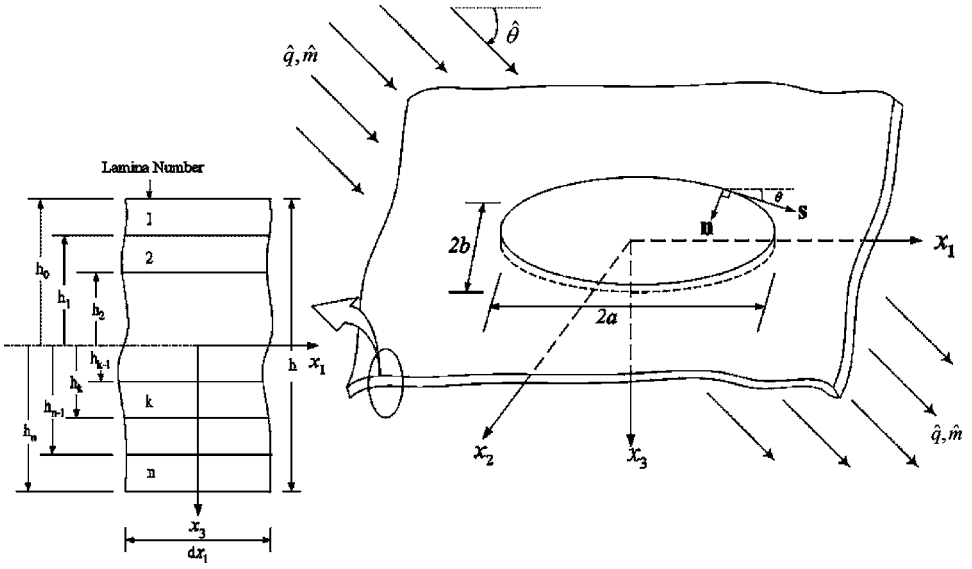


Fig. 1 Unsymmetric laminate weakened by an elliptical hole subjected to uniform heat flow and moisture transfer in x_1-x_2 plane

$$T = T^0 + x_3 T^*, \quad H = H^0 + x_3 H^*, \tag{3.8}$$

where T^0, T^*, H^0, and H^* are real constants independent of x_1 and x_2. With this assumption, the basic Eqs. (2.3) can be simplified as

$$\breve{q}_i = -K^t_{i3} T^*, \quad \breve{m}_i = -K^h_{i3} H^*,$$

$$N_{ij} = A_{ijkl} u_{k,l} + B_{ijkl} \beta_{k,l} - A^t_{ij} T^0 - A^h_{ij} H^0 - B^t_{ij} T^* - B^h_{ij} H^*,$$

$$M_{ij} = B_{ijkl} u_{k,l} + D_{ijkl} \beta_{k,l} - B^t_{ij} T^0 - B^h_{ij} H^0 - D^t_{ij} T^* - D^h_{ij} H^*,$$

$$N_{ij,j} = A_{ijkl} u_{k,lj} + B_{ijkl} \beta_{k,lj} = 0,$$

$$M_{ij,ij} = B_{ijkl} u_{k,lij} + D_{ijkl} \beta_{k,lij} = 0,$$

$$Q_i = M_{ij,j}, \quad i,j,k,l = 1,2, \tag{3.9}$$

and $\breve{q}_{i,i}=0$ and $\breve{m}_{i,i}=0$ are satisfied automatically. Note that the mathematical expressions of the governing equations written in Eq. (3.9), i.e., $N_{ij,j}=0$ and $M_{ij,ij}=0$ in terms of u_k and β_k, are exactly the same as those of the nonhygrothermal problems discussed in [6]. Therefore, the general solution for u_k and β_k should be exactly the same as that presented in [6]. Substituting this solution into the third and fourth equations of Eq. (3.9) and following the steps of [6], we can find the solution for N_{ij} and M_{ij}. The final solution expressed in the form of Stroh-like formalism can then be written as

$$\mathbf{u}_d = 2\,\mathrm{Re}\{\mathbf{A}\mathbf{f}(z)\},$$

$$\boldsymbol{\phi}_d = 2\,\mathrm{Re}\{\mathbf{B}\mathbf{f}(z)\} - x_1 \boldsymbol{\vartheta}_2 + x_2 \boldsymbol{\vartheta}_1, \tag{3.10a}$$

where

$$\boldsymbol{\vartheta}_i = \boldsymbol{\alpha}^t_i T^0 + \boldsymbol{\alpha}^h_i H^0 + \boldsymbol{\alpha}^{*t}_i T^* + \boldsymbol{\alpha}^{*h}_i H^*, \quad i = 1,2. \tag{3.10b}$$

In the above, $\boldsymbol{\alpha}^t_i$ and $\boldsymbol{\alpha}^h_i$ have been defined in (3.6*b*), whereas $\boldsymbol{\alpha}^{*t}_i$ are defined by

$$\boldsymbol{\alpha}^{*t}_i = \begin{Bmatrix} \boldsymbol{\alpha}^t_{Bi} \\ \boldsymbol{\alpha}^t_{Di} \end{Bmatrix}, \quad \boldsymbol{\alpha}^t_{Bi} = \begin{Bmatrix} B^t_{1i} \\ B^t_{2i} \end{Bmatrix}, \quad \boldsymbol{\alpha}^t_{Di} = \begin{Bmatrix} D^t_{1i} \\ D^t_{2i} \end{Bmatrix}, \quad i = 1,2, \tag{3.11}$$

and same expressions as Eq. (3.11) are defined for $\boldsymbol{\alpha}^{*h}_i$ only by replacing superscript from *t* to *h*.

With the general solution shown in Eq. (3.10) for the generalized displacement and stress function vectors, the middle surface displacements u_i and slopes β_i can be obtained directly from the components of the generalized displacement vector $\mathbf{u}_d$. As to the stress resultants N_{ij}, shear forces Q_i, effective shear forces V_i, and bending moments M_{ij}, we can utilize the relations shown in Eq. (3.3). Moreover, all the relations for the nonhygrothermal problems such as the eigenrelation (first equation of (3.4*b*)), force relation (3.7), and those shown in Eqs. (A1)–(A3) are all valid for the present case.

Note that unlike the general solution shown in Eq. (3.2) for case 1, the general solution shown in Eq. (3.10) for case 2 does not include the expressions for temperature, moisture content, heat flux resultant and moisture transfer since they have been given in the first and second equations of Eqs. (3.8) and (3.9) as *known* linear distributions and constant flows. In other words, for case 1 to find the plane distributions of T, H, $\breve{q}_i$, and $\breve{m}_i$ we need to prescribe their associated boundary conditions, while for case 2 their distributions are assumed to be known and hence the boundary conditions for these physical quantities will not be stated in the related problems.

4 Hole Problems

4.1 Case 1: Uniform Heat Flow and Moisture Transfer in x_1-x_2 Plane. In an infinite composite laminate, heat $\hat{q}$ and moisture $\hat{m}$ is flowing uniformly in the direction of angle $\hat{\theta}$ clockwise from the positive x_1 axis (Fig. 1). The uniform steady heat and moisture flow is disturbed by the presence of an insulated elliptic hole whose boundary is given by

$$x_1 = a\cos\varphi, \quad x_2 = b\sin\varphi, \tag{4.1}$$

where $2a$, $2b$ are the length of the major and minor axes of the ellipse and φ is a real parameter. If the hole is assumed to be free of tractions, the boundary conditions for this problem can be written as,

$$\breve{q}_i \to \hat{q}_i, \quad \breve{m}_i \to \hat{m}_i, \quad N_{ij} \to 0,$$

$$M_{ij} \to 0, \quad i,j = 1,2, \quad \text{at infinity}, \tag{4.2a}$$

$$\breve{q}_n = \breve{m}_n = 0, \quad N_n = N_{ns} = 0,$$

$$M_n = V_n = 0 \quad \text{along the hole boundary,} \tag{4.2b}$$

where $\breve{q}_n$ and $\breve{m}_n$ are the heat flux and moisture transfer in the direction of $\mathbf{n}$ which is normal to the surface of the elliptic hole. From the relations (3.7) we know that the traction boundary condition given in Eq. (4.2*a*) and (4.2*b*) can be expressed in terms of the generalized stress function vector $\boldsymbol{\phi}_d$ as

$$\boldsymbol{\phi}_d \to \mathbf{0}, \quad \text{at infinity,}$$

$$\boldsymbol{\phi}_{d,s} = \mathbf{0}, \quad \text{along the hole boundary.} \tag{4.3}$$

To consider the heat flux and moisture transfer boundary conditions given in Eq. (4.2*b*), we now use the coordinate transformation and employ the third and fourth equations of (3.2*a*) and Eq. (3.4*a*) to get

$$\begin{aligned}
\breve{q}_n &= -\breve{q}_1 \sin\theta + \breve{q}_2 \cos\theta = 2K_t \,\mathrm{Im}\{(\cos\theta + \tau_t \sin\theta) g_t''(z_t)\},\\
\breve{m}_n &= -\breve{m}_1 \sin\theta + \breve{m}_2 \cos\theta = 2K_h \,\mathrm{Im}\{(\cos\theta + \tau_h \sin\theta) g_h''(z_h)\},
\end{aligned} \tag{4.4}$$

where Im stands for the imaginary part; $K_t = K_{22}^t(\tau_t - \bar{\tau}_t)/2i$ and $K_h = K_{22}^h(\tau_h - \bar{\tau}_h)/2i$ are real constants and the overbar denotes the complex conjugates; θ is the angle between the x_1 axis and the tangent direction $\mathbf{s}$ (Fig. 1) and is related to φ by

$$\rho\cos\theta = a\sin\varphi, \quad \rho\sin\theta = -b\cos\varphi, \tag{4.5a}$$

in which

$$\rho = \sqrt{a^2\sin^2\varphi + b^2\cos^2\varphi}. \tag{4.5b}$$

Because the mathematical forms of the general solutions (3.2) and boundary conditions (4.2) together with the relations (4.3)–(4.5) are the same as those of the corresponding two-dimensional problems, by following the steps described in [3] the unknown functions in the general solution (3.2) can be assumed to be

$$\mathbf{f}(z) = \sum_{k=0}^{2} \langle f_k(z_\alpha, \mu_\alpha)\rangle \{\mathbf{A}^T \mathbf{q}_{ka} + \mathbf{B}^T \mathbf{q}_{kb}\},$$

$$g_t(z_t) = \sum_{k=0}^{2} \hat{g}_k^t f_k(z_t, \tau_t), \quad g_h(z_h) = \sum_{k=0}^{2} \hat{g}_k^h f_k(z_h, \tau_h), \tag{4.6a}$$

where the angular bracket $\langle\rangle$ stands for the diagonal matrix in which each component is varied according to its subscript α, for example., $\langle z_\alpha\rangle = \mathrm{diag.}[z_1, z_2, z_3, z_4]$, and

$$f_0(z,\mu) = \frac{1}{2}z^2,$$

$$f_1(z,\mu) = \frac{1}{a + i\mu b}\left\{\frac{1}{2}z^2 - \frac{1}{2}z\sqrt{z^2 - (a^2 + \mu^2 b^2)}\right\},$$

$$f_2(z,\mu) = \frac{a - i\mu b}{2}\log\{z + \sqrt{z^2 - (a^2 + \mu^2 b^2)}\}, \tag{4.6b}$$

and

$$\hat{g}_1^t = \hat{g}_2^t, \quad \hat{g}_1^h = \hat{g}_2^h. \tag{4.6c}$$

The problem now reduces to the determination of the unknown constants $\mathbf{q}_{ka}$, $\mathbf{q}_{kb}$, $\hat{g}_k^t$ and $\hat{g}_k^h$, $k = 0, 1, 2$, which should satisfy the boundary conditions set in Eq. (4.2) with useful relations given in Eqs. (4.3) and (4.4). Substituting the assumed functions (4.6) into the general solutions (3.2) and the boundary conditions (4.2), and following the steps described in [3,11], we get

$$\hat{g}_0^t = \frac{1}{2K_t^2}\{(K_{12}^t + iK_t)\hat{q}_2 - K_{22}^t \hat{q}_1\},$$

$$\hat{g}_0^h = \frac{1}{2K_h^2}\{(K_{12}^h + iK_h)\hat{m}_2 - K_{22}^h \hat{m}_1\},$$

$$\hat{g}_1^t = -\frac{1}{2K_t}\{b\hat{q}_1 + ia\hat{q}_2\}, \quad \hat{g}_1^h = -\frac{1}{2K_h}\{b\hat{m}_1 + ia\hat{m}_2\}, \tag{4.7a}$$

$$\mathbf{q}_{0a} = -2\,\mathrm{Re}\{\hat{g}_0^t \mathbf{d}_t + \hat{g}_0^h \mathbf{d}_h\},$$

$$\mathbf{q}_{0b} = 2\mathbf{N}_3^{\langle -1\rangle}\,\mathrm{Re}\{\hat{g}_0^t(\mathbf{N}_1^T - \tau_t\mathbf{I})\mathbf{d}_t + \hat{g}_0^h(\mathbf{N}_1^T - \tau_h\mathbf{I})\mathbf{d}_h\} + k^*\mathbf{m}(0), \tag{4.7b}$$

$$\begin{Bmatrix}\mathbf{q}_{1b}\\ \mathbf{q}_{1a}\end{Bmatrix} = 2[a\tilde{\mathbf{N}} - b\mathbf{N}]^{-1}\,\mathrm{Im}\begin{Bmatrix}\mathbf{L}^{-1}(\mathbf{S}^T - i\mathbf{I})\mathbf{d}_1^*\\ \mathbf{d}_1^*\end{Bmatrix},$$

$$\begin{Bmatrix}\mathbf{q}_{2b}\\ \mathbf{q}_{2a}\end{Bmatrix} = 2[a\tilde{\mathbf{N}} + b\mathbf{N}]^{-1}\,\mathrm{Im}\{\hat{g}_1^t(a - ib\tau_t)\boldsymbol{\eta}_t + \hat{g}_1^h(a - ib\tau_h)\boldsymbol{\eta}_h\}, \tag{4.7c}$$

where

$$\mathbf{d}_1^* = \hat{g}_1^t(a + ib\tau_t)\mathbf{d}_t + \hat{g}_1^h(a + ib\tau_h)\mathbf{d}_h. \tag{4.8}$$

In Eq. (4.7), $\mathbf{S}$ and $\mathbf{L}$ are real matrices and are generally called *Barnett-Lothe tensors* in two-dimensional problems whose detailed definitions can be found in [8] and are listed in Appendix B; $\tilde{\mathbf{N}}$ is the average fundamental matrix whose definition is also listed in Appendix B; the arbitrary constant k^* can be determined by substituting Eq. (4.7*b*) into the following equation which comes from the satisfaction of the first equation in Eq. (4.3) [11]

$$\mathbf{N}_1^{T(2)}\mathbf{q}_{0a} + \mathbf{N}_3^{(2)}\mathbf{q}_{0b} + 2\,\mathrm{Re}\{\hat{g}_0^t\tau_t^2\mathbf{d}_t + \hat{g}_0^h\tau_h^2\mathbf{d}_h\} = \mathbf{0}, \tag{4.9}$$

where

$$\mathbf{N}_1^{T(2)} = \mathbf{N}_3\mathbf{N}_2 + \mathbf{N}_1^T\mathbf{N}_1^T, \quad \mathbf{N}_3^{(2)} = \mathbf{N}_3\mathbf{N}_1 + \mathbf{N}_1^T\mathbf{N}_3, \tag{4.10}$$

and $\mathbf{m}(0) = (0\ 1\ 0\ 0)^T$.

One should note that the inverse matrix of $\mathbf{N}_3$ does not exist because it is a singular matrix [12,13]. The pseudoinverse matrix $\mathbf{N}_3^{\langle -1\rangle}$ shown in the second equation in Eq. (4.7*b*) is defined as

$$\mathbf{N}_3^{\langle -1\rangle}\mathbf{N}_3 = \mathbf{N}_3\mathbf{N}_3^{\langle -1\rangle} = \mathbf{I}_3, \tag{4.11a}$$

where

$$\mathbf{I}_3 = \begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}. \tag{4.11b}$$

The general solution (3.2) together with Eqs. (4.6)–(4.10) now provides us the complete full field solution for the present problem. With this solution any mechanical responses inside the body can be obtained directly from the relations provided in Secs. 2 and 3. In engineering application, one is usually interested in the hoop stress along the hole boundary. According to the experience of two-dimensional problems, by using the identities developed in the literature for converting complex form to real form [8], it is usually possible to get real form solutions for hoop stress. From Eq. (3.7), we see that the real form solution of the hoop stress can be obtained directly if we have the real form solution for $\boldsymbol{\phi}_{d,n}$, i.e., $(\boldsymbol{\phi}_{,n}, \boldsymbol{\psi}_{,n})$. Note that we do not need to calculate $\boldsymbol{\phi}_{,s}$ and $\boldsymbol{\psi}_{,s}$ along the hole boundary since they are identical to zero when the second equation in the boundary condition (4.3) is satisfied in our solution. Again, by following the steps described in [3], the explicit form solution of $\boldsymbol{\phi}_{d,n}$ for the present case is obtained as

$$\boldsymbol{\phi}_{d,n}=\frac{1}{K_t}(b\hat{q}_1\cos\varphi+a\hat{q}_2\sin\varphi)\boldsymbol{\gamma}_2^t(\theta)+\frac{1}{K_h}(b\hat{m}_1\cos\varphi+a\hat{m}_2\sin\varphi)\boldsymbol{\gamma}_2^h(\theta)+\frac{1}{\rho}\{\mathbf{G}_3(\theta)\mathrm{Re}\{ie^{-2i\varphi}[\hat{g}_1^t(a+ib\tau_t)\tilde{\boldsymbol{\gamma}}_2^{t*}+\hat{g}_1^h(a+ib\tau_h)\tilde{\boldsymbol{\gamma}}_2^{h*}]\}\},\tag{4.12a}$$

where

$$\mathbf{G}_3(\theta)=-\mathbf{N}_3(\theta)\mathbf{L}^{-1}.\tag{4.12b}$$

In Eq. (4.12), $\mathbf{N}_3(\theta)$ is one of the submatrices of the generalized fundamental matrix $\mathbf{N}(\theta)$ whose definition can be found in [8]; $\boldsymbol{\gamma}_2^t(\theta)$, $\boldsymbol{\gamma}_2^h(\theta)$, $\tilde{\boldsymbol{\gamma}}_2^{t*}$, and $\tilde{\boldsymbol{\gamma}}_2^{h*}$ are vectors related to thermal and moisture moduli and their detailed definitions are given in Appendix C.

4.2 Case 2: Uniform Heat Flow and Moisture Transfer in x_3 Direction. All the conditions are the same as case 1 except that the heat $\hat{q}$ and moisture $\hat{m}$ is now flowing uniformly in the thickness direction instead of plane direction. This may occur when the temperature and moisture content on the top and bottom surfaces of the laminate are different, for example, T_u and H_u on the top surface and T_l and H_l on the bottom surface. If the temperature and moisture content are assumed to vary linearly as shown by Eq. (3.8), we have

$$T^0=\frac{T_l+T_u}{2},\quad H^0=\frac{H_l+H_u}{2},\quad T^*=\frac{T_l-T_u}{h},\quad H^*=\frac{H_l-H_u}{h}.\tag{4.13}$$

By the first and second equations of Eq. (3.9), the heat flux $\hat{q}$ and moisture transfer $\hat{m}$ in the thickness direction are related to T^* and H^* by

$$\hat{q}=-K_{33}^tT^*,\quad \hat{m}=-K_{33}^hH^*,\tag{4.14}$$

which are constant throughout the entire plate.

If the hole is assumed to be free of traction, the boundary conditions for this problem can be written as

$$N_{ij}\to\hat{N}_{ij},\quad M_{ij}\to\hat{M}_{ij},\quad i,j=1,2,\quad \text{at infinity},$$
$$N_n=N_{ns}=0,\quad M_n=V_n=0,\quad \text{along the hole boundary}.\tag{4.15}$$

As stated in the last paragraph of Sec. 3, in Eq. (4.15) there is no need to describe the boundary conditions of heat and moisture flow as those shown in Eq. (4.2) for Case 1. The prescribed values $\hat{N}_{ij}$ and $\hat{M}_{ij}$ are the stress resultants and bending moments induced by the temperature and moisture which are related to T^0, H^0, T^*, and H^* by

$$\begin{Bmatrix}\hat{N}_{11}\\ \hat{N}_{12}\\ \hat{M}_{11}\\ \hat{M}_{12}\end{Bmatrix}=-\boldsymbol{\vartheta}_1,\quad \begin{Bmatrix}\hat{N}_{12}\\ \hat{N}_{22}\\ \hat{M}_{12}\\ \hat{M}_{22}\end{Bmatrix}=-\boldsymbol{\vartheta}_2,\tag{4.16}$$

where $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$ are given in Eq. (3.10*b*). Note that Eq. (4.16) is obtained by using Eq. (3.3) and (3.10) and knowing that no mechanical loading is applied in this problem.

From the relations (3.3), (3.7), and (4.16), the boundary conditions (4.15) can now be expressed in terms of the generalized stress function vector $\boldsymbol{\phi}_d$ as

$$\phi_d\to-x_1\boldsymbol{\vartheta}_2+x_2\boldsymbol{\vartheta}_1,\quad \text{at infinity},$$
$$\boldsymbol{\phi}_{d,s}=\mathbf{0},\quad \text{along the hole boundary}.\tag{4.17}$$

By comparing the general solution (3.10) and boundary condition (4.17) of this problem with those of nonhygrothermal problem solved in [9], without any further detailed derivation we can obtain directly the solution for this problem as

$$\mathbf{u}_d=\mathrm{Re}\{\mathbf{A}\langle\mathfrak{s}_\alpha^{-1}\rangle\mathbf{B}^{-1}(a\boldsymbol{\vartheta}_2-ib\boldsymbol{\vartheta}_1)\},$$
$$\boldsymbol{\phi}_d=\mathrm{Re}\{\mathbf{B}\langle\mathfrak{s}_\alpha^{-1}\rangle\mathbf{B}^{-1}(a\boldsymbol{\vartheta}_2-ib\boldsymbol{\vartheta}_1)\}-x_1\boldsymbol{\vartheta}_2+x_2\boldsymbol{\vartheta}_1,\tag{4.18a}$$

where

$$\mathfrak{s}_\alpha=\frac{z_\alpha+\sqrt{z_\alpha^2-a^2-\mu_\alpha^2b^2}}{a-i\mu_\alpha b},\quad \alpha=1,2,3.\tag{4.18b}$$

Again, from Eq. (3.7) we see that the real form solution of the forces and moments around the hole boundary can be obtained directly if we have the solution for $\boldsymbol{\phi}_{d,n}$. By referring to the corresponding nonhygrothermal solutions obtained in [9], without any further detailed derivation we get

$$\boldsymbol{\phi}_{d,n}=\cos\theta\left[\boldsymbol{\vartheta}_1+\mathbf{G}_1(\theta)\boldsymbol{\vartheta}_2+\frac{b}{a}\mathbf{G}_3(\theta)\boldsymbol{\vartheta}_1\right]+\sin\theta\left[\boldsymbol{\vartheta}_2-\mathbf{G}_1(\theta)\boldsymbol{\vartheta}_1+\frac{a}{b}\mathbf{G}_3(\theta)\boldsymbol{\vartheta}_2\right],\tag{4.19}$$

where $\mathbf{G}_3(\theta)$ is defined in Eq. (4.12*b*) and

$$\mathbf{G}_1(\theta)=\mathbf{N}_1^T(\theta)-\mathbf{N}_3(\theta)\mathbf{S}\mathbf{L}^{-1}.\tag{4.20}$$

5 Numerical Examples

To show the generality of the analytical solutions obtained in the last section, in the following examples several different kinds of laminations, symmetric or unsymmetric, are considered. Without losing generality for simplicity all of them are composed of four layers of graphite-epoxy fiber reinforced composite laminae. The laminate contains a through-thickness elliptical hole and the material properties of graphite-epoxy are

$$E_1=181\ \text{GPa},\quad E_2=10.3\ \text{GPa},\quad G_{12}=7.17\ \text{GPa},\quad \nu_{12}=0.28,$$
$$\alpha_{11}^t=0.02\times10^{-6}/^\circ\text{C},\quad \alpha_{22}^t=22.5\times10^{-6}/^\circ\text{C},$$
$$k_{11}^t=1.5\ \text{W/m}^\circ\text{C},\quad k_{22}^t=k_{33}^t=0.5\ \text{W/m}^\circ\text{C},$$

where E_1 and E_2 are the Young's moduli in fiber and its transverse directions, respectively; G_{12} is the shear modulus in the x_1x_2 plane; ν_{12} is the major Poisson's ratio; α_{11}^t and α_{22}^t are the coefficients of thermal expansion in fiber and its transverse directions, respectively; k_{ii}^t, $i=1,2,3$, are the coefficients of heat conduction. All the other values of α_{ks}^t, k_{ij}^t, $i\neq j$, are zero. Each lamina thickness is 1 mm.

Based upon the material properties of graphite epoxy given above for the 1–2 coordinate system (1 and 2, respectively, stand for the fiber and transverse directions of each layer), the material properties in the x_1-x_2 coordinate system for each layer such as C_{ijks}, α_{ks}^t, k_{ij}^t shown in the basic Eqs. (2.1), can be calculated through the transformation equations [4]. With these values, the sectional properties A_{ijkl}, B_{ijkl}, D_{ijkl}, A_{ij}^t, B_{ij}^t, D_{ij}^t, K_{ij}^t, and K_{ij}^{*t} can then be calculated through the definitions given in Eq. (2.4). Note that in the above we did not list the moisture properties because for hygrothermal similarity only thermal problems are illustrated in the following examples.

5.1 Example 1: Uniform Temperature Changes in a Unidirectional Laminate. For the purpose of verification, we first consider the simplest case that only in-plane stretching occurs under uniform temperature changes in a unidirectional laminate. Thus, our solutions can even be compared by using only plane elements. Consider a $[0]_4$ unidirectional laminate containing a

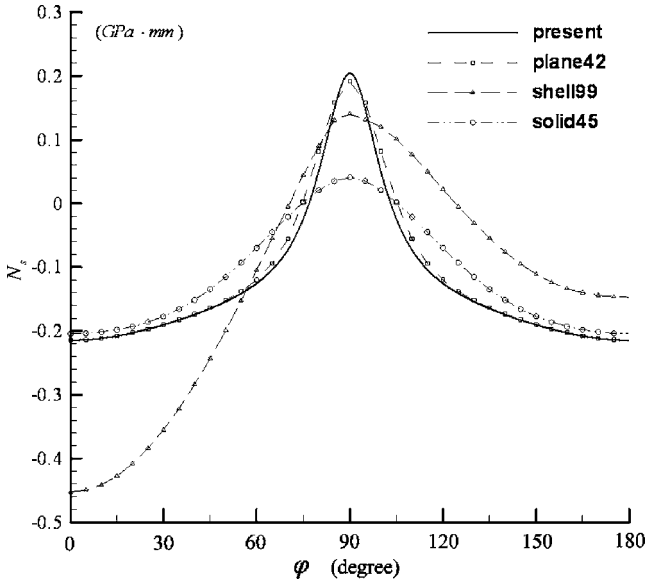


Fig. 2 Resultant hoop stress N_s around the circular hole boundary when the laminate is subjected to uniform temperature change

through-thickness circular hole subjected to temperature change from 0 to 100°C for the entire laminate, i.e., $T^0=100°C$ and $T^*=0$ in Eq. (3.8). Figure 2 shows the comparison of resultant hoop stress N_s around the hole boundary obtained by our solution and three different element types PLANE42, SHELL99, and SOLID45 in the commercial finite element software package ANSYS. In the ANSYS simulation, we use 72 nodes around the hole boundary and 1:100 hole/plate ratio to approximate the unbounded laminate. From Fig. 2, we see that the results of PLANE42 are well agreed to our solution while larger discrepancy occurs for the other element types. To find the reason, the numerical values of resultant tractions N_n/N_s and N_{ns}/N_s around the hole boundary, which should vanish for a traction-free hole, are plotted in Figs. 3(*a*) and 3(*b*). These two figures give us strong evidence why larger discrepancy occurs for SHELL99 and SOLID45 since they did not satisfy the traction-free hole boundary condition, which is satisfied by PLANE42 approximately and by our solution exactly.

Because the element type PLANE42 can only be used for in-plane problems, for general stretching-bending coupling problems one can only choose shell or solid element. From Figs. 2 and 3, it is expected that larger discrepancy between our results and those of ANSYS will occur for general unsymmetrtic laminates. That is to say, due to the approximate nature of finite element software, to avoid providing wrong solutions it is important to have a good reference such as the exact solutions for unbounded plates obtained in this paper.

5.2 Example 2: Uniform Heat Flow in x_1-x_2 Plane. In this example, we consider an unsymmetric laminate [+45/−45/+45/0]. Figure 4 is the plot of resultant hoop stress N_s around the circular hole boundary when only heat $\hat{q}_2=\hat{q}$ flows through the laminate in the x_2 direction. The three-diemnsional SOLID70 and SOLID45 element type of ANSYS that with 72 nodes around the hole boundary and 1:100 hole/plate ratio to approximate the unbounded laminate is used to compare our results. As expected the results of ANSYS shown in Fig. 4 do not agree well with our solutions. Like the discussion of Example 1, this discrepancy may come from the fact that our solution satisfies the hole boundary condition exactly, which can only be satisfied approximately by

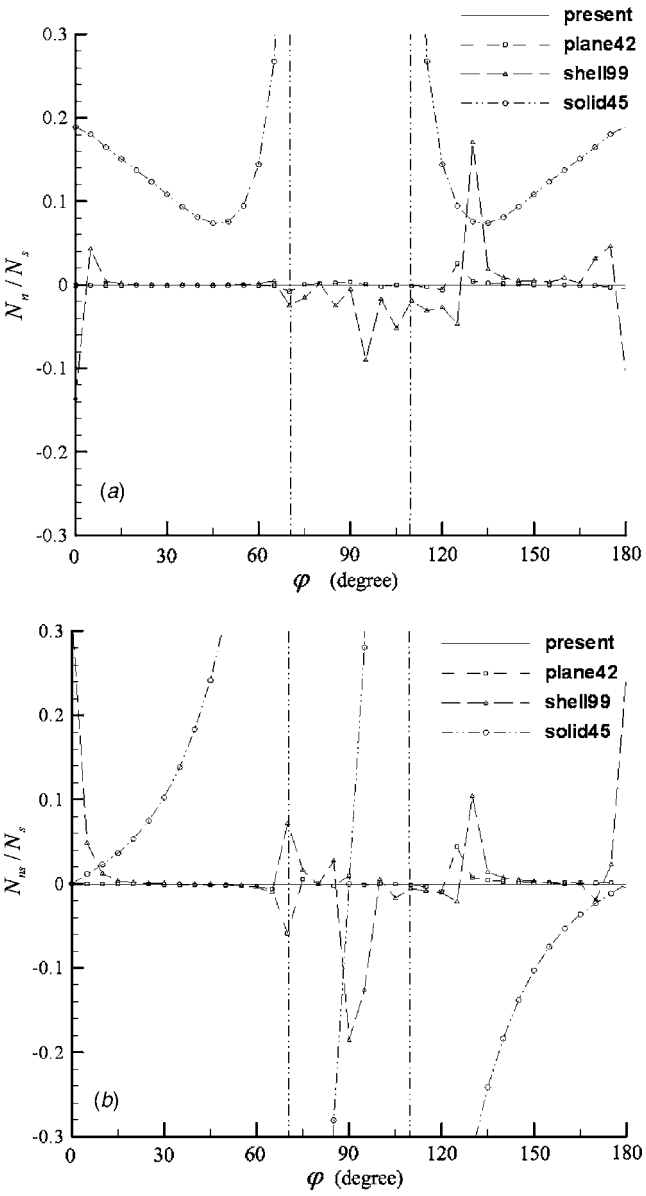


Fig. 3 Resultant tractions around the hole boundary when the laminate is subjected to uniform temperature changes (a) N_n/N_s (b) N_{ns}/N_s

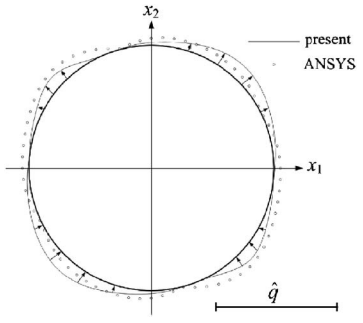


Fig. 4 Resultant hoop stress N_s around the circular hole boundary when the laminate is subjected to uniform heat flow $\hat{q}_2=\hat{q}$ in the positive x_2 direction

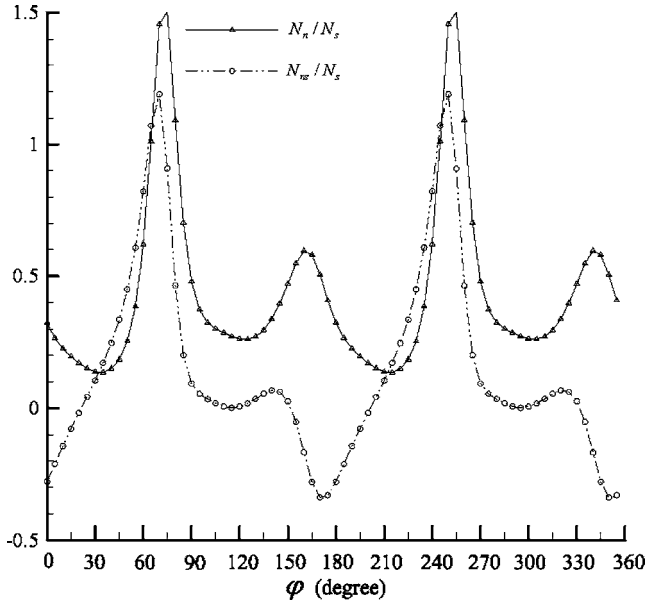


Fig. 5 Resultant tractions around the hole boundary in ANSYS simulation when the laminate is subjected to uniform heat flow $\hat{q}_2=\hat{q}$ in the positive x_2 direction

ANSYS. Moreover, our solution is for unbounded laminates, which is now approximated by 1:100 hole/plate ratio in ANSYS. Figure 5 shows the values of the resultant tractions N_n/N_s and N_{ns}/N_s around the hole boundary in ANSYS simulation which indicates clearly that they are still far from zero.

Since our results are not only valid for circular hole but also for elliptical hole, to see the shape effect of holes, we plot the distributions of resultant force N_s, moment M_s, and twist M_{ns} around the hole boundary for different ratios, $b/a=5,3,1,1/3,1/5$ in Figs. 6(*a*)–6(*c*). These figures are the plots of the considered laminate subjected to the heat flow $\hat{q}_1=\hat{q}$ in x_1-x_2 plane. From Figs. 6(*a*)–6(*c*), we see that when the ratio b/a is getting higher (i.e., the elliptical hole approaches to be a vertical crack perpendicular to the applied heat flow $\hat{q}_1=\hat{q}$), the maximum values of N_s, M_s, and M_{ns} will locate near $\varphi=90°$ or 270°, which is conformable to our engineering intuition. These results are reasonable and expectable.

Figure 7 shows the contour plot of temperature distribution near the hole when the laminates is subjected to uniform heat flow $\hat{q}_1=\hat{q}$ in x_1-x_2 plane. The temperature is nondimensionalized by $TK_t/a\hat{q}$. The uniform heat flow applied at infinity can be seen by the equidistant and parallel lines shown in this figure. However, the normal of these parallel lines is not necessarily the flow direction, which can be understood by first equation in the relation (3.1) if $K^t_{ij}\neq 0$ when $i\neq j$ ($K^t_{11}=4.5\times10^{-3}$ W/°C, $K^t_{12}=1\times10^{-3}$ W/°C, and $K^t_{22}=3.5\times10^{-3}$ W/°C).

5.3 Example 3: Uniform Heat Flow in x_3 Direction. Consider the unsymmetric laminate [+45/−45/+45/0] subjected to uniform heat flow in x_3 direction and keeps its entire upper and lower surfaces in a constant temperature $T_u=100$°C and $T_l=0$°C. Figure 8 compares the resultant hoop stress N_s around hole boundary calculated by Eq. (4.19) and ANSYS. As discussed in Example 2, due to the ill satisfaction of traction-free hole condition by ANSYS (see Fig. 9), the results of ANSYS are not well agreed with ours. Figure 10 shows the plot of resultant force N_s, moment M_s, and twist M_{ns} around the elliptical hole boundary. From Fig. 10, we see that the maximum values of N_s, M_s, and M_{ns} will locate near $\varphi=90°$ or 270° as the ratio b/a is getting higher and locate near $\varphi=0°$ or 180° as the ratio b/a is getting lower. The tendency of resultant forces and moments versus hole boundaryangle φ is similar to that of Example 2. It also shows that the hygrothermal changes of unsymmertic laminate will not only induce in-plane forces but also the bending moments around the hole boundary, which is well agreed to the engineering sense.

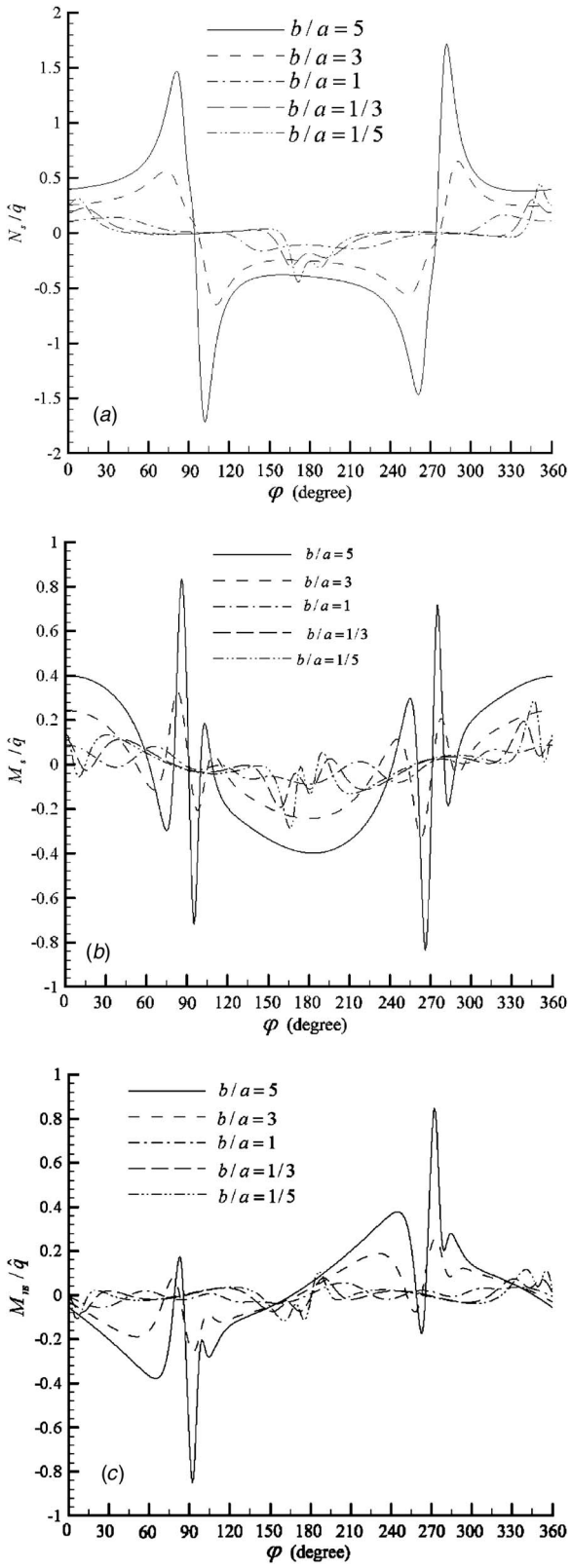


Fig. 6 Forces and moments around the elliptical hole in an unsymmetric laminate when the laminate is subjected to uniform heat flow $\hat{q}_1=\hat{q}$ in x_1-x_2 plane

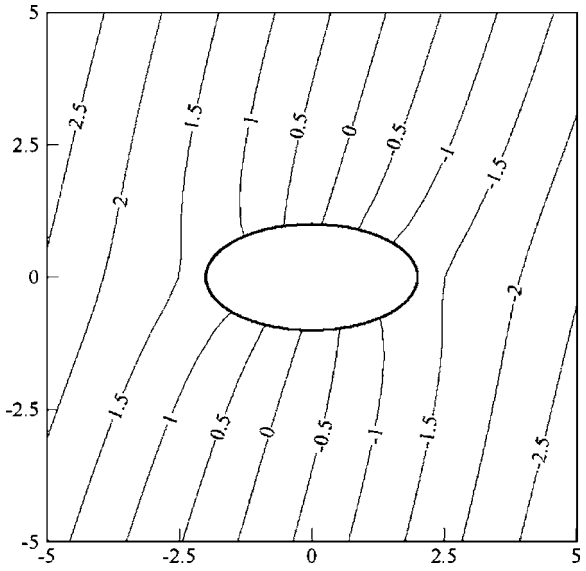


Fig. 7 Temperature distribution around the hole when the laminate is subjected to uniform heat flow $\hat{q}_1 = \hat{q}$ in x_1-x_2 plane

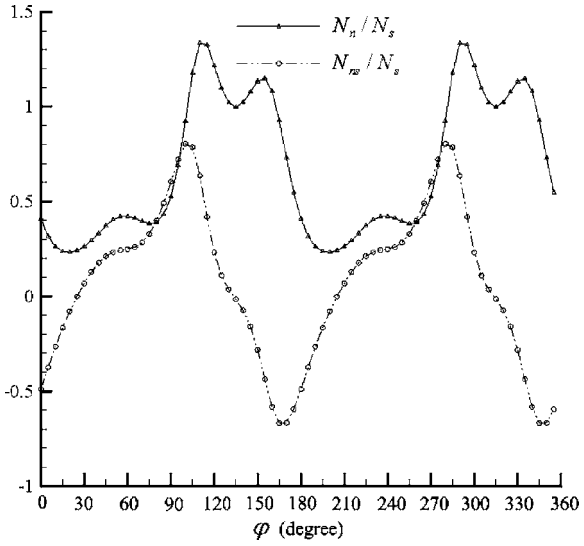


Fig. 9 Resultant tractions around the hole boundary in ANSYS simulation when the laminate is subjected to uniform heat flow in x_3 direction

6 Conclusions

In this work, the Stroh-like formalism for coupled stretching-bending analysis under hygrothermal environment is developed. By using this formalism, the solutions for an unbounded laminate, symmetric or unsymmetric, disturbed by an elliptical hole subjected to uniform heat flow and moisture transfer in the x_1-x_2 plane or the x_3 direction are obtained explicitly. To understand the hygrothermal effects on holes in general composite laminates, the explicit solutions of resultant forces and moments around the hole boundary are also obtained. Moreover, to show our solutions are simple, general, and exact, three numerical examples concerning the resultant forces and moments along the hole boundary subjected to different hygrothermal surrounding changes are presented and their comparisons with ANSYS finite element software package are also shown, which indicate the correctness and importance of our present solutions.

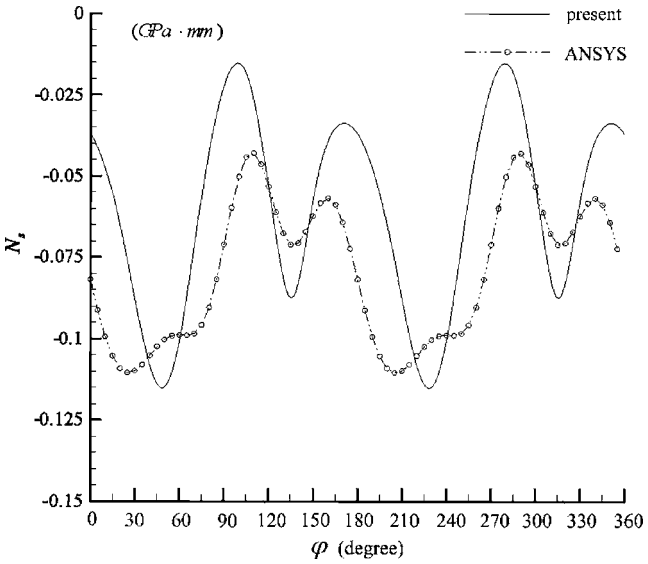


Fig. 8 Resultant hoop stress N_s around the circular hole boundary when the laminate is subjected to heat flow in x_3 direction

Acknowledgment

The author would like to thank the National Science Council, Taiwan, R.O.C. for support through Grant No. 92-2212-E-006-132.

Appendix A: Definition of the Fundamental Elasticity Matrix N for the Coupled Stretching-Bending Problems [6]

The 8×8 fundamental elasticity matrix $\mathbf{N}$ is defined as

$$\mathbf{N} = \left(\mathbf{L}_2 + \frac{1}{2}\mathbf{J}_2\right)^{-1}\left(\mathbf{L}_1 + \frac{1}{2}\mathbf{J}_1\right) = \mathbf{I}_t\mathbf{N}_m\mathbf{I}_t, \tag{A1a}$$

where

$$\mathbf{N} = \begin{bmatrix} \mathbf{N}_1 & \mathbf{N}_2 \\ \mathbf{N}_3 & \mathbf{N}_1^{\mathrm{T}} \end{bmatrix}, \quad \mathbf{N}_m = \begin{bmatrix} (\mathbf{N}_m)_1 & (\mathbf{N}_m)_2 \\ (\mathbf{N}_m)_3 & (\mathbf{N}_m)_1^{\mathrm{T}} \end{bmatrix}, \quad \mathbf{I}_t = \begin{bmatrix} \mathbf{I}_1 & \mathbf{I}_2 \\ \mathbf{I}_2 & \mathbf{I}_1 \end{bmatrix}, \tag{A1b}$$

$$\mathbf{L}_1 = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{R}^{\mathrm{T}} & -\mathbf{I} \end{bmatrix}, \quad \mathbf{L}_2 = -\begin{bmatrix} \mathbf{R} & \mathbf{I} \\ \mathbf{T} & \mathbf{0} \end{bmatrix},$$

$$\mathbf{J}_1 = \begin{bmatrix} -\mathbf{I}_{44} & -\mathbf{I}_{43} \\ \mathbf{I}_{34} & \mathbf{I}_{33} \end{bmatrix}, \quad \mathbf{J}_2 = \begin{bmatrix} -\mathbf{I}_{43} & \mathbf{I}_{44} \\ \mathbf{I}_{33} & -\mathbf{I}_{34} \end{bmatrix}. \tag{A1c}$$

and

$$(\mathbf{N}_m)_1 = -\mathbf{T}_m^{-1}\mathbf{R}_m^T, \quad (\mathbf{N}_m)_2 = \mathbf{T}_m^{-1} = (\mathbf{N}_m)_2^T,$$

$$(\mathbf{N}_m)_3 = \mathbf{R}_m\mathbf{T}_m^{-1}\mathbf{R}_m^T - \mathbf{Q}_m = (\mathbf{N}_m)_3^T, \tag{A1d}$$

$$\mathbf{I}_1 = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{I}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{A1e}$$

In the above, $\mathbf{I}$ denotes the identity matrix and $\mathbf{I}_{ij}$ stands for a matrix whose components are all zero except the ij component. Matrices $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{T}$ are three 4×4 real matrices which are related to the elastic constants and are the matrices defined for the displacement formalism developed in [6], while $\mathbf{Q}_m$, $\mathbf{R}_m$, and $\mathbf{T}_m$ are those defined for the mixed formalism [6]. They are defined by

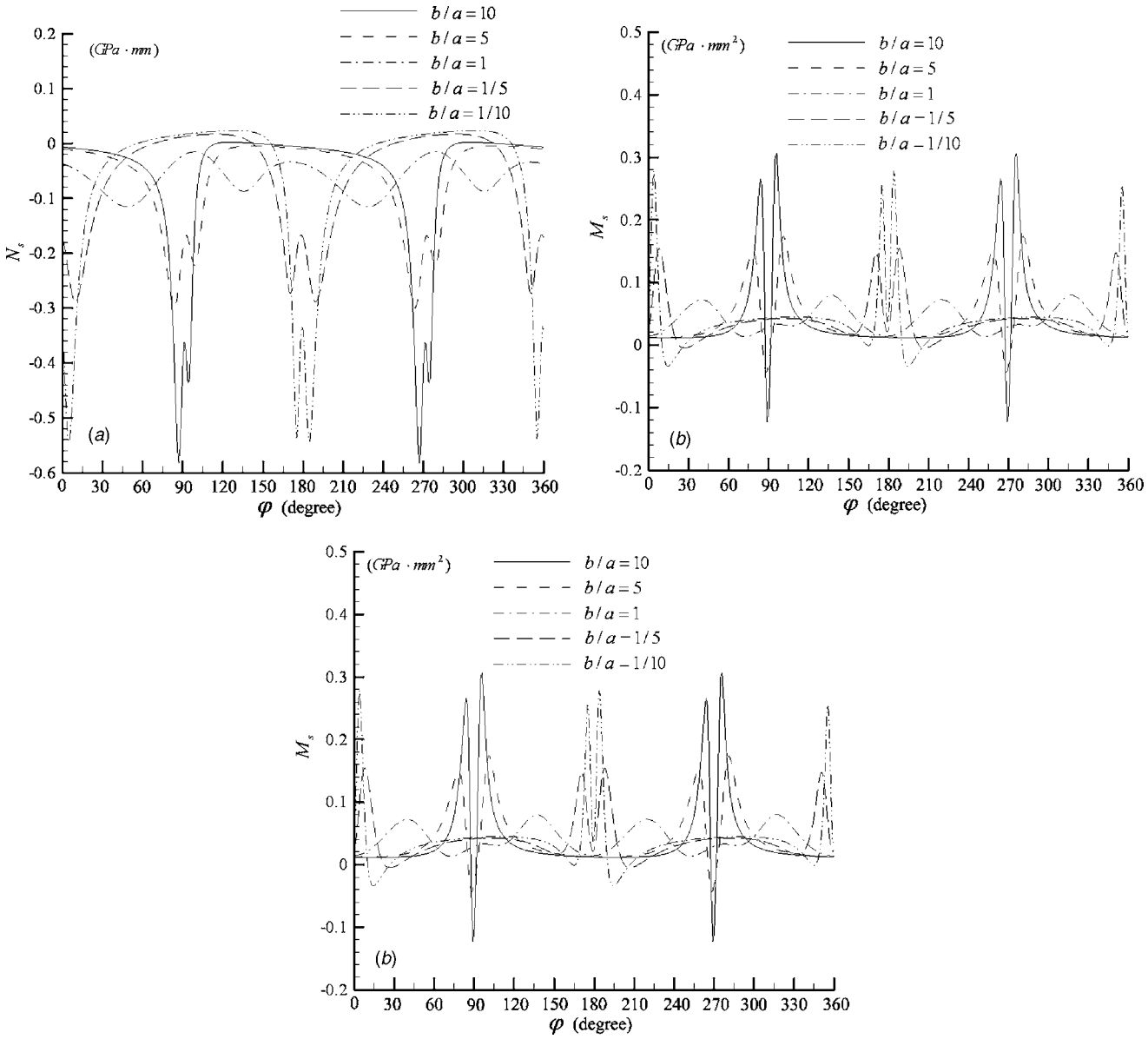


Fig. 10 Forces and moments around the elliptical hole in an unsymmetric laminate when the laminate is subjected to uniform heat flow in x_3 direction (T_u=100°C on the upper surface and T_l=0°C on the lower surface)

$$\mathbf{Q}=\begin{bmatrix}\mathbf{Q}_A & \mathbf{Q}_B\\ \mathbf{Q}_B & \mathbf{Q}_D\end{bmatrix},\quad \mathbf{R}=\begin{bmatrix}\mathbf{R}_A & \mathbf{R}_B\\ \mathbf{R}_B & \mathbf{R}_D\end{bmatrix},\quad \mathbf{T}=\begin{bmatrix}\mathbf{T}_A & \mathbf{T}_B\\ \mathbf{T}_B & \mathbf{T}_D\end{bmatrix},$$

$$\mathbf{T}_A=\begin{bmatrix}A_{66} & A_{26}\\ A_{26} & A_{22}\end{bmatrix},\quad \mathbf{T}_B=\begin{bmatrix}B_{66} & B_{26}\\ B_{26} & B_{22}\end{bmatrix},\quad \mathbf{T}_D=\begin{bmatrix}D_{66} & D_{26}\\ D_{26} & D_{22}\end{bmatrix}. \tag{A2b}$$

$$\mathbf{Q}_m=\begin{bmatrix}\mathbf{Q}_{\tilde{A}} & \mathbf{R}_{\tilde{B}}\\ \mathbf{R}_{\tilde{B}}^T & -\mathbf{T}_{\tilde{D}}\end{bmatrix},\quad \mathbf{R}_m=\begin{bmatrix}\mathbf{R}_{\tilde{A}} & -\mathbf{Q}_{\tilde{B}}\\ \mathbf{T}_{\tilde{B}}^T & \mathbf{R}_{\tilde{D}}^T\end{bmatrix},$$

$$\mathbf{T}_m=\begin{bmatrix}\mathbf{T}_{\tilde{A}} & -\tilde{\mathbf{R}}_{\tilde{B}}\\ -\tilde{\mathbf{R}}_{\tilde{B}}^T & -\mathbf{Q}_{\tilde{D}}\end{bmatrix}, \tag{A2a}$$

$$\mathbf{Q}_{\tilde{A}}=\begin{bmatrix}\tilde{A}_{11} & \tilde{A}_{16}\\ \tilde{A}_{16} & \tilde{A}_{66}\end{bmatrix},\quad \mathbf{Q}_{\tilde{B}}=\begin{bmatrix}\tilde{B}_{11} & \frac{1}{2}\tilde{B}_{16}\\ \tilde{B}_{61} & \frac{1}{2}\tilde{B}_{66}\end{bmatrix},$$

in which

$$\mathbf{Q}_A=\begin{bmatrix}A_{11} & A_{16}\\ A_{16} & A_{66}\end{bmatrix},\quad \mathbf{Q}_B=\begin{bmatrix}B_{11} & B_{16}\\ B_{16} & B_{66}\end{bmatrix},\quad \mathbf{Q}_D=\begin{bmatrix}D_{11} & D_{16}\\ D_{16} & D_{66}\end{bmatrix},$$

$$\mathbf{R}_A=\begin{bmatrix}A_{16} & A_{12}\\ A_{66} & A_{26}\end{bmatrix},\quad \mathbf{R}_B=\begin{bmatrix}B_{16} & B_{12}\\ B_{66} & B_{26}\end{bmatrix},\quad \mathbf{R}_D=\begin{bmatrix}D_{16} & D_{12}\\ D_{66} & D_{26}\end{bmatrix},$$

$$\mathbf{Q}_{\tilde{D}}=\begin{bmatrix}\tilde{D}_{11} & \frac{1}{2}\tilde{D}_{16}\\ \frac{1}{2}\tilde{D}_{16} & \frac{1}{4}\tilde{D}_{66}\end{bmatrix},$$

$$\mathbf{R}_{\tilde{A}}=\begin{bmatrix}\tilde{A}_{16} & \tilde{A}_{12}\\ \tilde{A}_{66} & \tilde{A}_{26}\end{bmatrix},\quad \mathbf{R}_{\tilde{B}}=\begin{bmatrix}\frac{1}{2}\tilde{B}_{16} & \tilde{B}_{12}\\ \frac{1}{2}\tilde{B}_{66} & \tilde{B}_{62}\end{bmatrix},$$

$$\mathbf{R}_{\tilde{D}}=\begin{bmatrix}\frac{1}{2}\tilde{D}_{16} & \tilde{D}_{12}\\ \frac{1}{4}\tilde{D}_{66} & \frac{1}{2}\tilde{D}_{26}\end{bmatrix},\quad \tilde{\mathbf{R}}_{\tilde{B}}=\begin{bmatrix}\tilde{B}_{61} & \frac{1}{2}\tilde{B}_{66}\\ \tilde{B}_{21} & \frac{1}{2}\tilde{B}_{26}\end{bmatrix}$$

$$\mathbf{T}_{\tilde{A}}=\begin{bmatrix}\tilde{A}_{66} & \tilde{A}_{26}\\ \tilde{A}_{26} & \tilde{A}_{22}\end{bmatrix},\quad \mathbf{T}_{\tilde{B}}=\begin{bmatrix}\frac{1}{2}\tilde{B}_{66} & \tilde{B}_{62}\\ \frac{1}{2}\tilde{B}_{26} & \tilde{B}_{22}\end{bmatrix},$$

$$\mathbf{T}_{\tilde{D}}=\begin{bmatrix}\frac{1}{4}\tilde{D}_{66} & \frac{1}{2}\tilde{D}_{26}\\ \frac{1}{2}\tilde{D}_{26} & \tilde{D}_{22}\end{bmatrix}. \tag{A2c}$$

A_{ij}, B_{ij}, and D_{ij}, are the extensional, coupling and bending stiffnesses, which are the contracted notations of A_{ijks}, B_{ijks}, and D_{ijks} defined in Eq. (2.4); $\tilde{A}_{ij}$, $\tilde{B}_{ij}$, and $\tilde{D}_{ij}$ are related to A_{ij}, B_{ij}, and D_{ij} by

$$\tilde{\boldsymbol{A}}=\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{B},\quad \tilde{\boldsymbol{B}}=\boldsymbol{B}\boldsymbol{D}^{-1},\quad \tilde{\boldsymbol{D}}=\boldsymbol{D}^{-1}. \tag{A3}$$

Note that in Eq. (A3), the symbols **A** and **B** have different representations from the eigenvector matrices **A** and **B** defined in Eq. (3.2*c*). The former is the traditional notation used in the community of mechanics of composite materials, while the latter is the notation generally used in the community of anisotropic elasticity. To let the readers from both communities see clearly what we express in this paper, we just use italic and roman fonts to distinguish these symbols.

Appendix B: Definition of Barnett-Lothe Tensors $\tilde{\mathbf{N}}$, S, H, and L [8]

S, **H**, and **L**, generally called *Barnett-Lothe tensors* in two-dimensional problems, are three real matrices defined as

$$\mathbf{S}=i(2\mathbf{A}\mathbf{B}^T-\mathbf{I}),\quad \mathbf{H}=2i\mathbf{A}\mathbf{A}^T,\quad \mathbf{L}=-2i\mathbf{B}\mathbf{B}^T. \tag{B1}$$

These three real matrices have also been proven to be the average fundamental submatrices $\tilde{\mathbf{N}}_1$, $\tilde{\mathbf{N}}_2$, and $-\tilde{\mathbf{N}}_3$, i.e.,

$$\mathbf{S}=\tilde{\mathbf{N}}_1=\frac{1}{\pi}\int_0^{\pi}\mathbf{N}_1(\omega)d\omega,$$

$$\mathbf{H}=\tilde{\mathbf{N}}_2=\frac{1}{\pi}\int_0^{\pi}\mathbf{N}_2(\omega)d\omega,$$

$$\mathbf{L}=-\tilde{\mathbf{N}}_3=-\frac{1}{\pi}\int_0^{\pi}\mathbf{N}_3(\omega)d\omega, \tag{B2}$$

in which $\mathbf{N}_1(\omega)$, $\mathbf{N}_2(\omega)$, and $\mathbf{N}_3(\omega)$ are the submatrices of the generalized fundamental matrix $\mathbf{N}(\omega)$ whose definition can be found in [8].

Since the Stroh-like formalism for the coupled stretching-bending problems has been purposely organized into the form of Stroh formalism for two-dimensional problems, the definitions given in Eqs. (B1) and (B2) usually used for the two-dimensional problems are also applicable for the coupled stretching-bending problems.

Appendix C: Definitions of Some Vectors Related to Thermal and Moisture Moduli

By following the steps of two-dimensional thermoelasticity developed by Hwu [3] and extended Stroh-like formalism described in Sec. 3, the thermal and moisture vectors γ_t and γ_h defined in Eq. (3.6) can be generalized to

$$\gamma_t(\omega)=\begin{Bmatrix}\boldsymbol{\gamma}_1^t(\omega)\\ \boldsymbol{\gamma}_2^t(\omega)\end{Bmatrix}=-\begin{bmatrix}\mathbf{I}_1 & \mathbf{I}_2\\ \mathbf{I}_2 & \mathbf{I}_1\end{bmatrix}\begin{bmatrix}\mathbf{0} & [\mathbf{N}_m(\omega)]_2\\ \mathbf{I} & [\mathbf{N}_m(\omega)]_1^T\end{bmatrix}\begin{Bmatrix}\tilde{\boldsymbol{\alpha}}_1^t(\omega)\\ \tilde{\boldsymbol{\alpha}}_2^t(\omega)\end{Bmatrix},$$

$$\gamma_h(\omega)=\begin{Bmatrix}\boldsymbol{\gamma}_1^h(\omega)\\ \boldsymbol{\gamma}_2^h(\omega)\end{Bmatrix}=-\begin{bmatrix}\mathbf{I}_1 & \mathbf{I}_2\\ \mathbf{I}_2 & \mathbf{I}_1\end{bmatrix}\begin{bmatrix}\mathbf{0} & [\mathbf{N}_m(\omega)]_2\\ \mathbf{I} & [\mathbf{N}_m(\omega)]_1^T\end{bmatrix}\begin{Bmatrix}\tilde{\boldsymbol{\alpha}}_1^h(\omega)\\ \tilde{\boldsymbol{\alpha}}_2^h(\omega)\end{Bmatrix}, \tag{C1a}$$

where

$$\tilde{\boldsymbol{\alpha}}_1^t(\omega)=\cos\omega\tilde{\boldsymbol{\alpha}}_1^t+\sin\omega\tilde{\boldsymbol{\alpha}}_2^t,\quad \tilde{\boldsymbol{\alpha}}_2^t(\omega)=-\sin\omega\tilde{\boldsymbol{\alpha}}_1^t+\cos\omega\tilde{\boldsymbol{\alpha}}_2^t,$$

$$\tilde{\boldsymbol{\alpha}}_1^h(\omega)=\cos\omega\tilde{\boldsymbol{\alpha}}_1^h+\sin\omega\tilde{\boldsymbol{\alpha}}_2^h,\quad \tilde{\boldsymbol{\alpha}}_2^t(\omega)=-\sin\omega\tilde{\boldsymbol{\alpha}}_1^h+\cos\omega\tilde{\boldsymbol{\alpha}}_2^h, \tag{C1b}$$

and $[\mathbf{N}_m(\omega)]_1$,$[\mathbf{N}_m(\omega)]_2$, and $[\mathbf{N}_m(\omega)]_3$ are the submatrices of the generalized fundamental matrices $\mathbf{N}_m(\omega)$ of the mixed formalism introduced in [6]. They are defined by

$$[\mathbf{N}_m(\omega)]_1=-\mathbf{T}_m^{-1}(\omega)\mathbf{R}_m^T(\omega),\quad [\mathbf{N}_m(\omega)]_2=\mathbf{T}_m^{-1}(\omega)=[\mathbf{N}_m(\omega)]_2^T,$$

$$[\mathbf{N}_m(\omega)]_3=\mathbf{R}_m(\omega)\mathbf{T}_m^{-1}(\omega)\mathbf{R}_m^T(\omega)-\mathbf{Q}_m(\omega), \tag{C2a}$$

where $\mathbf{Q}_m(\omega)$,$\mathbf{R}_m(\omega)$, and $\mathbf{T}_m(\omega)$ are related to $\mathbf{Q}_m$, $\mathbf{R}_m$, and $\mathbf{T}_m$ by

$$\mathbf{Q}_m(\omega)=\mathbf{Q}_m\cos^2\omega+(\mathbf{R}_m+\mathbf{R}_m^T)\sin\omega\cos\omega+\mathbf{T}_m\sin^2\omega,$$

$$\mathbf{R}_m(\omega)=\mathbf{R}_m\cos^2\omega+(\mathbf{T}_m-\mathbf{Q}_m)\sin\omega\cos\omega-\mathbf{R}_m^T\sin^2\omega,$$

$$\mathbf{T}_m(\omega)=\mathbf{T}_m\cos^2\omega-(\mathbf{R}_m+\mathbf{R}_m^T)\sin\omega\cos\omega+\mathbf{Q}_m\sin^2\omega. \tag{C2b}$$

The eigen-relations for the thermal and moisture properties have been shown in the second and third equations in Eq. (3.4*b*), which can be generalized as

$$\mathbf{N}(\omega)\eta_t(\omega)=\tau_t(\omega)\eta_t(\omega)+\gamma_t(\omega),$$

$$\mathbf{N}(\omega)\eta_h(\omega)=\tau_h(\omega)\eta_h(\omega)+\gamma_h(\omega), \tag{C3a}$$

where

$$\tau_t(\omega)=\frac{\tau_t\cos\omega-\sin\omega}{\tau_t\sin\omega+\cos\omega},\quad \tau_h(\omega)=\frac{\tau_h\cos\omega-\sin\omega}{\tau_h\sin\omega+\cos\omega},$$

$$\eta_t(\omega)=\chi_t(\omega)\eta_t,\quad \eta_h(\omega)=\chi_h(\omega)\eta_h, \tag{C3b}$$

and

$$\chi_t(\omega)=\cos\omega+\tau_t\sin\omega,\quad \chi_h(\omega)=\cos\omega+\tau_h\sin\omega. \tag{C3c}$$

Integrating the generalized eigenrelations (C3*a*) from 0 to π and using the definition (B2), we get

$$\begin{bmatrix}\mathbf{S} & \mathbf{H}\\ -\mathbf{L} & \mathbf{S}^T\end{bmatrix}\begin{Bmatrix}\mathbf{c}_t\\ \mathbf{d}_t\end{Bmatrix}=i\begin{Bmatrix}\mathbf{c}_t\\ \mathbf{d}_t\end{Bmatrix}+\begin{Bmatrix}\tilde{\boldsymbol{\gamma}}_1^{t^*}\\ \tilde{\boldsymbol{\gamma}}_2^{t^*}\end{Bmatrix},$$

$$\begin{bmatrix}\mathbf{S} & \mathbf{H}\\ -\mathbf{L} & \mathbf{S}^T\end{bmatrix}\begin{Bmatrix}\mathbf{c}_h\\ \mathbf{d}_h\end{Bmatrix}=i\begin{Bmatrix}\mathbf{c}_h\\ \mathbf{d}_h\end{Bmatrix}+\begin{Bmatrix}\tilde{\boldsymbol{\gamma}}_1^{h^*}\\ \tilde{\boldsymbol{\gamma}}_2^{h^*}\end{Bmatrix}, \tag{C4a}$$

where

$$\tilde{\gamma}_i^{t^*} = \frac{1}{\pi}\int_0^{\pi} \chi_t^{-1}(\omega)\gamma_i^t(\omega)d\omega,$$

$$\tilde{\gamma}_i^{h^*} = \frac{1}{\pi}\int_0^{\pi} \chi_h^{-1}(\omega)\gamma_i^h(\omega)d\omega, \quad i = 1,2. \tag{C4b}$$

References

[1] Savin, G. N., 1961, *Stress Concentration Around Holes*, Pergamon Press, London.

[2] Florence, A. L., and Goodier, J. N., 1959, "Thermal Stress at Spherical Cavities and Circular Holes in Uniform Heat Flow," ASME J. Appl. Mech., **26**, pp. 293–294.

[3] Hwu, C., 1990, "Thermal Stresses in an Anisotropic Plate Disturbed by an Insulated Elliptic Hole or Crack," ASME J. Appl. Mech., **57**, pp. 916–922.

[4] Jones, R. M., 1974, *Mechanics of Composite Materials*, Scripta, Washington, D.C.

[5] Herakovich, C. T., 1998, *Mechanics of Fibrous Composites*, Wiley, New York.

[6] Hwu, C., 2003, "Stroh-Like Formalism for the Coupled Stretching-Bending Analysis for Composite Laminates," Int. J. Solids Struct., **40**, pp. 3681–3705.

[7] Stroh, A. N., 1958, "Dislocations and Cracks in Anisotropic Elasticity," Philos. Mag., **7**, pp. 625–646.

[8] Ting, T. C. T., 1996, *Anisotropic Elasticity—Theory and Applications*, Oxford Science Publications, New York.

[9] Hsieh, M. C., and Hwu, C., 2003, "Explicit Solutions for the Coupled Stretching-Bending Problems of Holes in Composite Laminates," Int. J. Solids Struct., **40**, pp. 3913–3933.

[10] Nowacki, W., 1962, *Thermoelasticity*, Addison-Wesley, Reading, MA.

[11] Lin, C. C., and Hwu, C., 1993, "Uniform Heat Flow Disturbed by an Elliptical Rigid inclusion Embedded in an Anisotropic Elastic Matrix," J. Therm. Stresses, **16**, pp. 119–133.

[12] Hwu, C., 2003, "Stroh Formalism and its Extensions to Coupled Inplane-Bending Problems," Chin. J. Mech., Ser. A, Special Issue to Celebrate Prof. T.C.T. Ting's 70th Birthday, **19**, pp. 41–53.

[13] Hsieh, M. C., and Hwu, C., 2002, "Explicit Expressions of the Fundamental Elasticity Matrices of Stroh-Like Formalism for Symmetric/Unsymmetric Laminates," Chin. J. Mech., Ser. A, **18**, pp. 109–118.

Chaotic Phenomena Induced by the Fast Plastic Deformation of Metals During Cutting

Zoltan Palmai
COGITO Ltd.,
Budapest Alvinci ut 24,
Budapest 1022, Hungary
e-mail: palmay@axelero.hu

In the present study the examination of chip formation is focused on the primary shear zone, which is divided into two layers, and the variation of shear stress and temperature in time are given by two mechanical balance equations and three energy equations. All the five evolution differential equations are autonomous and nonlinear. The material characteristics are determined by an exponential constitutive equation. The mathematical model is suitable for the qualitative description of different types of chips, such as continuous chips and periodic or aperiodic shear localized chips, which is demonstrated by the general structure and typical solutions of the equation system.
[DOI: 10.1115/1.2074727]

1 Introduction

It is widely known that technologies based on the plastic formation of metals generally bring about high strain in materials, which often occurs rather rapidly. These processes are very difficult to describe by theoretical methods, even in cases where we know that deformation is largely concentrated on localized shear zones in a wide range of practical technologies. This can be well-illustrated by the example of cutting, in which the degree of shear strain is 2–5, the shear strain rate is about 10^4 s^{-1}, the magnitude of rise in temperature is 10^2 K and its rate of increase is 10^4–10^5 K/s in the so-called primary shear zone.

Under such circumstances, processes can only be described—even with simplifying assumptions—with nonlinear differential equations, which until recently could only be solved in exceptional cases. As a result of this, technological developments mainly relied on measurements, by means of which empirical correlations were established. It is inevitably useful for future research that a large inventory of experimental results have been collected so far with regard to specific technologies, such as those concerning regular, continuous chip formation in cutting. This data may prove useful for the theoretical discussion of plastic forming technologies, facilitated by the recent progress of computer technology.

Precise theoretical calculations are very important indeed, since—owing to the progress of technology—some characteristic trends of modern manufacturing processes impose new requirements for the exact description of processes:

- in order to reach higher productivity, *the intensity of processes must be increased*, while the *stability* of the cutting system must be preserved;
- in order to increase economic efficiency, more operations are often concentrated in a single cutting phase, so *the complexity of production processes is increasing*;
- the performance of new, *extreme technological tasks* has become necessary, for example, cutting with thin tools of a small diameter, or machining of hardened materials or special alloys etc.;
- for automation, the *regular inspection of technological operations* must be conducted, for example, we must monitor and evaluate the degradation (wear) of the tool and the chatter characteristics of the workpiece-machine-tool system, etc.;
- technology experts are facing increasingly strict requirements, including the demand for *high quality, dimensional accuracy and surface quality*.

As an inevitable consequence of these trends, the importance of modeling technological processes becomes more and more important, and today it is still the subject of intensive research, as reflected in contemporary scientific studies. The theoretical study of machining metals usually requires the examination of nonlinear processes, owing to the nature of these technologies:

(a) *The characteristics of the material* manufactured, which can generally be described by nonlinear functions;
(b) *the dynamic characteristics* of machines, equipment, tools and their complex system, which are often of a nonlinear nature;
(c) the nonlinear effects of control arising from the *feedback* in process control.

The above list also reflects the specific hierarchy of problems. The degrees of freedom increase from (a) to (c), and the behaviour of the system can only be described using an increasing number of "dimensions." The objective of the present study can be defined "on the lower level:" We aim to study the nonlinear phenomena arising from the characteristics of the workpiece material in cutting, in order to develop a model suitable for the description of machining technologies subject to fast plastic deformation in practice, which will also facilitate the practical study of so-called deterministic instability.

The examination of cutting can also prove fruitful, because both stable and unstable processes may be observed in this technology. Figure 1 shows austenitic steel chips, produced by machining at three different cutting speeds. In case (a) (the classical example of cutting) and in case (c), continuous chips and segmental chips were produced, respectively. Moreover, it is important that there are real technological situations in which the process becomes aperiodic, as is proven by case (b).

Both strain-hardening and softening resulting from increased temperature may occur in cutting characterized by massive deformation, and the process is so rapid that the deformation zone often behaves like a nearly adiabatic system. As a result of this, the deformation is localized in a narrow band, and a so-called ther-

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received September 23, 2004; final manuscript receivedwh July 7, 2005. Review conducted by A. Maniatty. 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.

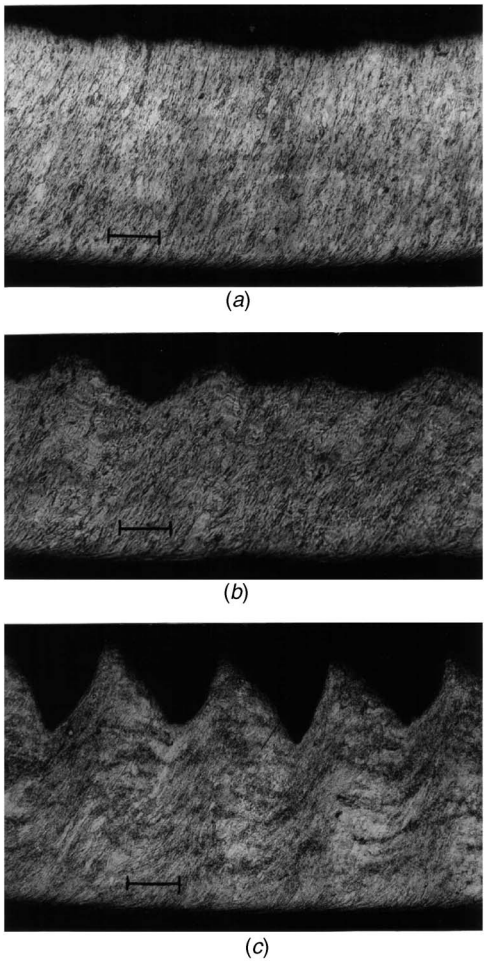


Fig. 1 Longitudinal cross sections of austenitic steel chips. (*a*) v=0.3 m/s, (*b*) *v*=0.58 m/s, (*c*) *v*=1.17 m/s)↔:0.1 mm.

moplastic shear instability can occur, as already shown by Zener and Hollomon [1]. This phenomenon—as Recht [2] proved—also emerges in the course of cutting, and shear-localized chips are formed.

The study of different types of chips goes back more than a hundred years, nevertheless, chips have still remained the focus of attention, which is perfectly shown by the historical overview of Davies and Burns [3]. The history of cutting Ti-alloys is a perfect example of this: Although Recht [2] already showed the relationship between shear instability and the formation of segmental chips, Komanduri and Hou [4] have recently published new research findings about the periodic segmental chip formation observed in Ti cutting. Our own experiments have shown that segmental discontinuous Ti chips can also be aperiodic. Roughly similar chip structure may be observed in the cutting of free cutting steel with low-carbon content, and we have some similar experiences in the field of cutting so-called invar alloys.

According to our view concerning the classification of various types of chips, chip formation is fundamentally either stable or unstable, with the latter being either periodic or aperiodic. Among practical technologies, aperiodic chip formation is a typical phenomenon of nonlinear dynamic processes, whose examination is the subject of "chaos theory." In the case of stable processes, evenly deformed chips of constant thickness, called continuous chips are produced (see Fig. 2(*a*)), which is significant because we have reliable knowledge about these processes owing to the several decades of intensive research into cutting.

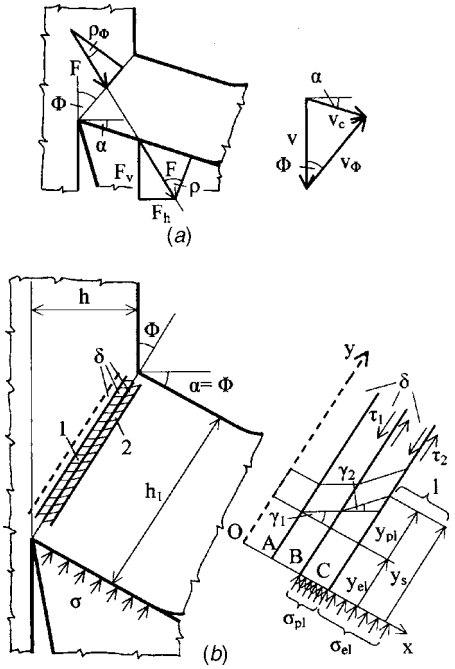


Fig. 2 The simplified model of chip formation

2 Basic Correlations of Cutting Theory and Initial Assumptions

The theory of continuous chip formation has long been elaborated and justified experimentally by many researchers. We know the characteristic parameters and the methods by means of which these parameters can be determined, either by calculation or measurement (see, for example, Shaw [5]).

With the symbols of Fig. 2(*a*), shear strain ε_Φ in the shear zone can be established by the following formula:

$$\varepsilon_\Phi = \tan(\Phi - \alpha) + \cot \Phi \tag{1}$$

where α is the rake angle of the tool, Φ is the angle of the so-called shear plane. The v_c velocity of the chip is

$$v_c = v.\sin \Phi/\cos(\Phi - \alpha) \tag{2}$$

Due to the high degree of deformation, it is important to note that ε_Φ is the tangent of the angle of distortion of the material in the shear zone.

The average velocity of shear strain is

$$\frac{d\gamma_{\text{avg}}}{dt} = \dot{\varepsilon}_\Phi = \frac{v_s}{\delta} = \frac{v}{2\delta}\frac{\cos \alpha}{\cos(\Phi - \alpha)}, \tag{3}$$

which is the temporal change of the tangent of the angle of distortion.

Finally, an additional important factor is shear stress τ_Φ, which is an average value that can be determined from the components of the cutting force for the whole deformation zone as follows:

$$\tau_\Phi = \frac{F}{q} \sin \Phi \,.\, \sin \varsigma_\Phi, \tag{4}$$

where q is the original cross section of the chip's material, $\varsigma_\Phi = \pi/2 - \Phi - \alpha + \varsigma$, and ς can be determined by dynamometry (Fig. 2(*a*)).

The main objective of cutting theory has long been to establish a useful correlation between Φ, ς, α, and a certain material strength factor (τ_Φ, etc.). This is an issue that scientists still try to resolve, as shown by the research conducted by Bayard, who has collected 49 different kinds of models to support his findings [6]. We do not wish to discuss these static models. We will only need

formulas (1)–(4). Naturally, these features of the shear zone are average values, which characterize the shear strain of the material going through the primary deformation layer. They may be regarded as a reference point or value when calculating actual characteristics changing in time.

We need to establish the mathematical model of the deformation process in order to describe chaotic phenomena occurring during chip formation. It is widely known that chip formation is the result of complex phenomena, therefore, we must make simplified assumptions acceptable for the theoretical examination of the topic. In the present case, we will apply the following thermomechanical model for the description of chip formation:

(a) The processes can be described with satisfactory precision in the plane ("free cutting").

(b) The examination is limited to the so-called primary strain zone, i.e., the shear zone, which is surrounded by parallel planes forming an angle of Φ with the cutting speed, and we assume that Φ=constant.

(c) Inertial forces and elastic deformation in energy balance are ignored.

(d) The shear zone is composed of 1+2 layers of equal thickness δ. In the first, or "quasi" layer, there is no plastic deformation, nevertheless, this quasi layer plays an important role in thermal processes. The plastic deformation that leads to the formation of chips occurs in the other two layers. In Fig. 2(*b*), the quasi layer is situated at $0 \leqslant x \leqslant x_A = \delta$ on axis x, while temperature T_0 is higher than temperature T_w of the workpiece. In the first deformation layer of temperature T_1 $x_A = \delta < x \leqslant x_B = 2\delta$, while in the second deformation layer of temperature T_2 $x_B = 2\delta < x \leqslant x_C = 3\delta$. Temperature is the same everywhere within layers of thickness δ, it only changes as a function of time.

(e) The rake angle of the tool equals the angle of the shear plane: $\alpha = \Phi$ (Fig. 2(*b*)). Thus

$$\varepsilon_\Phi = \cot \Phi, \tag{5a}$$

$$v_c = v.\sin \Phi, \tag{5b}$$

$$\dot{\varepsilon}_\Phi = \frac{v}{\delta} \cos \Phi. \tag{5c}$$

(f) Stress conditions in chip formation are illustrated by Fig. 2(*b*). Applying the method of Burns and Davies [7], we can assume that elastic compressive stress σ_{el} of even distribution occurs on the tool face on length l, and this keeps balance at point $x = x_C$ with shear stress τ developing in the shear zone, which is in a plastic state. As for mechanical balance at point $x = x_B$, we also have to consider stress σ_{pl} affecting the second layer being in a plastic state.

(g) Let us assume that a significant part of the mechanical work (L_{pl}) appears as heat in plastic deformation, the change of internal energy is insignificant, and the system can be regarded as nearly adiabatic. In accordance with Theorem I of thermodynamics, the change in heat content is $\mathrm{d}q \approx r.\mathrm{d}L_{\text{pl}}$, where $r \approx 0.95-0.98$. With the symbols of Fig. 2(*b*), we can establish

$$\frac{dq_{Li}}{dt} = \dot{q}_{Li} = r\tau_i\dot{\gamma}_i \quad (i = 1,2), \tag{6}$$

where τ_i is the shear stress and $\dot{\gamma}_i$ stands for the velocity of shear strain in the deformation layers.

(h) Thermal flow is one-dimensional, with thermal conduction only occurring on the boundary surfaces of the layers. Here we can use the simplified form of the differential equation of thermal conduction

$$\frac{dq_i}{dt} = \dot{q}_i \approx \frac{4\lambda}{\delta^2}(T_i - T_{i-1}), \quad i = 0,1,2 \tag{7}$$

where $\dot{q}_i = c\varsigma\dot{T}_i$ and in the quasi layer (i=0) $T_{i-1} = T_w$, which is the temperature of the workpiece.

(i) The material progressing in direction x with velocity $v_c = v.\sin\Phi$ conducts heat itself, which is

$$\dot{q}_{vi} = c\varsigma(T_i - T_w).v.\sin\Phi \quad i = 0,1,2 \tag{8}$$

3 The Constitutive Equation

A constitutive equation that describes the behavior of the material under the extreme strain conditions of the shear layer is essential for the mathematical model of chip formation. Generally, this would be $f(\varepsilon, \dot{\varepsilon}, \tau, T) = 0$, however, we will disregard strain-hardening in the first approach. Thus, after rearrangement, the equation becomes $\dot{\varepsilon} = f_1(\tau, T)$.

We use the term "velocity-modified temperature" to establish this function, which has already been used and confirmed experimentally by MacGregor and Fisher [8]. According to them, the correlation between temperature T of the material and deformation velocity $\dot{\varepsilon}$ during plastic deformation can be determined from the following empirical formula:

$$\frac{T_m}{T} = 1 - k.\ln\frac{\dot{\varepsilon}}{\dot{\varepsilon}_{\text{ref}}}\frac{T_{\text{ref}}}{T} \tag{9}$$

Here T_m refers to velocity-modified temperature, T_{ref} and $\dot{\varepsilon}_{\text{ref}}$ are reference data determined in familiar conditions and k is a constant characteristic of the material, which is—according to their measurements—$k = 0.008-0.045$.

Furthermore, we assume that the material becomes softer during temperature increase within the temperature range which is characteristic of the shear zone, and plastic shear stress τ can be described by equation

$$\tau - \tau_\Phi \cong \nu(T - T_\Phi) \tag{10}$$

with satisfactory precision, where ν is a constant characteristic of the material's thermal softening. Formula (4) defines average shear stress τ_Φ, where T is the actual temperature, while T_Φ refers to the estimated value of the shear zone temperature, which has been determined by several different methods (e.g., Ref. [5]), and concrete calculations are also available [9]. Thus, in this case $T_{\text{ref}} = T_\Phi$ and $\dot{\gamma}_{\text{ref}} = \dot{\varepsilon}_\Phi$, which can be calculated by formula (3). The constitutive equation can be established by connecting Eqs. (9) and (10). It is advisable to study our subject without dimensions, introducing the following variables

$$\hat{T} = \frac{T - T_w}{T_w}, \quad \hat{T}_\Phi = \frac{T_\Phi - T_w}{T_w} = C, \quad \hat{\tau} = \frac{\tau}{\tau_\Phi}, \tag{11}$$

then, after contraction and rearrangement, discarding the special symbols showing their dimensionless nature, we can establish

$$\frac{\dot{\varepsilon}}{\dot{\varepsilon}_\Phi} = \frac{T+1}{C+1}.\exp\frac{\tau - 1 + a\cdot(T - C)}{b(T+1)} = F(\tau, T) \tag{12}$$

$$\text{where } a = \nu.T_w/\tau_\Phi, \quad b = k.a \tag{13}$$

Constitutive Eq. (12) is similar to the results derived from an Arrhenius-type function in a different way by Kubin et al. [10], and specifically includes the constitutive equation of Burns [11].

4 The Mathematical Model of Chip Formation

On the basis of assumptions (a)–(i), two equations of mechanical balance and three energy equations can be established. In addition to constitutive Eq. (12), these equations constitute the mathematical model of chip formation applied in the present study.

As a result of elastic tension σ_{el} occurring in range $x_C \leqslant x \leqslant x_C + l$ of the tool face (see Fig. 2(*b*)), the chip of h_1 thickness gets compressed in direction y. The displacement of point C becomes

$$y_{el} = h_1.\varepsilon_y = h_1 \frac{\sigma_{el}}{\sin \Phi} = \frac{h}{\sin \Phi} \frac{\sigma_{el}}{E} \tag{14}$$

Plastic forming also occurs in range $x_B < x \leqslant x_C$, which can be expressed with shear strain γ_1, γ_2 as

$$y_{pl} = \delta(\tan \gamma_1 + \tan \gamma_2). \tag{15}$$

Thus, the total displacement can be written as

$$y_{tool} = y_{el} + y_{pl} = \frac{h}{\sin \Phi} \frac{\sigma_{el}}{E} + \delta(\tan \gamma_1 + \tan \gamma_2) \tag{16}$$

After differentiation with respect to time and rearrangement, the equation becomes

$$\dot{\varepsilon}_{el} = \frac{E}{h_1}[\dot{y}_{tool} - \delta(\dot{\varepsilon}_{xy1} + \dot{\varepsilon}_{xy2})] \tag{17}$$

where

$$\dot{y}_{tool} = v_\Phi = v.\ \cos \Phi \tag{18a}$$

$$\dot{\varepsilon}_{x,yi} = \frac{d}{dt}(\tan \gamma_i) \quad i = 1,2 \tag{18b}$$

Using the average deformation shear velocity Eq. (3), after rearrangement from Eqs. (16)–(18), we get

$$\dot{\sigma}_{el} = \frac{E}{h_1} \delta.\dot{\varepsilon}_\Phi \left[1 - \left(\frac{\varepsilon_{x,y1}}{\dot{\varepsilon}_\Phi} + \frac{\dot{\varepsilon}_{x,y2}}{\dot{\varepsilon}_\Phi} \right) \right]. \tag{19}$$

It follows from the mechanical equilibrium that $\sigma_{el}.l = \tau_2.h_1$ in the environment of A, thus

$$\dot{\tau}_2 = \frac{E.l}{h^2} v.\ \sin^2 \Phi.\ \cos \Phi \left[1 - \left(\frac{\dot{\varepsilon}_{xy1}}{\dot{\varepsilon}_\Phi} + \frac{\dot{\varepsilon}_{xy2}}{\dot{\varepsilon}_\Phi} \right) \right]. \tag{20}$$

On the boundary of the two deformed layers, the equilibrium is modified at point B on Fig. 2(*b*) and

$$\sigma_{el}.l + \sigma_{pl}.\ \delta = \tau_1 h_1, \tag{21}$$

thus

$$\tau_1 = \sigma_{el} \frac{l}{h_1} + \sigma_{pl} \frac{\delta}{h_1}. \tag{22}$$

It is well-known that in a plastic state we can say that $\sigma_{pl} \approx \sqrt{3}\tau_{pl}$, and here $\tau_{pl} = \tau_2$, which is the shear stress occurring in the second layer of δ thickness, exposed to shear strain, therefore, out of Eq. (22) we can establish

$$\dot{\tau}_1 = \sigma_{el} \frac{l}{h_1} + \dot{\tau}_2 \frac{\sqrt{3}\delta}{h_1}. \tag{22a}$$

Finally, using Eqs. (3), (19), and (20), after contraction, the equation becomes

$$\dot{\tau}_1 = p.\dot{\tau}_2, \tag{23}$$

where

$$p = 1 + \frac{\sqrt{3}\delta}{h} \sin \Phi. \tag{24}$$

Let us also introduce dimensionless form for time

$$\hat{t} = \frac{t}{K}, \tag{25}$$

then using formula (11) and constitutive equation (12), the equation becomes

$$\frac{d\hat{\tau}}{d\hat{t}} \frac{\tau_\Phi}{K} = \frac{E.l}{h^2} v.\sin^2 \Phi.\cos \Phi[1 - (F_1 + F_2)]. \tag{26}$$

Finally, introducing coefficient

$$K = \frac{\tau_\Phi h^2}{E.l.v.\sin^2 \Phi.\cos \Phi} \tag{27}$$

the dimensionless form of Eqs. (20) and (23), omitting its individual sign, we get

$$\dot{\tau}_1 = p.[1 - (F_1 + F_2)], \tag{28a}$$

$$\dot{\tau}_2 = 1 - (F_1 + F_2). \tag{28b}$$

The variation of the material's heat content ($c\rho\dot{T}$) at chip formation is composed of three elements:

- mechanical power in accordance with condition (g),
- thermal conduction on the basis of assumption (h),
- heat transferred by the moving material according to assumption (i).

On the basis of this, the energy equation in the quasi layer is

$$c\rho\dot{T}_0 = \frac{4\lambda}{\delta^2}[(T_1 - T_0) - (T_0 - T_w)] - c\rho(T_0 - T_w)v.\sin \Phi, \tag{29a}$$

while in the two deformation layers

$$c\rho\dot{T}_1 = r\tau_1\dot{\varepsilon}_{xy1} + \frac{4\lambda}{\delta^2}[(T_2 - T_1) - (T_1 - T_0)] - c\rho(T_1 - T_0)v.\sin \Phi \tag{29b}$$

and

$$c\rho\dot{T}_2 = r\tau_2\dot{\varepsilon}_{xy2} + \frac{4\lambda}{\delta^2}(T_1 - T_2) - c\rho(T_2 - T_1)v.\sin \Phi. \tag{29c}$$

By using constitutive Eq. (12), $\dot{\varepsilon}_{xyi} = \dot{\varepsilon}_\Phi.F_i(\tau_i, T_i)$, where $i = 1, 2$. It is also advisable to use dimensionless forms here, therefore, based on Eq. (11), we can establish

$$\hat{T}_w = 0, \quad \hat{T}_i = \frac{T_i - T_w}{T_w}, \quad i = 0,1,2 \tag{30}$$

After substitution and rearrangement, the three dimensionless energy equations, can also be written without specific signs as

$$\dot{T}_0 = \varsigma(T_1 - 2T_0) - \xi.T_0 \tag{31a}$$

$$\dot{T}_1 = \eta\tau_1 F_1(\tau_1, T_1) - \varsigma(2T_1 - T_2 - T_0) - \xi(T_1 - T_0) \tag{31b}$$

$$\dot{T}_2 = \eta\tau_2 F_2(\tau_2, T_2) - (\xi + \varsigma)(T_2 - T_1), \tag{31c}$$

where

$$\eta = \frac{r.K.\tau_\Phi.v}{c\rho T_w \delta} \cos \Phi, \quad \xi = \frac{K.v}{\delta} \sin \Phi, \quad \varsigma = \frac{4K.\lambda}{c\rho\delta^2} \tag{32}$$

So the mathematical model of chip formation consists of evolution equations (28*a*) and (28*b*) and Eq. (31*a*)–(31*c*), to which constitutive equation (13) is connected. This autonomous differential equation system, comprising five equations, can be solved easily by means of the MATHCAD program.

5 Discussion

In order to establish under what conditions we can expect the occurrence of chaotic chip formation, we examined the structure of the mathematical model's solutions. In spite of the fact that assumptions (a)–(i) considerably simplify the real conditions of chip formation, the thermomechanic model still includes several parameters:

- we applied parameters τ_Φ, c, E, ρ, λ, ν, T_Φ, and k to characterize the material,
- it was necessary to define the machining with parameters h, v, α, Φ, and T_w,
- r, l, and δ also appeared in the model.

These 17 parameters are concentrated in seven dimensionless coefficients (p, a, b, K, ξ, η, and ζ) by the mathematical model, meaning that the characteristics of the mathematical model should be examined in this seven-dimensional context, which is a rather complicated task. The examination of the physical and technological meaning of these parameters may offer some help in reducing the range of variable coefficients for the purposes of practical tasks. Coefficients a and b of the constitutive equation, which are constant in the case of a specific workpiece, serve the same purpose. The p expresses the geometrical conditions of chip formation, therefore, it can often be considered constant as well. Time scale K, however, is a complex function, since it includes material characteristics, technological parameters and model constants. Although it can be used as fixed data, we have to be aware of its complex nature. In accordance with formula (32), coefficient ξ is the dimensionless velocity of the chip, which is an important technological parameter. The amount of heat produced during plastic deformation is described by η, but it is independent of the technology according to the model. Finally, ζ is related to both thermal conduction and technology.

Taking all these into consideration, the principal characteristics of the mathematical model on plane $\xi-\zeta$ are shown by Fig. 3, in addition to specifying the other coefficients of a concrete calculation ($p=1.03$, $a=0.3$, $b=0.012$, $C=1$, and $\eta=4.5$).

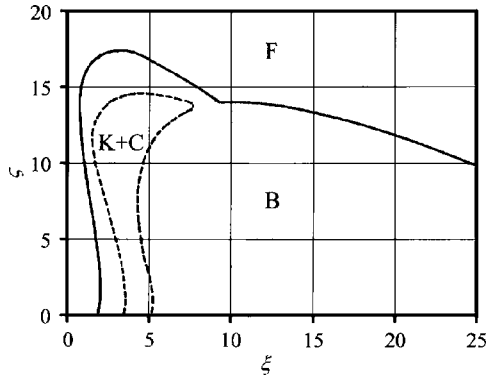


Fig. 3 The structure of the mathematical model's solutions. *F*: Fixed point, *B*: Bifurcation, *K+C*: Chaotic and *n*-cyclic ($n>1$), $p=1.03$, $\eta=4.5$, $a=0.3$, $b=0.012$, and $C=1$.

Following the standard method of linear stability analysis, it is recommended to calculate the boundary points determined by equations

$$\dot{\tau}_1 = 0, \quad \dot{\tau}_2 = 0 \quad \text{and } \dot{T}_i = 0, \quad i = 0,1,2 \tag{33}$$

first. These boundary points divide the stable solutions (fixed points) of the original system of equations from the cyclic (Hopf bifurcation) solutions. In Fig. 3, a continuous line marks this boundary curve. Another boundary curve can be determined within the range of cyclic solutions, where regular cycles cease to

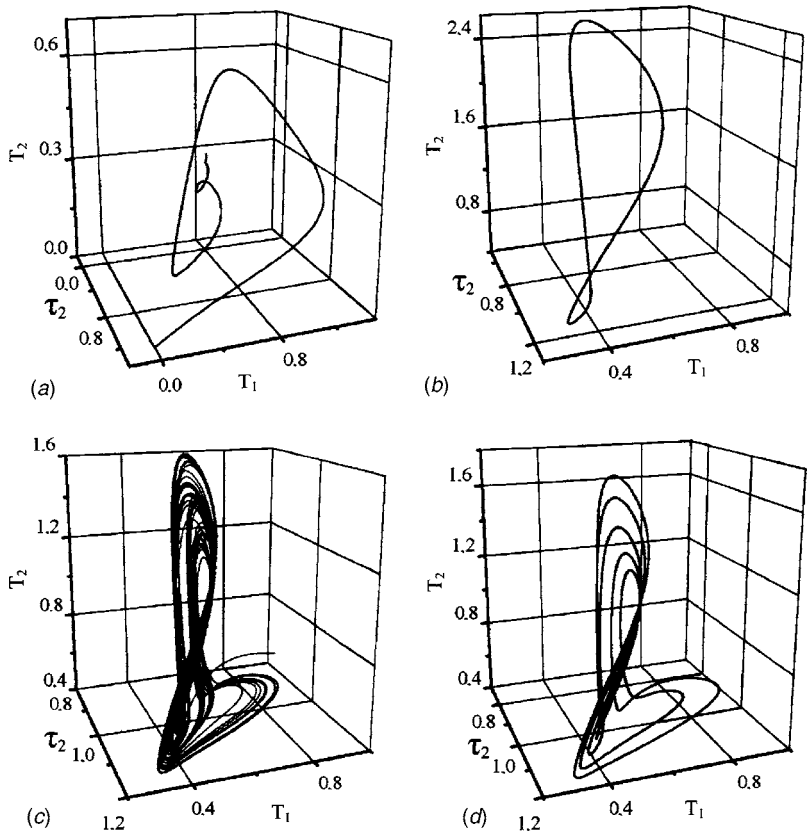


Fig. 4 The typical solutions of the mathematical model ($p=1.03$, $\xi=2.6$ $\eta=4.5$, $a=0.3$, $b=0.012$, $C=1$). (*a*) Fixed point $\zeta=22.5$, (*b*) bifurcation $\zeta=5.8$, (*c*) chaotic $\zeta=11.65$, (*d*) five-cyclic $\zeta=11.33$.

exist, and the solution of the equation system becomes chaotic. This is the range of deterministic instability, which is market by the dashed line in Fig. 3.

The variation of shear stress τ_i and temperature T_i shows the behavior of the model under different technological circumstances. Figure 4 shows functional relationship $T_2[\tau_2(t), T_1(t)]$ at value $\xi=2.6$ and coefficient ζ, calculated for various cutting speeds v. If $\zeta=22.5$, we can see that the solution of the equation system is a typical fixed point (Fig. 4(*a*)), and if $\zeta=5.8$ we get a cyclic solution (Fig. 4(*b*)). However, there is a range between the two values where the solution of the equation reveals chaotic characteristics. This is shown in Fig. 4(*c*), if $\zeta=11.65$.

Such cutting conditions may occur in certain practical technological situations, as we can see in Fig. 1. Apart from the so-called continuous chips of constant width, shown above (Fig. 1(*a*)), and cyclically segmented chips, shown below (Fig. 1(*c*)), aperiodic chips of chaotic nature may also be produced (Fig. 1(*b*)) [12].

The simplified model of chip formation shown above indicates complex characteristics, even if they are reduced to plane $\xi-\zeta$ and only two variables. It can be shown that—even within the range of chaotic-type chips surrounded by the dashed line (Fig. 3)—there are smaller fields in which the equation system has exact, though periodical solutions, containing several cycles. This is indicated for $\zeta=11.33$ in Fig. 4(*d*), which shows five cycles (Figs. 4(*c*) and 4(*d*)) do not reveal the initial, rather short section of curves related to values $\zeta=11.65$ and 11.33, in order to make their relevant character more apparent. This indicates the variability of nonlinear systems, which is somewhat important for practical purposes, since the difference $\Delta\zeta=0.32$, for instance, may even occur as a result of a 2.8% change of the cutting speed. Such change can also be brought about by the fluctuation of the speed of the workpiece, which may even arise from the mains voltage variation.

Finally, another characteristic of the mathematical model's solutions must be mentioned. The calculations often show that there is a considerable difference between the actual numeric values of constitutive equations $F_1(\tau_1, T_1)$ and $F_2(\tau_2, T_2)$ in certain moments of cyclic and chaotic solutions. This means that the difference between the velocities of deformation in the two assumed layers of the shear zone in proportionately the same. This is in harmony with the findings of many researchers, according to which strongly deformed and hardly deformed material occur cyclically in chips as a result of thermoplastic instability in segmental chip formation. This can also be seen in Fig. 1(*c*).

6 Summary

On the basis of the simplified thermomechanic model of chip formation, a mathematical model was established, which comprises five nonlinear autonomous differential equations and a constitutive equation describing the thermoplastic behavior of the workpiece material. The numerical solutions of the mathematical model qualitatively correspond to the various types of chips depending on technology. Besides continuous chips and segmental chips produced frequently, the model also describes the formation of aperiodic chips of irregularly, i.e., chaotically changing thickness, which is supported by empirical findings.

References

[1] Zener, C., and Hollonom, J. H., 1946, "Problems in Non-Elastic Deformation of Metals," J. Appl. Phys., **17**, Febr. 2. pp. 69–82.

[2] Recht, R. F., 1964, "Catastrophic Thermoplastic Shear," J. Appl. Mech., **31**, pp. 189–193.

[3] Davies, M. A., and Burns, T. 2002, "Thermomechanical Oscillations in Material Flow During High-Speed Machining," Philos. Trans. R. Soc. London, Ser. A **359**, pp. 821–846.

[4] Komanduri, R., and Hou, Zhen-Bing 2002, "On Thermoplastic Shear Instability in the Machining of a Titanium Alloy (Ti-6Al-4V)," Metall. Mater. Trans. A **33A**, pp. 2295–3010.

[5] Shaw, M. C., 1984, *Metal Cutting Principles*, Clarendon Press, Oxford, UK, pp. 24–29.

[6] Bayard, Ove, 2000, "Investigation of the Verification Techniques for Modelling Turning Processes," Royal Institute of Technology. Dept of Materials Processing, Production Engineering, Stockholm, pp. 2.3–2.38.

[7] Burns, T. J., and Davies, M. A. 1997, "Nonlinear Dynamics Model Chip Segmentation in Machining," Phys. Rev. Lett. **79**, No. 3. pp. 447–450.

[8] MacGregor, C. W., and Fisher, J. C. 1946, "A Velocity-Modified Temperature for the Plastic Flow of Metals," J. Appl. Mech. March, pp. A-11-A-16.

[9] Hou, Zhen-Bing, and Komanduri, R. 1997, "Modeling of Thermomechanical Shear Instability in Machining," Int. J. Mech. Sci. **39**, pp. 1273–1314.

[10] Kubin, L. P., and Estrin, Y., Spiesser, Ph. 1984, "Low-Temperature Plastic Deformation of Metals and the Bifurcation Theory," Res. Mech. **10**, pp. 25–38.

[11] Burns, T. J., 1990, "A Mechanism for Shear Band Formation in the High Strain-Rate Torsion Test," J. Appl. Mech. **57**, pp. 836–844.

[12] Pálmai, Z., 2004, "Chaotic Phenomena in Chip Formation," *Proceedings of the Fourth Conference on Mechanical Engineering*, Budapest, Hungary, pp. 643–647.

Size Effects on Cavitation Instabilities

Christian F. Niordson

Viggo Tvergaard

Department of Mechanical Engineering,
Solids Mechanics,
Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark

In metal-ceramic systems the constraint on plastic flow leads to so high stress triaxialities that cavitation instabilities may occur. If the void radius is on the order of magnitude of a characteristic length for the metal, the rate of void growth is reduced, and the possibility of unstable cavity growth is here analyzed for such cases. A finite strain generalization of a higher order strain gradient plasticity theory is applied for a power-law hardening material, and the numerical analyses are carried out for an axisymmetric unit cell containing a spherical void. In the range of high stress triaxiality, where cavitation instabilities are predicted by conventional plasticity theory, such instabilities are also found for the nonlocal theory, but the effects of gradient hardening delay the onset of the instability. Furthermore, in some cases the cavitation stress reaches a maximum and then decays as the void grows to a size well above the characteristic material length. [DOI: 10.1115/1.2074747]

1 Introduction

Unstable growth of a small void in an elastic-plastic solid has been studied by [1] for spherically symmetric conditions and by [2,3] for axisymmetric stress conditions. A cavitation instability is predicted when the stress level is sufficiently high such that the work released in the field surrounding the expanding void is enough to drive continued expansion, also for cases of axisymmetric conditions as long as the ratio of the transverse stress and the axial tensile stress is near unity. Related spherically symmetric studies in the context of nonlinear elasticity [4–6] have interpreted a cavitation instability either as a bifurcation from a homogeneously stressed solid or as the growth of a preexisting void.

In material systems where metals are bonded to ceramics, the constraint on plastic flow gives rise to high stress triaxiality, resulting in rapid void growth, as has been observed in experiments for Al_2O_3 reinforced by Al particles [7] and in experiments for a lead wire well bonded to a thick outer glass cylinder [8]. The cavitation in the constrained lead wire has been recently analyzed by [9]. For a thin ductile metal layer used to bond two ceramic blocks together, highly constrained plastic flow occurs under tension normal to the layer, leading to rapid void growth. Unit cell models have been used to analyze the growth of such voids through cavitation instabilities, using remeshing techniques to be able to account for void volume increases by a factor of up to 10^{10} [10,11].

When the void radius is sufficiently small, the growth behavior is not well described by conventional plasticity theory. Then the material model must account for observed size-effects [12–17]. For very small voids under purely hydrostatic tension and axisymmetric loading it has been shown [14,18,19] that the rate of growth is much reduced when the void radius is smaller than the characteristic material length incorporated in the nonlocal plasticity theory, and Tvergaad and Niordson [20] have studied the effect of a material size scale on the interaction of different size voids.

In the present paper, the nonlocal plasticity model used is the finite strain generalization developed by [21] for the strain gradient plasticity theory of [18]. An implementation for power-law hardening has recently been given by [22]. A unit cell model containing a single spherical void is applied, thus allowing for interaction with neighboring voids in a periodic array. But the main focus here is on the effect of the characteristic material length scale on predictions of cavitation instabilities, i.e., on very small void volume fractions and on stress states with a very high stress triaxiality.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received April 26, 2005; final manuscript received May 10, 2005. Review conducted by R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Material Model

A finite strain generalization [21] for the strain gradient plasticity theory by [18] is used to model the material numerically using an updated Lagrangian formulation.

The theory models gradient hardening through three quadratic invariants of the gradient of the plastic strain rate $\rho_{ijk}=\rho_{jik}=\dot{\epsilon}^P_{ij,k}$. Plastic work in the material is performed due to the effective plastic strain measure E^P, defined incrementally through

$$\dot{E}^{P2} = \dot{\epsilon}^{P2} + l_1^2 I_1 + 4 l_2^2 I_2 + \frac{8}{3} l_3^2 I_3 \tag{1}$$

Here, $\dot{\epsilon}^{P2}=\frac{2}{3}\dot{\epsilon}^P_{ij}\dot{\epsilon}^P_{ij}$, is the conventional measure of effective plastic strain, I_1, I_2, and I_3 are three invariants of ρ_{ijk}, and l_1, l_2, and l_3 are three material length parameters.

An alternate form of Eq. (1) using just a single material length parameter can be defined by

$$\dot{E}^{P2} = \dot{\epsilon}^{P2} + l_*^2 \dot{\epsilon}^P_{,i}\dot{\epsilon}^P_{,i} \tag{2}$$

where l_* is a new material length parameter [18]. This effective plastic strain measure results in a strain gradient theory related to the theory by Aifantis [23].

The plastic strain increment is defined according to the usual relation for J_2 flow theory

$$\dot{\epsilon}^P_{ij} = \frac{3}{2}\frac{S_{ij}}{\sigma_{(e)}}\dot{\epsilon}^P = m_{ij}\dot{\epsilon}^P \tag{3}$$

where S_{ij} is the deviator of the Cauchy stress σ_{ij}, $\sigma_{(e)}=\sqrt{\frac{3}{2}S_{ij}S_{ij}}$ is von Mises' effective stress, and m_{ij} denotes the direction of the stress deviator.

Rewriting Eq. (1) in terms of m_{ij} and $\dot{\epsilon}^P$ results in

$$\dot{E}^{P2} = \dot{\epsilon}^{P2} + A_{ij}\dot{\epsilon}^P_{,i}\dot{\epsilon}^P_{,j} + B_i\dot{\epsilon}^P_{,i}\dot{\epsilon}^P + C\dot{\epsilon}^{P2} \tag{4}$$

where the tensors A_{ij}, B_i, and C depend on the three material length parameters l_1, l_2, and l_3 as well as on the spatial gradients of m_{ij} (for details see [18]).

Using the current configuration as reference the principle of virtual work can be expressed as

$$\int_V (\overset{\nabla}{s}_{ij}\delta\dot{\epsilon}_{ij} - \sigma_{ij}(2\dot{\epsilon}_{ik}\delta\dot{\epsilon}_{kj} - \dot{e}_{kj}\delta\dot{e}_{ki}) + (\dot{q} - \dot{\sigma}^S_{(e)})\delta\dot{\epsilon}^P + \overset{\vee}{\rho}_i\delta\dot{\epsilon}^P_{,i})dV$$

$$= \int_S (\dot{T}_{0i}\delta\dot{u}_i + \dot{t}_0\delta\dot{\epsilon}^P)dS \tag{5}$$

where $\overset{\nabla}{s}_{ij}$ is the Jaumann rate of the Kirchhoff stress, $\dot{q}$ is work conjugate to ϵ^P, and $\overset{\vee}{\rho}_i$ is the convected derivative of the higher order Kirchhoff stress. With the displacement vector denoted by u_i, the total strain rate is denoted by $\dot{\epsilon}_{ij}=\frac{1}{2}(\dot{u}_{i,j}+\dot{u}_{j,i})$, and the rate of the displacement gradient is denoted $\dot{e}_{ij}=\dot{u}_{i,j}$. The Kirchhoff stress is related to the Cauchy stress by $s_{ij}=J\sigma_{ij}$, and the higher order Kirchhoff stress is related to the true higher order stress by $\rho_i=J\tau_i$, where J is the determinant of the metric tensor. The effective stress enters the principle of virtual work through $\sigma^S_{(e)}=J\sigma_{(e)}$.

In the right-hand side of Eq. (5) $\dot{T}_{0i}$ and $\dot{t}_0$ are nominal traction rates conjugate to u_i and ϵ^P, respectively. In addition to conventional boundary conditions, conditions on the higher order traction rate $\dot{t}_0$ or the effective plastic strain rate $\dot{\epsilon}^P$ are needed. For the present problem $\dot{t}_0=0$ is imposed for all higher order boundary conditions, on internal boundaries as well as external boundaries. For further discussions on higher order boundary conditions see [18,24,25]

The constitutive equations for the various Kirchhoff stress-measures are

$$\overset{\nabla}{s}_{ij} = \mathcal{R}_{ijkl}(\dot{\epsilon}_{kl} - \dot{\epsilon}^P m_{kl}) = \dot{s}_{ij} - \dot{\omega}_{ik}\sigma_{kj} - \sigma_{ik}\dot{\omega}_{jk} \tag{6}$$

$$\dot{q} = h\left(\dot{\epsilon}^P + \frac{1}{2}B_i\dot{\epsilon}^P_{,i} + C\dot{\epsilon}^P\right) \tag{7}$$

$$\overset{\vee}{\rho}_i = h\left(A_{ij}\dot{\epsilon}^P_{,j} + \frac{1}{2}B_i\dot{\epsilon}^P\right) = \dot{\rho}_i - \dot{e}_{ik}\rho_k \tag{8}$$

where $h=h[E^p]$ is the hardening modulus, $\dot{\omega}_{ij}$ is the antisymmetric part of $\dot{e}_{ij}$ and

$$\mathcal{R}_{ijkl} = \frac{E}{1+\nu}\left(\frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + \frac{\nu}{1-2\nu}\delta_{ij}\delta_{kl}\right) \tag{9}$$

is the elastic stiffness tensor. Here, E is Young's modulus, ν is Poisson's ratio, and δ_{ij} is Kronecker's delta.

For details on the strain gradient theory and the finite strain generalization see [18,21].

3 Numerical Method and Problem Description

The numerical solutions are obtained using a two-field finite element method similar to that used by [26,27] to model the gradient theory by [23]. This method has been used by [28] to model the small strain theory by [18], while [21,22] have used the approach to solve problems at finite strains.

In the results to be presented, power-law hardening material behavior is assumed, so the serendipity elements developed in [22] have been used. Further details on the numerical implementation can be found in [28,21,22].

An axisymmetric cell model is used to model a hexagonal arrangement of spherical voids (Fig. 1(*a*)). The voids are aligned in planes as shown in Fig. 1(*b*). Exploiting symmetry the cell used for computations is shown in Fig. 1(*c*). The void and cell radii are denoted R_v and R_c, respectively, and the length of the cell is denoted L_c. Hence the in-plane void spacing is $2R_c$, while the out-of-plane void spacing is $2L_c$. For the results presented throughout this paper $L_c/R_c=1$ is used so that the in-plane and out-of-plane spacings are equal. The boundary conditions applied to the cell are specified by

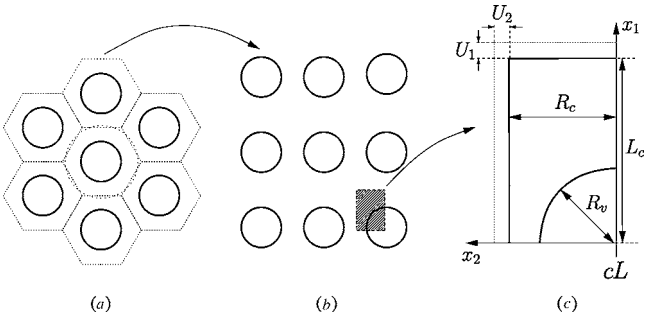


Fig. 1 Cell model for a material with an array of voids. (*a*) Hexagonal distribution of voids with a cylindrical cell indicated by the circular dashed line. (*b*) A part of a layer of voids. (*c*) Using the symmetry of the problem half a void can be modeled in an axisymmetric cell. The cell radius and length are denoted R_c and L_c, respectively, and the void radius is denoted R_v.

$$\begin{aligned}
&\dot{u}_1 = 0 \quad \text{and} \quad \dot{T}_{02} = 0, \quad \text{for} \quad x_1 = 0\\
&\dot{u}_1 = \dot{U}_1 \quad \text{and} \quad \dot{T}_{02} = 0, \quad \text{for} \quad x_1 = L_c + U_1\\
&\dot{u}_2 = 0 \quad \text{and} \quad \dot{T}_{01} = 0, \quad \text{for} \quad x_2 = 0\\
&\dot{u}_2 = \dot{U}_2 \quad \text{and} \quad \dot{T}_{01} = 0, \quad \text{for} \quad x_2 = R_c + U_2
\end{aligned} \tag{10}$$

where $\dot{U}_1$ and $\dot{U}_2$ are determined through a Rayleigh-Ritz procedure ensuring that a constant ratio of the average true stresses is maintained

$$\frac{\sigma_2}{\sigma_1} = \rho \tag{11}$$

while specifying increasing void growth. In addition to these conventional boundary conditions, $\dot{t}_0=0$ is specified along the entire surface of the material modeled. At the surface of the cell this constitutes the appropriate symmetry boundary condition, while at the void surface it models that dislocations are free to pass from the material through the free surface.

In the present study the trial functions for the Rayleigh-Ritz procedure are based on prescribed unit displacements of each of the two cell sides in addition to the displacement of the material point located at $(0,R_v)$ along the direction of the x_1-axis (see Fig. 2(*a*)).

In Fig. 2 two typical finite elements meshes used for the analyses are shown. Figure 2(*a*) shows a mesh for a large void, while Fig. 2(*b*) shows a mesh for a very small void, where the mesh is

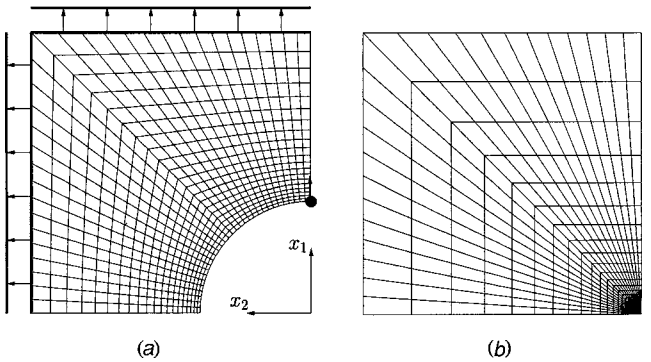


Fig. 2 Typical finite element meshes used for the analyses for (*a*) a large void and (*b*) a small void. In (*a*) the two sides and the node used as control degrees of freedom for the Rayleigh-Ritz procedure are highlighted.

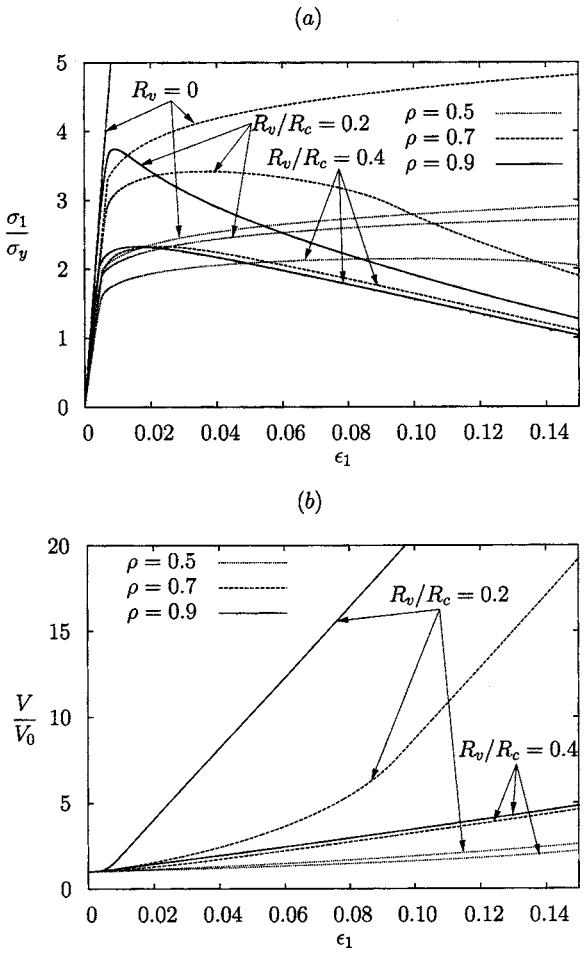


Fig. 3 Conventional results for a material without voids and for two materials containing voids where the in-plane spacing equals the out-of-plane spacing (L_c/R_c=1). For one of the materials R_v/R_c=0.2 and for the other material R_v/R_c=0.4. The analyses are carried out for three different values of the ratio of transverse stress to axial stress ρ. The material parameters are given by σ_y/E=0.004, ν=1/3, and n=10. (*a*) Shows the overall response in terms of the true stress as a function of logarithmic strain, and (*b*) shows the relative void growth.

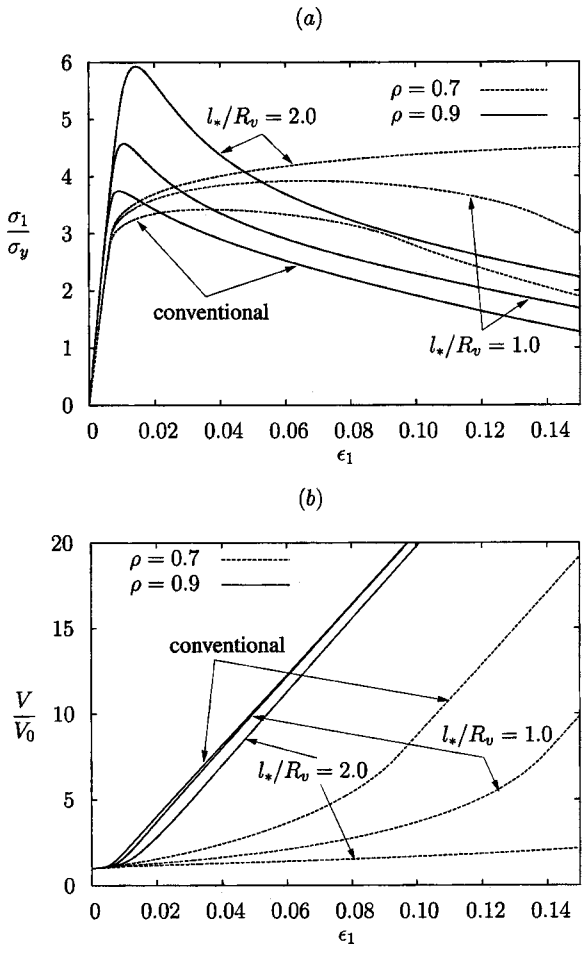


Fig. 4 Size dependent results for a material with large voids (R_v/R_c=0.2) with equal in-plane and out-of-plane spacing (L_c/R_c=1). Both conventional and gradient dependent results with l_*/R_v=1.0 and l_*/R_v=2.0 are shown for two different stress ratios. The conventional material parameters are given by σ_y/E=0.004, ν=1/3, and n=10. (*a*) Shows the overall response in terms of the true stress as a function of strain, and (*b*) shows the relative void growth.

highly refined around the void. In Fig. 2(*a*) the two sides along which the Rayleigh-Ritz degrees of freedom are imposed are highlighted together with the node at the void boundary on the x_1 axis, which is used to control increasing void growth through mechanical instabilities.

4 Results

Power-law hardening material behavior is assumed with the following expression for the tangent modulus

$$E_T = \frac{E}{n}\left(\frac{E^P}{\epsilon_0} + 1\right)^{(1/n)-1} \tag{12}$$

which deviates very little from that corresponding to a standard power-law. In this expression $\epsilon_0 = \sigma_y/E$ is the uniaxial yield strain, with σ_y denoting the initial yield stress, and n is the hardening exponent. The hardening modulus is then calculated through

$$h[E_T] = \left(\frac{1}{E_T} - \frac{1}{E}\right)^{-1} \tag{13}$$

The true stress as a function of the logarithmic strain are shown in Fig. 3(*a*) for conventional materials at different load ratios ρ. Results are presented for a homogeneous material as well as for two materials with rather large voids. The aspect ratio of the computational cells is unity corresponding to equal in-plane and out-of-plane void spacing. The void radii for the two materials with voids are 20% and 40% of the cell radii, which corresponds to initial void volume fractions of around 0.533% and 4.27%, respectively. Figure 3(*b*) shows the relative void growth as a function of the logarithmic strain, for the two materials containing voids. In Fig. 3 results are presented for different values of the stress ratio ρ, namely, for ρ=0.5, 0.7, and 0.9. It is seen that the overall response depends strongly on the presence of voids, especially for the higher triaxialities, where significant void growth is accompanied by a dramatic softening of the response. Figure 3(*b*) shows that the relative void growth generally is larger for smaller voids at a given value of the overall strain, which can be explained due to the fact that the elastic-plastic material is almost incompressible, since only elastic deformations contribute to volume change. Hence, for a given overall deformation (and overall volume change) a smaller void will have grown relatively more than a larger void.

In Fig. 4 conventional results are compared to gradient dependent results for two different values of l_*/R_v at two values of the stress ratio ρ. The size of the void compared to the size of the cell is given by R_v/R_c=0.2, which corresponds to an initial void volume fraction of around 0.533%. The solid curves show results for

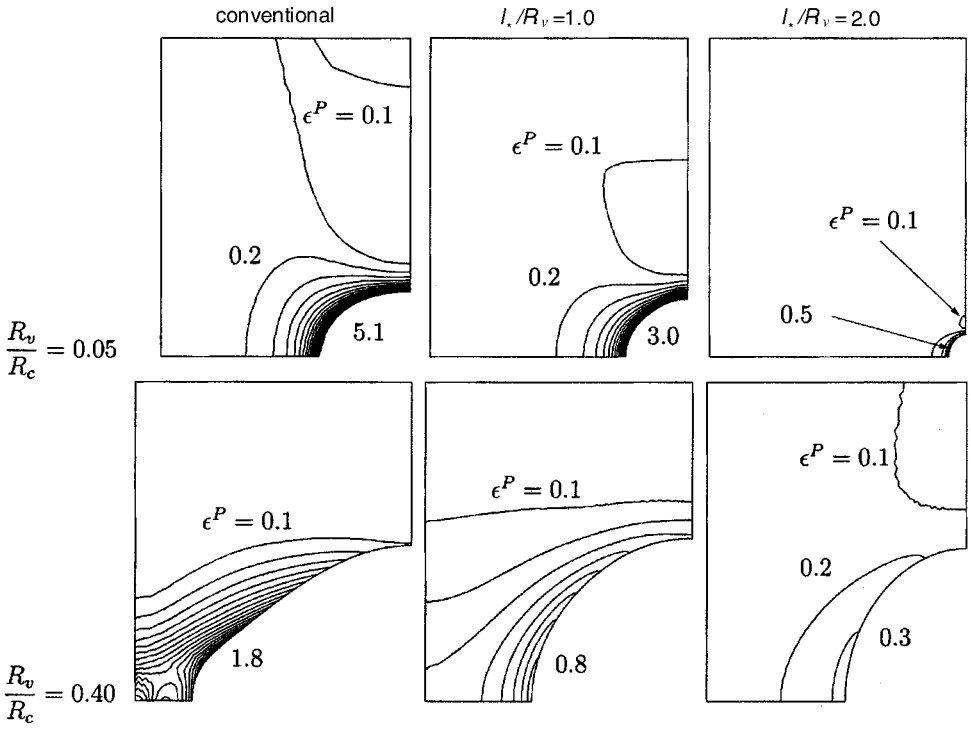


Fig. 5 Contours of effective plastic strain for ρ=0.7 at an overall strain of ϵ_1=0.15. The conventional material parameters are given by σ_y/E=0.004, ν=1/3, n=10. The contours of effective plastic strain differs by $\Delta\epsilon^P$=0.1.

ρ=0.9, while the dashed curves show results for ρ=0.7. It is seen from Fig. 4(*a*) that a higher overall stress level is obtained for ρ =0.9 when compared to the lower value of ρ for all the different values l_*. Furthermore, for ρ=0.9, the peak load is more narrow for the higher triaxiality and the overall material softening is more pronounced for both the conventional material and the gradient dependent materials. The reason for this can be found in Fig. 4(*b*) which shows the relative void growth as a function of strain. Here, it is seen that the material softening can be attributed to significant void growth. For ρ=0.7, the relative void growth is significantly limited by strain gradient effects. Hence, for l_* comparable to the void radius, there is only little void growth and therefore no significant overall softening of the material response in Fig. 4(*a*). On the other hand, for ρ=0.9 the amount of void growth is very large already at small strains and it has a rather weak dependence on the material length, which explains why there is pronounced softening for all values of l_*.

Figure 5 shows contours of effective plastic strain, ϵ_p, at the overall strain ϵ_1=0.15 for R_v/R_c=0.05 and R_v/R_c=0.40. These void radii correspond to initial void volume fractions of $8.33\cdot10^{-5}$ and $4.27\cdot10^{-2}$, respectively. The stress ratio is given by ρ=0.7. The first row shows results for the smaller void, while the second row shows results for the larger void. The first column shows results for conventional materials, while the second and third column show results for l_*/R_v=1.0 and l_*/R_v=2.0, respectively. For all the analyses the maximum plastic strain level is attained at the void. It is seen that increasing the material length parameter l_* leads to smaller gradients of plastic strain. For the initially smaller void, the figure illustrates that the relative void growth is significantly decreased when increasing the material length parameter. For the initially larger void, V/V_0 is much less sensitive to gradient hardening, and the relative void growth is primarily determined by the overall deformation ϵ_1. However, the shape of the void is affected by l_*/R_v, also for the initially larger void, where the analyses have been carried out to the range where coalescence becomes relevant. For both void sizes the shape of the void changes from oblate for the conventional analysis toward prolate when increasing l_*/R_v.

The response for materials containing sparsely distributed voids is shown in Fig. 6(*a*) where R_v/R_c=0.001 has been used for the calculations so that the initial void volume fraction is around $6.67\cdot10^{-10}$. Results for two values of the stress ratio ρ are presented. For ρ=0.9 plasticity does not occur on the macroscopic scale, since the stress level does not attain large enough values. For this value of ρ a stress level of 10 times the yield stress must be applied for macroscopic plasticity to occur. However, for the smaller value of the stress ratio ρ=0.7, plastic deformation does occur at the macroscopic scale, since the stress level exceeds the yield value of approximately 3.33 times the yield stress. The shape of the response curves are practically unaffected by gradient effects, as the gradient hardening only influences the material locally around the void in a region scaling with l_*. The relative void growth is shown in Fig. 6(*b*) as a function of the overall strain ϵ_1. Gradient effects clearly have a large influence on the relative void growth for ρ=0.7, where void growth is suppressed by gradient effects. Although this is less obvious, gradient hardening also has important consequences for void growth when ρ=0.9. Figure 6(*b*) indicates that the curve of relative void growth has a vertical asymptote, as in fact a cavitation instability is obtained [1–3]. In Fig. 6(*a*) this corresponds to the response curve approaching a limit point from below (increasing stress and strain). For the two size-dependent analyses, with l_*/R_v=1.0 and l_*/R_v=2.0, the maximum strain level exceeds that of the conventional material, and thereafter the overall strain decreases as the voids grow and approaches the void growth curve for the conventional material under decreasing overall strain. In Fig. 6(*a*), for the size-dependent analyses, this corresponds to the response curves exceeding the limit point of the conventional material, to attain a maximum value of overall stress and strain, and subsequently start reversing in order to approach the limit point of the conventional

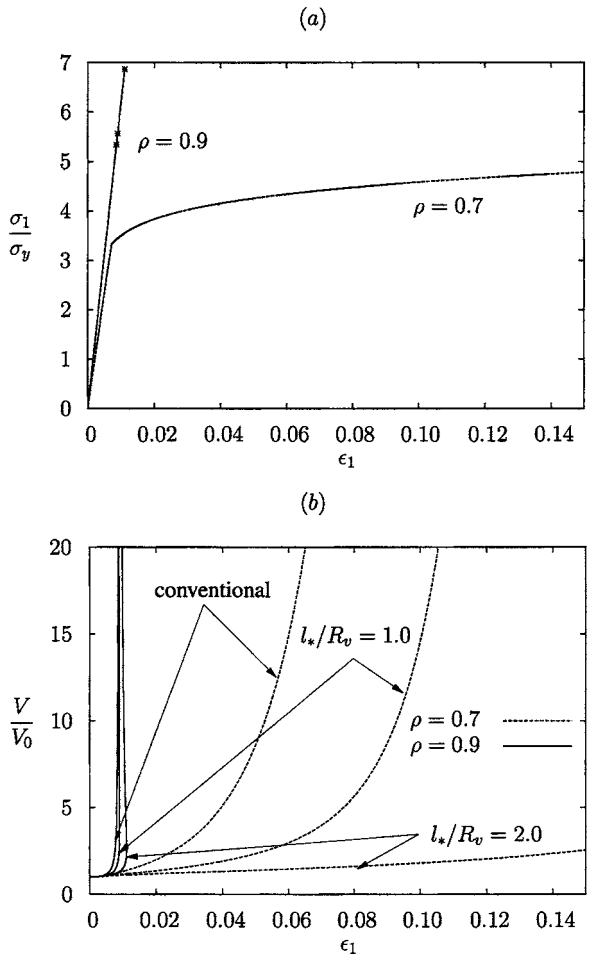


Fig. 6 Size dependent results for sparsely distributed voids (R_v/R_c=0.001). Both conventional and gradient dependent results with l_*/R_v=1.0 and l_*/R_v=2.0 are shown for two different stress ratios. The conventional material parameters are given by σ_y/E=0.004, ν=1/3, and n=10. (a) Shows the overall response in terms of the true stress as a function of strain, and (b) shows the relative void growth.

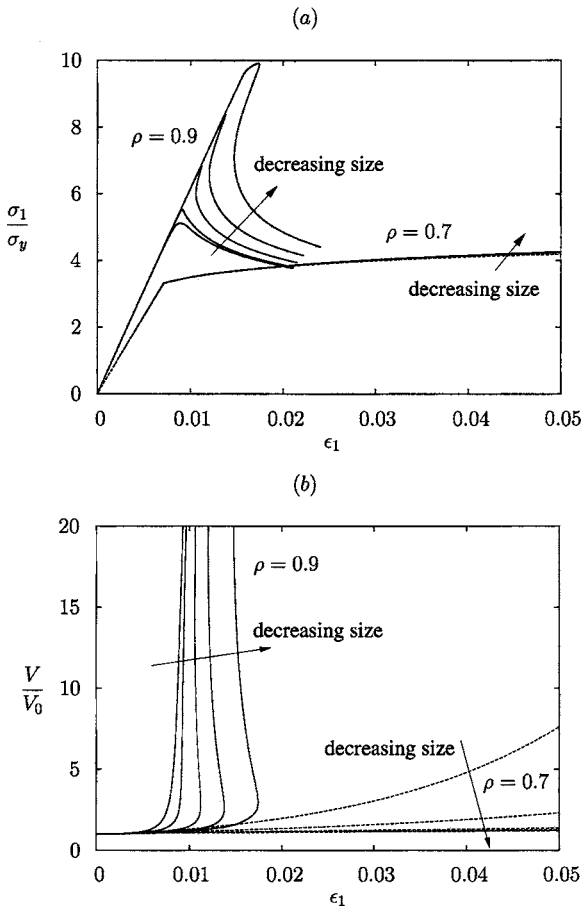


Fig. 7 Size dependent results for a material with R_v/R_c=0.05 and L_c/R_c=1. Both conventional and gradient dependent results with l_*/R_v=1.0, 2.0, 3.0, and 4.0 are shown for two different stress ratios. The curves correspond to a material with a fixed length parameter and a fixed initial void volume fraction, but with different void size and void spacing. The conventional material parameters are given by σ_y/E=0.004, ν=1/3, and n=10. (a) Shows the overall response in terms of the true stress as a function of strain, and (b) shows the relative void growth.

material from above (under decreasing stress and strain). In Fig. 6(*a*) the markings on the response curves for ρ=0.9 show the limit point for the conventional material (lowest value of stress and strain), and the two points at which the responses for the gradient dependent materials reverse (the point for the largest stress and strain is the marking for l_*/R_v=2.0). For the gradient dependent materials, the material length parameter l_* decreases relative to the deformed void size, and therefore as the void grows larger, the gradient effects lose significance at distances comparable to the deformed void size. Hence, the response, and even the deformations close to the void become less affected by gradient effects, as the overall stress and deformation level approaches that of the conventional material (for decreasing stress and strain). In this sense the cavitation instability for a gradient dependent materials is at first more dramatic than for a conventional material, as it is accompanied by decreasing stresses and strains at the macroscopic level. However, as the voids grow large, the cavitation instability strains become comparable.

Void growth at a larger initial void volume fraction is studied in Fig. 7, where the initial void volume fraction is $8.33 \cdot 10^{-5}$(R_v/R_c=0.05, L_c/R_c=1.0). Results are presented for two different values of ρ, and for five different values of l_*/R_v (0, 1.0, 2.0, 3.0, and 4.0). The arrows in the figures show how the results are affected by decreasing the initial void size R_v, and simultaneously the void spacing (or conversely increasing the material length parameter l_*). Figure 7(*b*) shows that void growth is seen to be suppressed by gradient effects for both ρ=0.7 and ρ=0.9. In Fig. 7(*a*) it is seen that the overall response is almost unaffected by size effects for ρ=0.7. This can be explained due to the fact that the relative void growth remains rather small for the overall deformation levels considered here, and hence, the overall properties of the material are unaffected by the small faults in the material. For ρ=0.9 the void starts to grow rapidly at $\epsilon \approx 0.009$ for the conventional material. Even though the void growth is dramatic, a cavitation instability is not reached as for the smaller void discussed in relation to Fig. 6. This can be seen from Fig. 7(*a*) since the overall response neither approaches a limit point nor goes through a mechanical instability and in Fig. 7(*b*) since the curve does not become parallel with the vertical axis. Increasing the material length parameter leads to delayed but more dramatic void growth (Fig. 7(*b*)), and for the three larger values of l_*/R_v the overall strain level even goes through a maximum and then starts to decrease. In Fig. 7(*a*) this is illustrated by the response curves having sections where both overall stress and strain decrease. For the largest value of the material length parameter l_*/R_v=4.0, plastic deformation in a major part of the cell occurs, which explains why the response curve goes through a short phase of reduced slope before passing through the mechanical instabil-

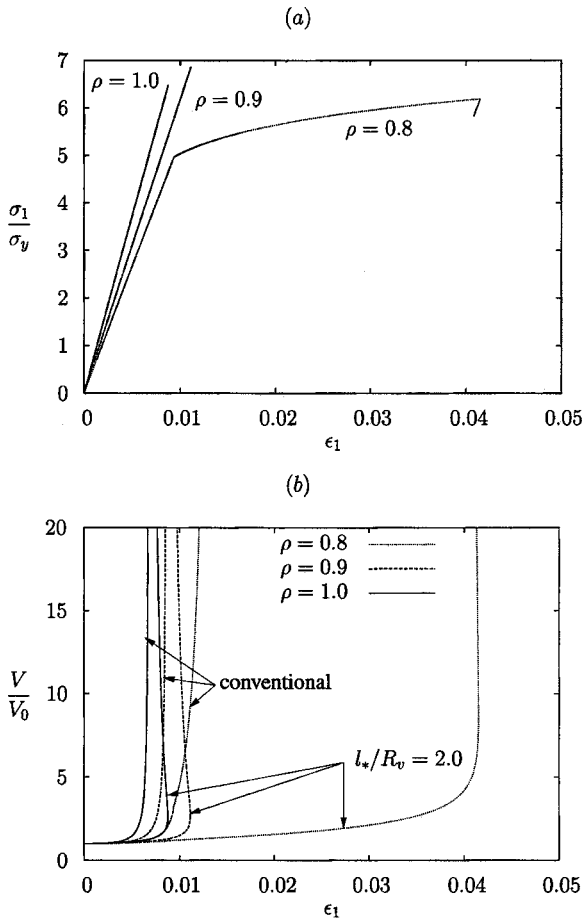


Fig. 8 Size dependent results for sparsely distributed voids (R_v/R_c=0.001) for different values of ρ. Both conventional and gradient dependent results with l_*/R_v=2.0 are shown. The conventional material parameters are given by σ_y/E=0.004, ν=1/3, and n=10. (a) Shows the overall response in terms of the true stress as a function of strain, and (b) shows the relative void growth.

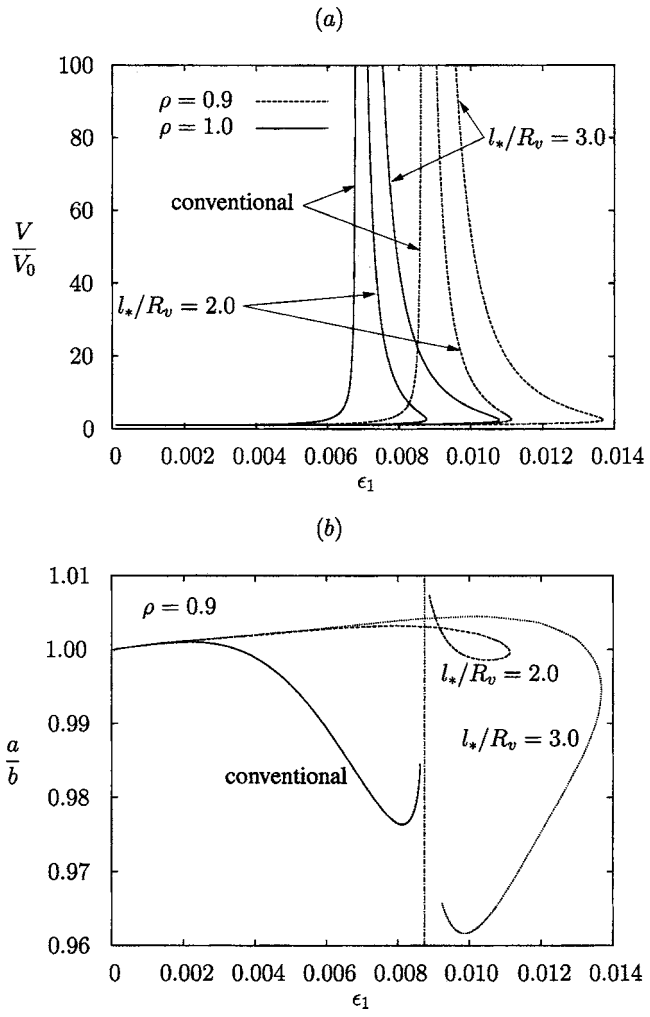


Fig. 9 Size dependent results for sparsely distributed voids (R_v/R_c=0.001). Both conventional and gradient dependent results with l_*/R_v=2.0 and 3.0 are shown for two different stress ratios. The conventional material parameters are given by σ_y/E=0.004, ν=1/3, and n=10. (*a*) Shows the relative void growth as a function of strain, and (*b*) shows the ratio of the void size in the x_1 direction to the void size in the x_2 direction for ρ=0.9.

ity. The fact that the strain level goes through a maximum at the onset of instability was also noted by [19] for spherically symmetric loading conditions.

Sparsely distributed voids (R_v/R_c=0.001) at different large values of ρ are investigated in Fig. 8. For the overall loading conditions specified by ρ=0.9 and ρ=1.0 the response curves show that the material behaves elastically on the macroscopic scale. For the conventional material under these loading conditions a cavitation instability is reached in the elastic regime, and for the gradient dependent analyses with l_*/R_v=2.0 this state is approached from above (with decreasing overall stress and strain). For ρ=0.8 overall plasticity sets in long before a cavitation instability is reached for the conventional material. Within the range considered here, the cavitation instability is not reached as the analysis is terminated when V/V_0=200 as the finite element mesh around the void becomes too distorted. For the size-dependent material with l_*/R_v=2.0 void growth is suppressed such that the analysis can be carried out to a larger overall deformation level. At an overall strain of just above ϵ_1=0.04 the cavitation instability is reached and the response curve shows that the material goes through a mechanical instability. For this value of ρ the cavitation instability of the conventional material does not serve as a limit state for the gradient dependent material since the instabilities are reached at different states in the plastic regime.

Figure 9(*a*) shows the relative void growth for ρ=0.9 and ρ=1.0. Both conventional results and gradient dependent results with l_*/R_v=2.0 and 3.0 are shown. Also here it is seen how the curves for the gradient dependent analyses first reach a peak strain, and then approach the curve for the conventional analyses under decreasing overall strain ϵ_1. For ρ=0.9 the ratio of the void size in the main tensile direction (the x_1 direction), denoted by a, to the void size in the x_2 direction, denoted by b, is shown as a function of overall strain in Fig. 9(*b*). For $a/b>1$ the void shape is prolate, while the void shape is oblate for $a/b<1$. For all values of l_* the void moves toward a prolate shape for small deformation levels. At an overall strain of around ϵ_1=0.0035 the void in the conventional material becomes oblate. For l_*/R_v=2.0 the void shape is prolate until around the maximum load point, where is becomes oblate, and then again becomes prolate upon further deformation. For l_*/R_v=3.0 the void shape is prolate until shortly before the mechanical instability is reached, where it becomes oblate until the termination of the analysis.

Until now only the single parameter version of the gradient theory has been discussed. In Fig. 10 the general three parameter version of the theory is studied for sparsely distributed voids (R_v/R_c=0.001), for ρ=0.9. The relative void growth is shown as a

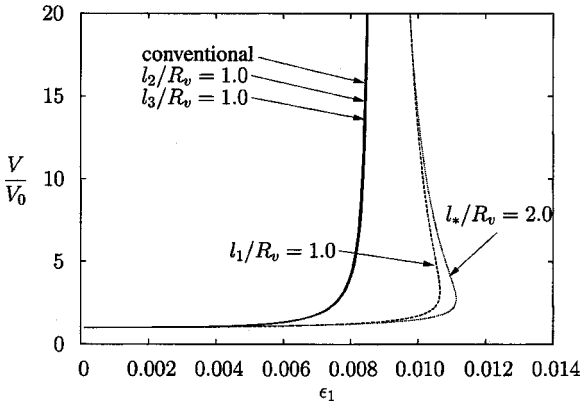


Fig. 10 Relative void growth for sparsely distributed voids (R_v/R_c=0.001) using the general theory with each of the three length parameter activated one by one keeping the other two equal to zero. For comparison results for the single parameter version is included. The conventional material parameters are given by σ_y/E=0.004, ν=1/3, and n=10.

function of strain. It is seen from the figure that l_2 and l_3 has an insignificant influence on void growth for this specific problem. This is in agreement with the findings in [18], where it is reported that void growth for a single void in an infinite medium under hydrostatic tension shows no dependence on l_2 and only minor dependence on l_3. Figure 10 shows that l_1 is an important parameter, and that in fact l_1 has a comparable influence on void growth as l_*, when the value of l_1 is chosen to be around half that of l_*. This also is in excellent agreement with the finding in [18] for a single void under hydrostatic tension, where it is reported that l_* has the same influence as l_1, apart from a factor of about 2.

5 Concluding Remarks

For a single void in an infinite solid made of a conventional elastic-plastic material a cavitation instability may be predicted when the stress level is sufficiently high, such that the work released in the field surrounding the expanding void is sufficient to drive the void expansion. This typically occurs when the material is subject to highly constrained plastic flow. The unstable void growth occurs while the stress and strain states at infinity stay constant. The analyses in the present paper have shown that this type of behavior is noticeably changed when the material follows a strain gradient plasticity theory.

In [29] the development of the void shape during void growth and void collapse in power-law creeping solids were studied. In the present paper it is illustrated that the transition from a prolate to an oblate void shape depends on the material length parameter in an elastic-plastic power-law hardening material.

For small voids with a radius comparable to or smaller than the characteristic material length incorporated in a nonlocal plasticity theory it has been shown [14,20] that the rate of void growth is much reduced relative to that in a conventional elastic-plastic material. This is also found in the present analyses, but it is seen that the delay due to nonlocal effects is more pronounced for low to moderately high stress triaxiality, as specified by values 0.5 or 0.7 of the stress ratio $\rho=\sigma_2/\sigma_1$. For high stress ratios ρ=0.9 or ρ =1.0, where a cavitation instability is likely to occur if the void volume fraction is sufficiently low, the behavior is apparently dominated by the tendency towards unstable growth, and the delay in void growth is less pronounced, e.g. see Figs. 6–8. However, also when a cavitation instability develops as clearly seen in Figs. 6, 7, and 9, the onset of the instability is delayed by the nonlocal effects.

All the investigations of void growth under the influence of a characteristic material length (e.g., [14,20]) have shown that the effect on void growth is smaller when the ratio between the material length and the void radius becomes smaller. This is important in the case of cavitation instabilities, where the value of the current void radius suddenly grows significantly (in some calculations in the present paper the void growth is followed up to a radius increase by a factor 5, and [10] has followed unstable growth where the radius has increased by a factor larger than 1000). The material length remains fixed and therefore a large increase of the void radius will reduce the ratio of the material length and the void radius to a smaller and smaller value, where the reduction of the rate of void growth gradually disappears as the void grows large. This is the reason for the behavior illustrated in Figs. 6–9, where the nonlocal materials show that a peak strain is reached before the remote strain and stress levels are again reduced according to the quasi-static analysis, while the void size continues to increase.

For the conventional material, accounting for no material length, the cavitation instability is characterized by the fact that the remote strain level becomes constant, while the void grows very large. However, for the nonlocal material the present analyses show that the remote strain level and the corresponding stress level will reach a peak value so that here the instability will occur at a snap point, where the work released in the field surrounding the void will exceed that needed to drive the void growth. In these circumstances it is expected that the instability will in reality occur more dramatically, with some of the released elastic energy transformed into kinetic energy.

Acknowledgment

This work is financially supported by the Danish Technical Research Council in a project titled Modeling Plasticity at the Micron Scale.

References

[1] Bishop, R. F., Hill, R., and Mott, N. F., 1945, "The Theory of Indentation and Hardness Tests," Proc. Phys. Soc. **57**, pp. 147-159.
[2] Huang, Y., Hutchinson, J., and Tvergaard, V., 1991, "Cavitation Instabilities in Elastic-Plastic Solids," J. Mech. Phys. Solids, **39**, pp. 223-241.
[3] Tvergaard, V., Huang, Y., and Hutchinson, J., 1992, "Cavitation Instabilities in a Power Hardening Elastic-Plastic Solid," Eur. J. Mech. A/Solids, **11**, pp. 215–231.
[4] Ball, J. M., 1982, "Discontinuous Equilibrium Solutions and Cavitation in Nonlinear Elasticity," Philos. Trans. R. Soc. London, Ser. A, **306**, pp. 557-611.
[5] Horgan, C., and Abeyaratne, R., 1966, "A Bifurcation Problem for a Compressible Nonlin-Early Elastic Medium: Growth of a Micro-void," J. Elast., **16**, pp. 189-200.
[6] Chou-Wang, M.-S., and Horgan, C., 1989, "Void Nucleation and Growth for a Class of Incompressible Nonlinearly Elastic Materials," Int. J. Solids Struct., **25**, pp. 1239-1254.
[7] Flinn, B., Ruhle, M., and Evans, A., 1989, "Toughening in Composites of Al_2O_3 Reinforced with Al," Acta Metall., **37**, pp. 3001-3006.
[8] Ashby, M., Blunt, F., and Bannister, M., 1989, "Flow Characteristics of Highly Constrained Metal Wires," Acta Metall., **37**, pp. 1847-1857.
[9] Tvergaard, V., 2004, "Effect of Residual Stress on Cavitation Instabilities in Constrained Metal Wires," Trans. ASME, J. Appl. Mech., **71**, pp. 560-566.
[10] Tvergaard, V., 1997, "Studies of Void Growth in a Thin Ductile Layer Between Ceramics," Comput. Mech., **20**, pp. 186-191.
[11] Tvergaard, V., 2000, "Interface Failure by Cavity Growth to Coalescence," Int. J. Mech. Sci., **42**, pp. 381-395.
[12] Fleck, N. A., Muller, G. M., Ashby, M. F., and Hutchinson, J. W., 1994, "Strain Gradient Plasticity: Theory and Experiment," Acta Metall. Mater., **42**(2), pp. 475-487.
[13] Ma, Q. and Clarke, D. R., 1995, "Size Dependent Hardness of Silver Single Crystals," J. Mater. Res., **10**, pp. 853-863.
[14] Fleck, N. A., and Hutchinson, J. W., 1997, "Strain Gradient Plasticity," *Advances in Applied Mechanics*, J. W. Hutchinson and T. Y. Wu, **33**, Academic Press, New York, pp. 295-361.
[15] Stölken, J. S., and Evans, A., 1998, "A Microbend Test Method for Measuring the Plasticity Length Scale," Acta Mater. **46**(14), pp. 5109-5115.
[16] Begley, M. R., and Hutchinson, J. W., 1998, "Mechanics of Size-Dependent Indentation," J. Mech. Phys. Solids, **46**(10), pp. 2049-2068.
[17] Haque, M. A., and Saif, M. T. A., 2003, "Strain Gradient Effect in Nanoscale Thin Films," Acta Mater. **51**(11), pp. 3053-3061.
[18] Fleck, N. A., and Hutchinson, J. W., 2001, "A Reformulation of Strain Gradient Plasticity," J. Mech. Phys. Solids, **49**, pp. 2245-2271.
[19] Huang, Y., Gao, H., Nix, W., and Hutchinson, J., 2000, "Mechanism-Based Strain Gradient Plasticity. II Analysis. J. Mech. Phys. Solids, **48**, pp. 99-128.

[20] Tvergaard, V., and Niordson, C. F., 2004, "Nonlocal Plasticity Effects on Interaction of Different Size Voids," Int. J. Plast., **20**, 107-120.
[21] Niordson, C. F., and Redanz, P., 2004, "Size-Effects in Plane Strain Sheet-Necking," J. Mech. Phys. Solids, **52**, pp. 2431-2454.
[22] Niordson, C. F., and Tvergaard, V., 2005, "Instabilities in Power Law Gradient Hardening Materials," Int. J. Solids Struct., **42**, pp. 2559–2573.
[23] Aifantis, E. C., 1984, "On the Microstructural Origin of Certain Inelastic Models," ASME J. Eng. Mater. Technol., **106**(4), pp. 326-330.
[24] Gurtin, M., and Needleman, A., 2005, "Boundary Conditions in Small-Deformation, Single-Crystal Plasticity that Account for the Burgers Vector," J. Mech. Phys. Solids, **53**, pp. 1-33.
[25] Niordson, C. F., 2003, "Strain Gradient Plasticity Effects in Whisker-Reinforced Metals," J. Mech. Phys. Solids, **51**, pp. 1863-1883.
[26] De Borst, R., and Mühlhaus, H.-B., 1992, "Gradient-Dependent Plasticity: Formulation and Algorithmic Aspects," Int. J. Numer. Methods Eng., **35**, pp. 521-539.
[27] De Borst, R., and Pamin, J., 1996, "Some Novel Developments in Finite Element Procedures for Gradient-Dependent Plasticity," Int. J. Numer. Methods Eng., **39**, pp. 2477-2505.
[28] Niordson, C. F., and Hutchinson, J. W., 2003, "Non-Uniform Plastic Deformation of Micron Scale Objects," Int. J. Numer. Methods Eng., **56**, pp. 961-975.
[29] Budiansky, B., Hutchinson, J., and Slutsky, S., 1982, "Void Growth and Collaps in Viscous Solids," edited by H. Hopkins and M. Sewell, *Mechanics of Solids, The Rodney Hill 60th Anniversary Volume*, Pergamon Press, Oxford, pp. 13-45.

Surface Stress and Reversing Size Effect in the Initial Yielding of Ultrathin Films

G. Gioia

X. Dai

Department of Theoretical and Applied Mechanics,
University of Illinois,
Urbana, IL 61801

Very recent experiments indicate that in free-standing metallic films of constant grain size the initial yield stress increases as the film becomes thinner, it peaks for a thickness on the order of 100 *nm, and then starts to decrease. This reversing size effect poses two challenges: (1) It cannot be explained using currently available models and (2) it appears to contradict the classical experimental results due to J. W. Beams [1959, "Mechanical Properties of Thin Films of Gold and Silver," in Structure and Properties of Thin Films, Wiley, New York, pp. 183–198]. Here we show that the reversing size effect can be explained and the contradiction dispelled by taking into account how the initial yielding is affected by the surface stress. We also predict that the mode of failure of a film changes from ductile to brittle for a thickness on the order of* 100 nm*, in accord with experiments.* [DOI: 10.1115/1.2074767]

1 Introduction

The mechanical behavior of tiny metallic bodies has long been known to be subject to size effects [1]. For example, the yield stress of crystalline whiskers may exceed the yield stress of large crystals of the same material by a factor of 10 or more [2]. With the development of nanotechnologies in recent years, much new research has been devoted to elucidating size effects in polycrystalline ultrathin films.

One size effect that has drawn much attention pertains to the large strain gradients that arise, for example, in films subjected to bending. This size effect has been ascribed to the high density of geometrically necessary dislocations induced by the strain gradient [3,4]. Another size effect pertains to the texture (or preferential grain orientation) that is characteristic of thin films grown on crystalline substrates. Because a texture frequently leads to a higher yield stress [5,6], this size effect can be readily explained. Still another size effect pertains to the grain size, which in annealed films tends to scale with the thickness of the film [5,6]. Because smaller grains lead to a higher yield stress (the Hall-Petch relation [7–9]) or perhaps to a lower yield stress (the reverse Hall-Petch relation, valid for grains smaller than about 10 nm [10]), this size effect can be readily explained.

In a very recent experimental study [11], the yield stress of gold films of constant texture and grain size subjected to uniaxial applied tension showed a peculiar size effect. The yield stress increased with diminishing film thickness, up to a thickness $h=500$ nm. Then, for $h=300$ nm, the yield stress remained the same as for $h=500$ nm, indicating that the yield stress had attained a maximum value for $h\approx 400$ nm. In another recent, similar experimental study [12], the yield stress of pure aluminum films increased up to a thickness $h=100$ nm; then, for $h<100$ nm, the yield stress started to decrease. This reversing (first hardening, then softening) size effect cannot be explained by any of the models proposed so far, because those models predict a hardening size effect (for constant grain size).

Here we model the film using continuum mechanics. We start by establishing expressions for the compressive stresses induced in the film by the surface stress. After adding these stresses to the applied stress, we use the von Mises yield condition to ascertain the value of the applied stress at initial yielding or *apparent yield stress*. Our results indicate that in films subjected to uniaxial applied tension the surface stress causes a reversing size effect on the apparent tensile yield stress. Using the values of surface stress determined in recent years via atomistic methods [13], we estimate that this size effect reverses for a thickness on the order of 100 nm, in accord with the experimental results summarized above.

Then, we use the well-known failure criterion proposed by Hancock and Mackenzie [14] to ascertain the mode of failure of the film. Our results indicate that the mode of failure changes from ductile to brittle for thicknesses close to the thickness for which the size effect reverses. These results are in accord with the experimental results of Ref. [11].

Last, we show that the surface stress may lead to entirely disparate size effects depending on the applied stress. In particular, we find that in films subjected to biaxial applied tension the surface stress does not lead to a reversing size effect. This finding reconciles the recent experimental results summarized above with J. W. Beams's experiments on gold and silver films, in which the size effect did not reverse even for $h=20$ nm [15,16].

Our work joins a growing body of research in which the surface stress has been found to play a crucial role in several problems at ultrasmall lengthscales, including the blunting of a crack tip [17] and the nanoindentation of a crystal [18].

2 Surface Stress

The surface stress is a second-rank tensor $T_{\alpha\beta}$, where the indices α and β run from 1 to 2 and denote in-plane coordinates defined on the surface. (For detailed discussions of the surface stress see, for example, [19,20].) To relate the surface stress to the surface energy γ using Eulerian coordinates [21], consider an element of surface of area A and energy γA that is stretched by an in-plane *elastic* strain $\varepsilon_{\alpha\beta}$. Then, the work performed by the surface stress is $dW=AT_{\alpha\beta}\varepsilon_{\alpha\beta}$, and the energy γA changes by $d(\gamma A)=\gamma dA+Ad\gamma=\gamma A\varepsilon_{\mu\mu}+A(\partial\gamma/\partial\varepsilon_{\alpha\beta})\varepsilon_{\alpha\beta}$, where repeated indices imply summation. Equating dW to $d(\gamma A)$ results in the desired expression for the surface stress, $T_{\alpha\beta}=\gamma\delta_{\alpha\beta}+\partial\gamma/\partial\varepsilon_{\alpha\beta}$, where $\delta_{\alpha\beta}$ is the second-rank Kroenecker delta. The second term in the expression for $T_{\alpha\beta}$ represents the change in surface energy associated with the elastic stretching of the surface. When the area of a

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received April 30, 2005; final manuscript received July 19, 2005. Review conducted by H. D. Espinosa. 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.

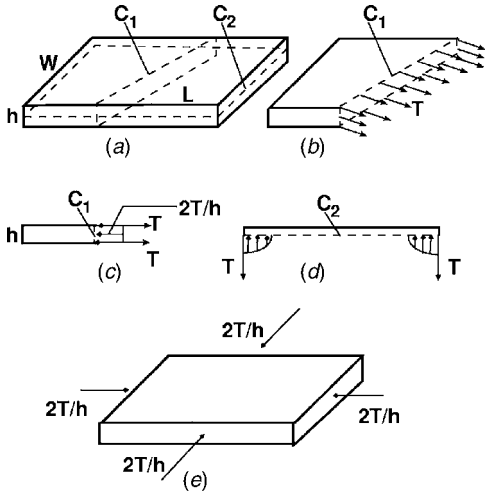


Fig. 1 (*a*) **A free-standing thin film.** C_1 **and** C_2 **are cuts performed for stress analysis.** (*b*) **The surface stress** T **acting on the perimeter of** C_1**.** (*c*) **The compressive stress induced by** T **on the surface of** C_1**.** (*d*) **The compressive stress induced by** T **on the surface of** C_2**.** (*e*) **Applied traction that gives the same stresses as** T**.**

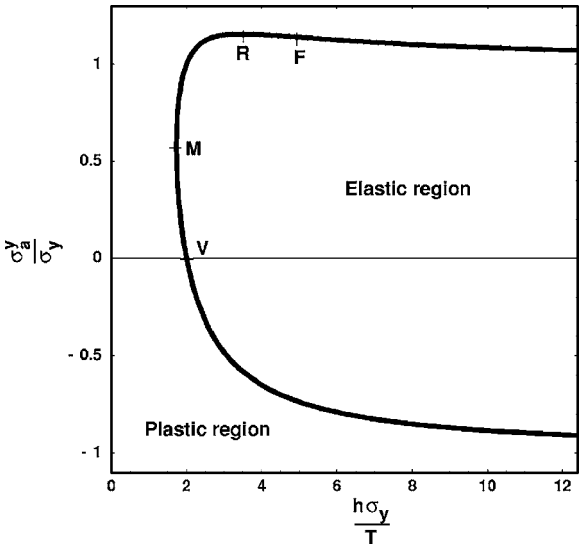


Fig. 2 A plot of the dimensionless apparent yield stress σ_a^y/σ_y **versus the dimensionless thickness** $h\sigma_y/T$**. See Eq. (1). The points** *F*, *R*, *M*, **and** *V* **are referred to in the text. The size effect of the apparent tensile yield stress reverses from hardening to softening at the point R.**

liquid surface is increased, the surface does not stretch elastically, because the atoms in the interior of the liquid are mobile and may readily migrate to the surface. Thus in liquids the second term in the expression for $T_{\alpha\beta}$ vanishes, and the surface stress is isotropic and equal to the surface energy. This is not the case in solids, because in solids the long-range order in the positions of the atoms makes it infeasible for the atoms to migrate to the surface, in particular when the strain applied to the surface amounts to a displacement of a small fraction of the lattice constant. Thus in solids the second term in the expression for $T_{\alpha\beta}$ may not necessarily vanish, and the surface stress is in general anisotropic.

Consider now the crystalline surface of a free-standing crystal. The surface accommodates the lattice constant of the bulk of the material by means of a spontaneous elastic stretching. If the crystalline surface possesses a threefold or higher rotational axial symmetry, then in the expression for $T_{\alpha\beta}$ the term $\partial\gamma/\partial\varepsilon_{\alpha\beta}$ associated with this elastic stretching is isotropic [13]. That is the case for (111) surfaces in fcc metals. In the experiments of interest here the films had a preferred $\langle 111\rangle$ crystallographic texture normal to the faces of the films, and we are justified in assuming an isotropic surface stress, $T_{\alpha\beta}=T\delta_{\alpha\beta}$. The surface stress may in principle be either positive (tensile) or negative, but it is positive for fcc metals.

3 Surface Stress in Thin Films

Consider a free-standing film of length $L\gg h$ and width $W\gg h$ [Fig. 1(*a*)]. Suppose that the film is severed through its thickness along an arbitrary in-plane direction. [The cut is marked C_1 in Fig. 1(*a*)]. Then, the surface stress, which we assume to be positive and isotropic, becomes manifest as a tensile force T per unit length of the perimeter of the cut, acting normal to the surface of the cut, as indicated in Fig. 1(*b*). If the severed parts of the film are to remain in equilibrium, the surface stress must induce a compressive stress on the surface of the cut; because the film is very thin, the induced stress is uniform and of value $-2T/h$ on the surface of the cut (Fig. 1(*c*)). Thus the surface stress induces a compressive stress of value $-2T/h$ in all in-plane directions [19].

Suppose now that the film is severed parallel to its upper and lower faces. (The cut is marked C_2 in Fig. 1(*a*)). Then, the surface stress must again induce a compressive stress on the surface of the cut, but now the stress is confined to a very narrow strip (of width $\sim h$) parallel to the lateral edges of the film, as indicated in Fig. 1(*d*). Thus in most of the film the surface stress induces no stress in the direction of the thickness of the film.

From our discussion so far, we conclude that in a film of thickness h the stresses induced by the surface stress may be approximately simulated by (i) applying an in-plane compressive traction of value $-2T/h$ on all the lateral edges of the film and (ii) leaving the upper and lower faces of the film traction-free (Fig. 1(*e*)). This conclusion is valid where the film is thin, i.e., where $L\gg h$ and $W\gg h$.

4 Apparent Yield Stress

Consider now a free-standing thin film to which a uniaxial stress σ_a is applied in the direction of the length of the film. Then, the film is uniformly subjected to principal stresses $\sigma_1=\sigma_a-2T/h$, $\sigma_2=-2T/h$, and $\sigma_3=0$ in the direction of the length, width, and thickness, respectively. We may ascertain the value of the applied stress at initial yielding or *apparent yield stress* σ_a^y by substituting the principal stresses in the von Mises yield condition, $2\sigma_y^2=(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2$, where σ_y is the yield stress [22]. The result is

$$\frac{\sigma_a^y}{\sigma_y}=\frac{T}{h\sigma_y}\pm\sqrt{1-3\left(\frac{T}{h\sigma_y}\right)^2}. \tag{1}$$

Figure 2 shows a graphical rendition of Eq. (1) in the form of a plot of the dimensionless apparent yield stress σ_a^y/σ_y versus the dimensionless thickness $h\sigma_y/T$. In the plot there is a single curve separating the elastic region (which the curve embraces) from the plastic region. The curve consists of two branches touching at their leftmost points (marked M in Fig. 2). The upper branch corresponds to the $+$ sign in Eq. (1) and gives the apparent tensile yield stress. On the other hand, the lower branch corresponds to the $-$ sign in Eq. (1) and gives the apparent compressive yield stress. The branches are supported on $h\geqslant h_M\equiv\sqrt{3}T/\sigma_y$ [because the discriminant of Eq. (1) is negative for $h<h_M$]. Therefore, a film of thickness $h<h_M$ cannot be poised between the elastic region and the plastic region, regardless of the applied stress; such a film is always in the plastic region.

Figure 2 indicates that for $h\gg T/\sigma_y$ the apparent tensile yield stress is σ_y, and the apparent compressive yield stress $-\sigma_y$. Thus

for $h \gg T/\sigma_y$ the initial yielding may be attained by applying a tensile stress σ_y or a compressive stress $-\sigma_y$. This is the expected asymptotic behavior.

Next, we discuss in turn the two branches of Fig. 2. Consider first the lower branch, starting with a thin film of thickness $h \gg T/\sigma_y$. If the thickness of the film diminishes, the absolute value of the apparent compressive yield stress decreases (i.e., $d|\sigma_a^y|/dh > 0$), and we say that there is a softening size effect. If the thickness continues to diminish, then, for a thickness $h=h_V \equiv 2T/\sigma_y$, the apparent compressive yield stress vanishes, $\sigma_a^y=0$ (point V in Fig. 2). Thus a free-standing film of thickness $h=h_V$ spontaneously attains the initial yielding by virtue of the compressive stresses induced by the surface stress. When a tensile stress is applied to this free-standing film, the film departs from the initial yielding, enters the elastic region, and re-attains the initial yielding at the apparent tensile yield stress given by the upper branch of Fig. 2. If the thickness continues to diminish beyond h_V, the apparent compressive yield stress becomes *positive* (i.e., the lower branch of Fig. 2 gives a *positive* value of σ_a^y). This positive value of σ_a^y is the minimum tensile stress that must be applied to the film if the film is not to yield by virtue of the *compressive* stresses induced by the surface stress. (Therefore, the name "apparent *compressive* yield stress" remains appropriate, even though this stress is positive.) Last, for a thickness $h=h_M \equiv \sqrt{3}T/\sigma_y$, we reach the leftmost point of the lower branch.

Consider now the upper branch of Fig. 2, starting with a thin film of thickness $h \gg T/\sigma_y$. If the thickness of the film diminishes, the apparent tensile yield stress increases (i.e., $d|\sigma_a^y|/dh < 0$), and we say that there is a hardening size effect. If the thickness of the film continues to diminish, then, for a thickness $h=h_F \approx 5T/\sigma_y$, the size effect is at its most hardening (i.e., $d^2|\sigma_a^y|/dh^2=0$; point F in Fig. 2). If the thickness continues to diminish beyond h_F, then the hardening size effect starts to lessen (i.e., $d|\sigma_a^y|/dh$ starts to become less negative). Eventually, for a thickness $h=h_R \equiv 2\sqrt{3}T/\sigma_y \approx 3.5T/\sigma_y$, the apparent tensile yield stress attains its maximum value $\sigma_a^y=\sigma_{aR}^y \equiv 2\sigma_y/\sqrt{3} \approx 1.15\sigma_y$ and the size effect vanishes (i.e., $d|\sigma_a^y|/dh=0$; point R in Fig. 2). If the thickness continues to diminish beyond h_R, the apparent tensile yield stress decreases (i.e., $d|\sigma_a^y|/dh > 0$), and we say that there is a softening size effect. Thus for a thickness $h=h_V \equiv 2T/\sigma_y$ the apparent tensile yield stress equals its original value $\sigma_a^y=\sigma_y$. Last, for a thickness $h=h_M \equiv \sqrt{3}T/\sigma_y \approx 1.73T/\sigma_y$, the apparent tensile yield stress equals its minimum value $\sigma_a^y=\sigma_{aM}^y \equiv \sigma_y/\sqrt{3} \approx 0.58\sigma_y$ and we reach the leftmost point of the upper branch.

From our discussion of Fig. 2 we conclude that the surface stress causes a size effect on the apparent tensile yield stress. For thin films of thickness $h \gg T/\sigma_y$ there is a hardening size effect, but the size effect reverses from hardening to softening for a thickness $h_R \equiv 2\sqrt{3}T/\sigma_y$. Thus the stresses induced in a thin film by the surface stress lead to a size effect of the type recently observed in experiments.

5 Size Effects and the Yield Condition

We have predicated Eq. (1) on the von Mises yield condition $\sigma_e=\sigma_y$. Here σ_e, the *equivalent stress*, quantifies the forcing that tends to produce plastic deformation; it is defined by the expression $2\sigma_e^2 \equiv (\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2$, where σ_1, σ_2, and σ_3 are the principal stresses. This expression for σ_e suitably quantifies the forcing if the plastic deformation occurs by the relative slip of adjacent planes in the material, regardless of the specific mechanisms whereby the slip is effected. (The slip need not be affected by dislocation motion, for example.) In fact, it is the yield stress σ_y and not the equivalent stress σ_e that depends on the specific mechanisms whereby the slip is effected. Thus the elucidation of size effects consists in determining the dependence of σ_y on the size. Yet this is not the case for the size effect caused by the surface stress, because this size effect is unrelated to the material. Instead, it is related to the stresses that act *on the bulk of the material* and to the fact that these stresses differ from the applied stresses. The elucidation of this size effect is not a problem in materials science, but rather a problem in solid mechanics.

In Eq. (1) the size effect caused by the surface stress is coupled to other size effects only through the value of σ_y. Therefore, *in Eq. (1) the yield stress σ_y is not the yield stress of the bulk material, but the yield stress of the bulk material enhanced by any pertinent size effects other than the size effect caused by the surface stress*.

6 Comparison with Experiments

In Sec. 4, we concluded that the size effect caused by the surface stress reverses from hardening to softening for a thickness $h_R \equiv 2\sqrt{3}T/\sigma_y$. To compare the predicted value of h_R with the experimental results, we recall that for $h=h_R$ the (maximum) apparent tensile yield stress is $\sigma_{aR}^y \equiv 2\sigma_y/\sqrt{3}$, and write an expression for h_R in terms of σ_{aR}^y, with the result $h_R=4T/\sigma_{aR}^y$.

(Note that $\sigma_y=\sigma_{aR}^y\sqrt{3}/2$ is the yield stress of the bulk material enhanced by any size effects other than the size effect caused by the surface stress; see Sec. 5. Note also that the expression $h_R=4T/\sigma_{aR}^y$ can give only a rough estimate of the thickness for which the observed size effect reverses, not only because we have predicated this expression on a number of simplifying assumptions, but also because (a) the value of T may be strongly affected by subtle changes in environmental conditions and (b) σ_{aR}^y is difficult to measure, and tends to be overestimated both in experiments and in atomistic simulations; see, e.g., [23].)

For the pure gold films of the experimental study of Ref. [11], the reported maximum apparent tensile yield stress was $\sigma_{aR}^y=170$ MPa. Using the surface stress of gold given in Ref. [13], $T=3.42$ N/m (see the Appendix), we compute $h_R=80$ nm, which is on the order of magnitude of the thickness for which the observed size effect reversed in that study, $h \approx 400$ nm.

For the pure aluminum thin films of the experimental study of Ref. [12], the reported *peak* stress was $\sigma_a^p=750$ MPa, and we estimate $\sigma_{aR}^y=\sigma_a^p/2=375$ MPa. Using the surface stress of aluminum given in Ref. [13], $T=2.29$ N/m (see the Appendix), we compute $h_R=24$ nm, which is on the order of magnitude of the thickness for which the observed size effect reversed in that study, $h \approx 100$ nm.

7 Failure and the Ductile-to-Brittle Transition

Upon attaining the initial yielding, the bulk of the film undergoes permanent deformation in the form of plastic strain increments $\Delta\varepsilon_1=s_1\Delta\lambda$, $\Delta\varepsilon_2=s_2\Delta\lambda$, and $\Delta\varepsilon_3=s_3\Delta\lambda$ in the direction of the length, width, and thickness of the film, respectively [22]. Here $\Delta\lambda$ is a dimensionless scalar factor, $s_1=(\sigma_1-p)/\sigma_y$, $s_2=(\sigma_2-p)/\sigma_y$, $s_3=(\sigma_3-p)/\sigma_y$, and $p=(\sigma_1+\sigma_2+\sigma_3)/3$. (Note that the plastic deformation is isochoric, $\Delta\varepsilon_1+\Delta\varepsilon_2+\Delta\varepsilon_3=0$.) By substituting $\sigma_1=\sigma_a-2T/h$, $\sigma_2=-2T/h$, $\sigma_3=0$, and $\sigma_a=\sigma_a^y$ (where σ_a^y is the apparent tensile yield stress given by the upper branch of Fig. 2), we obtain

$$s_1 = \frac{2}{3}\sqrt{1-3\left(\frac{T}{h\sigma_y}\right)^2}, \tag{2}$$

$$s_2 = -\frac{T}{h\sigma_y} - \frac{1}{3}\sqrt{1-3\left(\frac{T}{h\sigma_y}\right)^2}, \tag{3}$$

and

$$s_3 = \frac{T}{h\sigma_y} - \frac{1}{3}\sqrt{1-3\left(\frac{T}{h\sigma_y}\right)^2}. \tag{4}$$

Figure 3 shows a graphical rendition of Eqs. (2)–(4) in the form of plots of the dimensionless quantities s_1, s_2, and s_3 versus the di-

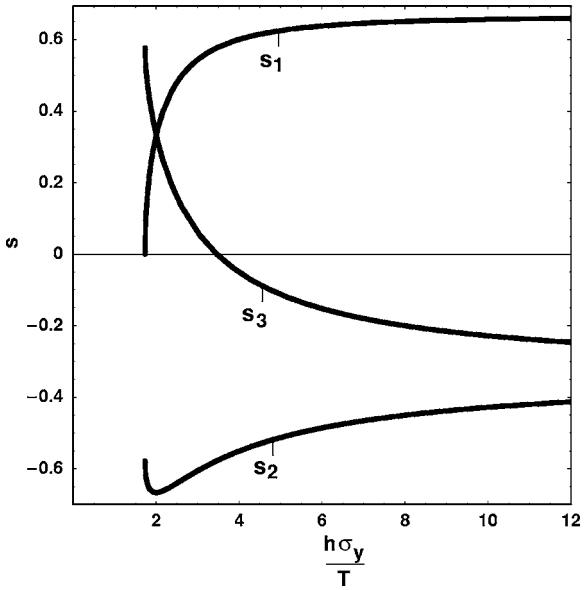


Fig. 3 A plot of the dimensionless quantities s_1, s_2, and s_3 versus the dimensionless thickness, $h\sigma_y/T$. See Eqs. (2)–(4).

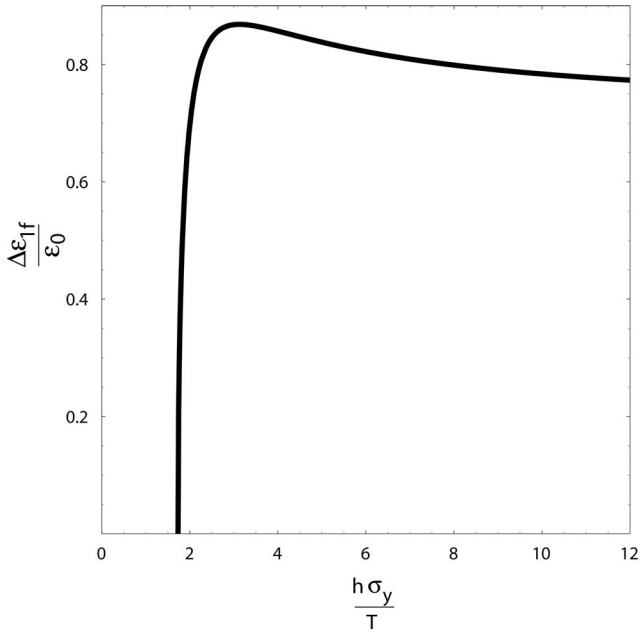


Fig. 4 A plot of the normalized plastic strain at failure in the direction of the applied stress versus the dimensionless thickness, $h\sigma_y/T$. See Eq. (6).

mensionless thickness $h\sigma_y/T$. As was the case for Fig. 2, the plots in Fig. 3 are supported on $h \gg h_M \equiv \sqrt{3}T/\sigma_y$.

Consider now the process whereby the film accumulates plastic deformation, eventually leading to failure in the form of fracture. According to a well-known failure criterion [14], the onset of failure occurs when the *equivalent plastic strain* ε_e attains a critical value ε_{ef} that depends on the triaxiality of the stress in the form

$$\varepsilon_{ef} = \varepsilon_0 \exp(-p/\sigma_e), \tag{5}$$

where the subscript "*f*" stands for "at failure," ε_0 is a dimensionlesss constant, p/σ_e is a measure of the triaxiality of the stress, and the equivalent plastic strain is defined by the expression $9\varepsilon_e^2/2=(\Delta\varepsilon_1-\Delta\varepsilon_2)^2+(\Delta\varepsilon_2-\Delta\varepsilon_3)^2+(\Delta\varepsilon_3-\Delta\varepsilon_1)^2$. By evaluating ε_e with $\Delta\varepsilon_1=s_1\Delta\lambda$, $\Delta\varepsilon_2=s_2\Delta\lambda$, $\Delta\varepsilon_3=s_3\Delta\lambda$, and the expressions for s_1, s_2, and s_3 given by Eqs. (2)–(4), we obtain $\varepsilon_e=2\Delta\lambda/3$; therefore, the value of $\Delta\lambda$ at failure is $\Delta\lambda_f=3\varepsilon_{ef}/2$. On the other hand, by setting $\sigma_e=\sigma_y$ and evaluating p with $\sigma_1=\sigma_a-2T/h$, $\sigma_2=-2T/h$, $\sigma_3=0$, and $\sigma_a=\sigma_a^y$ (where σ_a^y is the apparent tensile yield stress given by the upper branch of Fig. 2), we obtain $p/\sigma_e=-s_3$. Since $\Delta\lambda_f=3\varepsilon_{ef}/2$ and $p/\sigma_e=s_3$, we can recast Eq. (5) in the form $\Delta\lambda_f=(3/2)\varepsilon_0\exp(s_3)$, and write an expression for the plastic strain at failure *in the direction of the applied stress*, $\Delta\varepsilon_{1f}$, as follows:

$$\Delta\varepsilon_{1f} = (3/2)\varepsilon_0 s_1 \exp(s_3). \tag{6}$$

Figure 4 shows a graphical rendition of Eq. (6) in the form a plot of $\Delta\varepsilon_{1f}/\varepsilon_0$ versus the dimensionless thickness $h\sigma_y/T$. From the plot in Fig. 4, we conclude that the plastic strain at failure in the direction of the applied stress (a measure of the apparent ductility of the film) reaches a maximum thickness $h=h_R$, it diminishes as the film becomes thinner, and it vanishes for a thickness $h=h_m$.

Our conclusions from the previous paragraph are in accord with the results of recent experiments on gold films [11], in which a ductile-to-brittle transition was documented for a thickness on the order of 100 nm.

8 Biaxial Loading

To inquire further into the size effect caused by the surface stress, we now consider a type of experiment known as the *bulge test*. In the bulge test, a film of thickness h is placed across the open end of a circular tube of radius $R \gg h$. Then, the pressure of the air in the tube is increased to a value P, whereupon the film deflects to form a bulge of height $b \ll R$. As a result, the film is subjected to a *biaxial*, in-plane isotropic applied stress $\sigma_a = PR^2/4bh \gg P$. In a classic experimental study, J. W. Beams used the bulge test to determine the apparent yield stress of polycrystalline gold and silver films of thicknesses in the range of 200 to 20 nm [15,16]. He concluded that the apparent yield stress increased monotonically with diminishing film thickness. Thus, in contrast with the recent experimental studies summarized above, a reversing size effect was *not* observed in Beam's classical experimental study.

To understand this discrepancy, we substitute the principal stresses of the bulge test ($\sigma_1=\sigma_a-2T/h$, $\sigma_2=\sigma_a-2T/h$, and $\sigma_3=0$) in the von Mises yield condition, with the result

$$\frac{\sigma_a^y}{\sigma_y} = \pm 1 + 2\frac{T}{h\sigma_y}, \tag{7}$$

where σ_a^y is the apparent yield stress [24][1]. Equation (7) predicts a hardening size effect in the apparent tensile yield stress measured in bulge tests, as expected. In a 1959 review paper [16], Beams is said to have first ascribed the hardening size effect observed in his experiments to the surface *energy*. In the same paper, a plot is shown and attributed to Beams which might be a graphical rendition of Eq. (7). Unfortunately, Beams appears not to have published the equations leading to this plot. (He subsequently came to the conclusion that the size effect observed in his experiments could not be ascribed to the surface energy, because for relatively thick thin films the predicted size effect fell short of the observed size effect [16]. It was thought at the time that the observed size effect should be ascribed to a *single* reason.)

From our discussion of the bulge test we conclude that, in contrast with our results for thin films subjected to a uniaxial applied stress, the surface stress does not lead to a reversing size effect in thin films subjected to a biaxial, isotropic applied stress.

[1]It is straightforward to show that this same size effect is valid for wires. For atomistic simulations in ultrathin wires, see, for example, Ref. [24].

9 Discussion

We have concluded that in ultrathin, polycrystalline metallic films the surface stress leads to a size effect in the initial yielding that depends strongly on the applied stress, in accord with experiments. Where the applied stress is uniaxial, the size effect reverses for a film thickness h_R that can be estimated using values of the surface stress determined via atomistic methods. The result, $h_R \approx 100$ nm is in accord with experiments. In addition, we have predicted that the mode of failure of the film changes from ductile to brittle for thicknesses close to h_R, also in accord with experiments.

To reach these conclusions, we have used continuum mechanics. Given that the film thickness for which the size effect reverses is only about 100 times a typical lattice parameter, our conclusions add to a growing realization of the robustness of continuum mechanics at ultrasmall lengthscales, a realization that has been commented upon by a number of authors. (For a recent, eloquent example from the field of microfluidics see [25].)

Acknowledgment

The authors thank Professor K. Jimmy Hsia for a number of discussions and for his encouragement in the course of this research. Professor James W. Phillips kindly read our manuscript and suggested ways of improving it.

Appendix

In the calculations of Sec. 6 we use the (111) unrelaxed surface stress computed by the modified embedded atom method and reported in [13]. (In this useful reference, the surface stresses and surface energies obtained by a number of methods are given for all fcc metals: Cu, Ag, Au, Ni, Pd, Pt, Al, and Pb. Different methods lead to comparable results, and the results are in good agreement with the few available experimental measurements.) Note that this surface stress corresponds to a free-standing crystal and does not account for the additional elastic stretching undergone by the surface (as well as by the bulk of the material) as the film is stressed to the initial yielding. The required correction is negligible, however. In fact, a straightforward application of the atomistic model of Nix and Gao [21] allows us to estimate the required correction as $\Delta T \approx 2E\varepsilon a$, where E is the Young's modulus, ε is the strain associated with the additional elastic stretching, and a is the lattice constant; for gold we use $E\varepsilon = \sigma_y = 170$ MPa and $a = 0.3$ nm, with the result $\Delta T \approx 0.11$ N/m $\ll T = 3.41$ N/m.

Note also that a small increment in the *plastic* deformation brings additional atoms to the surface of the film but does not cause an additional elastic stretching of the surface (or of the bulk of the material) [21]. We conclude that the area of the surface of a film may change as a result of a small increment in plastic deformation, but the surface retains the same structure and remains equally stretched, so that the energy of the surface changes by TdA, where T is the surface stress of the free-standing film, and dA is the change in surface area. As an example of application of this conclusion, consider a film that undergoes plastic strain increments $d\varepsilon_1$, $d\varepsilon_2$, and $d\varepsilon_3$ in the direction of L, W, and h, respectively, where $d\varepsilon_3 = -(d\varepsilon_1 + d\varepsilon_2)$. The energy of the surface changes by $dW_s = TdA = T[2LW(d\varepsilon_1 + d\varepsilon_2) + 2(L+W)hd\varepsilon_3] = 2T[LW - (L+W)h](d\varepsilon_1 + d\varepsilon_2)$, the stresses in the bulk of the film perform a plastic work $dW_p = (\sigma_1 d\varepsilon_1 + \sigma_2 d\varepsilon_2 + \sigma_3 d\varepsilon_3)LWh$, and the applied traction performs a work $dW_a = \sigma_a LWhd\varepsilon_1$. Equating $dW_a = dW_s + dW_p$ leads to $\sigma_1 = \sigma_a - 2T/h[1 - (1 + L/W)h/L]$, $\sigma_2 = -2T/h[1 - (1 + L/W)h/L]$, and $\sigma_3 = 0$, which under the assumption $h/L \ll 1$ simplifies to $\sigma_1 = \sigma_a - 2T/h$, $\sigma_2 = -2T/h$, and $\sigma_3 = 0$, as we concluded before under the same assumption.

References

[1] Bažant, Z. P., and Chen, E. P., 1997, "Scaling of Structural Failure," Appl. Mech. Rev., **10**, pp. 593–527.

[2] Brenner, S. S., 1956, "Tensile Strength of Whiskers," J. Appl. Phys., **27**, pp. 1484–1491.

[3] 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.

[4] Fleck, N. A., and Hutchinson, J. W., 1997, "Strain Gradient Plasticity," Adv. Appl. Mech., **33**, pp. 295–261.

[5] Lejeck, P., and Sima, V., 1983, "Orientational Relationships in the Secondary Recrystallization of Pure Nickel," Mater. Sci. Eng., **60**, pp. 121–124.

[6] Grant, E. M., Hansen, N., Jensen, D. J., Ralph, B., and Stobbs, W. M., 1988, "Texture Development During Grain Growth in Thin Films," *Proceedings of the Eighth International Conference on Texture of Materials*, J. S. Kallend and G. Gottstein, eds., Springer-Verlag, New York.

[7] Griffin, A. J., Brotzen, F. R., and Dunn, C. F., 1987, "Mechanical-Properties and Microstructures of Al-1-Percent-Si Thin-Film Metallizations," Thin Solid Films, **150**, pp. 237–244.

[8] Venkatraman, R., and Bravman, J. C., 1992, "Separation of Film Thickness and Grain Boundary Strengthening Effects in Al Thin Films on Si," J. Mater. Res., **7**, pp. 2040–2048.

[9] Thompson, C. V., 1993, "The Yield Stress of Polycrytalline Thin Films," J. Mater. Res., **8**, pp. 237–238.

[10] Schiotz, J., and Jacobsen, K. W., 2003, "A Maximum in the Strength of Nanocrystalline Copper," Science, **301**, pp. 1357–1359.

[11] Espinosa, H. D., Prorok, B. C., and Peng, B., 2004, "Plasticity Size Effects in Free-Standing Submicron Polycrystalline FCC Films Subjected to Pure Tension," J. Mech. Phys. Solids, **52**, pp. 667–689.

[12] Saif, T., 2004, "Scaling the Depths," Mech. Eng. (Am. Soc. Mech. Eng.), **126**, pp. 8–11. See also Haque, A., 2002, "Length-Scale Effects on Nano-Scale Materials Behavior," Ph.D. thesis, Department of Mechanical Engineering, University of Illinois at Urbana-Champaign.

[13] Wan, J., Fan, Y. L., Gong, D. W., Shen, S. G., and Fan, X. Q., 1999, "Surface Relaxation and Stress of FCC Metals: Cu, Ag, Au, Ni, Pd, Pt, Al and Pb," Modell. Simul. Mater. Sci. Eng., **7**, pp. 189–206.

[14] Hancock, J. W., and Mackenzie, A. C., 1976, "On the Mechanisms of Ductile Failure in High-Strength Steels Subjected to Multi-Axial Stress States," J. Mech. Phys. Solids, **24**, pp. 147–169.

[15] Beams, J. W., 1959, "Mechanical Properties of Thin Films of Gold and Silver," in *Structure and Properties of Thin Films*, C. A. Neugebauer, C. A. Newkirk, and D. A. Vermilyea, eds., Willey, New York, pp. 183–198.

[16] Menter, J. W., and Pashley, D. W., 1959, "The Microstructure and Mechanical Properties of Thin Films," in *Structure and Properties of Thin Films*, C. A. Neugebauer, C. A. Newkirk, and D. A. Vermilyea, eds., Willey, New York, pp. 111–150.

[17] Carlsson, A. E., and Thomson, R., 1988, "Fracture Toughness of Materials: From Atomistics to Continuum Theory," Solid State Phys., **51**, pp. 233–280.

[18] Knap, J., and Ortiz, M., 2003, "Effect of Indenter-Radius Size on Au(001) Nanoindentation," Phys. Rev. Lett., **90**, pp. 226102.

[19] Herring, C., 1953, in *Structure and Properties of Solid Surfaces*, R. Gomer, and C. S. Smith, edds., The University of Chicago Press, Chicago, IL.

[20] Cammarata, R. C., 1994, "Surface and Interface Stress Effects in Thin-Films," Prog. Surf. Sci., **46**, pp. 1–38.

[21] Nix, W. D., and Gao, H., 1998, "An Atomistic Interpretation of Interface Stress," Scr. Mater., **39**, pp. 1653–1661.

[22] Calladine, C. R., *Plasticity*, *Horwood Publishing Ltd.*, Chichester, UK, 2000, p. 48.

[23] Schiotz, J., Vegge, T., Di Tolla, F. D., and Jacobsen, K. W., 1999, "Atomic-Scale Simulations of the Mechanical Deformation on Nanocrystalline Metals," Phys. Rev. B, **60**, pp. 11971–11983.

[24] Gall, K., Diao, J., and Dunn, M. L., 2004, "The Strength of Gold Nanowires," Nano Lett., **4**, pp. 2431–2436.

[25] Sharp, K. V., and Adrian, R. J., 2004, "Transition From Laminar to Turbulent Flow in Liquid Filled Microtubes," Exp. Fluids, **36**, pp. 741–747.

Control of Bending Vibrations Within Subdomains of Thin Plates—Part II: Piezoelectric Actuation and Approximate Solution

Michael Krommer
Institute for Technical Mechanics,
Johannes Kepler University Linz,
Altenbergerstr. 69,
A-4040 Linz, Austria
e-mail: krommer@mechatronik.uni-linz.ac.at

Vasundara V. Varadan
Department of Electrical Engineering,
University of Arkansas,
3217 Bell Engineering Center,
Fayetteville, 72701
e-mail: vvvesm@engr.uark.edu

In the first part of this paper, we presented the theoretical basics of a new method to control the bending motion of a subdomain of a thin plate. We used continuously distributed sources of self-stress, applied within the subdomain, to exactly achieve the desired result. From a practical point of view, continuously distributed self-stresses cannot be realized. Therefore, we discuss the application of discretely placed piezoelectric actuators to approximate the continuous distribution in this part. Using piezoelectric patch actuators requires the consideration of electrostatic equations as well. However, if the patches are relatively thin, the electromechanical coupling can be incorporated by means of piezoelastic (instead of elastic) stiffness (piezoelastically stiffended elastic constants). The placement of the patches is based on the discretization of the exact continuous distribution by means of piece-wise constant functions. These are calculated from a convolution integral representing the deviation of the bending motion in the controlled case from the desired one. A proper choice of test loadings allows us to eliminate representative mechanical quantities exactly and to make the resulting bending motion to match the desired one very closely; hence, to find a suboptimal approximate solution. In Part I of this paper we presented exact solutions for the axisymmetric bending of circular plates; it is also considered in Part II. For axisymmetric bending, only the radial coordinate is discretized. Hence, ring-shaped piezoelectric patch actuators are considered in this paper. [DOI: 10.1115/1.2083790]

1 Introduction

Shape control of structures is concerned with methods that will result in a desired shape by applying a suitable actuation. A desired shape may be a prescribed new shape or may be the nondeformed shape for a structure that is under the influence of external disturbances. No matter what the actual desired shape is, the problem formulation will always result in the question: given an external disturbance, how does a suitable distribution of the actuation result in the structure assuming the desired shape?

In a fundamental contribution, Haftka and Adelman [1] were the first to introduce the notion of shape control into the journal literature. They noted that disturbances that affect the shape of the structure could be subdivided into two types; one type is transient, whereas the other type is associated with fixed deformations or those that vary slowly in time. The first refers to dynamic shape control, while the latter refers to static shape control. Haftka and Adelman addressed the problem of static shape control of a large spacecraft structure by applying temperature as an actuating mechanism. Irschik and Pichler [2] reported on results for dynamic shape control of solids and structures by applied thermal expansion strains. Irschik [3] gave a detailed discussion and review of both static and dynamic shape control using piezoelectric eigenstrains as the actuation mechanism. In addition to thermal expansion strains and piezoelectric strains, other actuation methods may be used. However, all of these actuation methods belong to the general class of what is usually referred to as eigenstrain actuation or self-stress actuation methods. Early reports on self-stresses were given by Reissner [4] and Nemenyi [5]. Presently, self-stresses are discussed in connection with micromechanics of solids (Mura [6]) and, more recently, in connection with shape control and active (noise) control of structures (Rao and Sunar [7], Gopinathan et al. [8], Irschik [3]). Hence, a unified approach can be used for all members of the general class of self-stress actuation methods.

In Part I of this paper (Krommer and Varadan [9]), we presented the theoretical basics of a new method to control the bending vibrations of a subdomain of a thin plate; suitably distributed sources of self-stress were only applied in the subdomain itself. We calculated exact solutions in the framework of thin circular plates, which bend into an axisymmetric surface. We considered control of arbitrary subdomains; however, we did not pay any attention to the problem of how to practically achieve the required distributed control. We only assumed the control agency to be any possible type of sources of self-stress. In Part II of this paper, we especially focus on this latter practical aspect of the problem; especially, on the use of piezoelectric patch actuators and on how to calculate values for the actuation to be applied at these patches resulting in a suboptimal solution, when compared to the exact one.

We begin the second part of our paper with an introduction to the use of piezoelectric patch actuators. Piezoelectricity is well known in the engineering of smart materials and structures. Typically, piezoelectric materials, which exhibit the direct piezoelectric effect and the converse piezoelectric effect, are used to design sensors and actuators. An enormous amount of literature is avail-

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received May 9, 2005; final manuscript received June 2, 2005. Review conducted by R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

able, including many review papers. Recently published reviews are the papers by Rao and Sunar [7], Tani et al. [10], and Tzou [11]. It is well known that a continuously distributed piezoelectric actuation is not easy to be achieved. The only way reported in the literature is to vary the polarization profile of the piezoelectric material. However, the use of discretely acting piezoelectric patch actuators is more practical. For an early discussion on distributed actuation and sensing, see the fundamental paper by Lee [12] on the theory of thin piezoelectric composite plates with actuators and sensors. We discuss piezoelectric patches of arbitrary geometry, which are fully equipped with electrodes at their horizontal surfaces. Approximations with respect to which component of the electric–field vector is predominant result into a simple electromechanically coupled thin plate theory, which we apply in this part of the paper.

We present a brief summary of the theoretical basics of our methodology as developed in Part I; the goal is to control the bending motion of a subdomain of a linear elastic plate by applying control by means of self-stress actuation. The result of Part I was a continuously distributed self-moment tensor, which represents an exact solution of the problem. In Part II, we focus on the implementation of this actuation by using piezoelectric patch actuators. If we have to use discretely acting patch actuators, we are not able to achieve the desired control goal for the subdomain exactly. Therefore, Part II of this paper is mainly devoted to the problem of how to discretize the exact solution of Part I.

We use three methods of discretization in this paper to approximate the exact solution. The first method, which we denote as the direct method, finds discrete values for the self-moment tensor by demanding that the area underneath the distributed self-moment match the area underneath the piece-wise constant self-moment. The direct method lacks a mechanical interpretation of the resulting bending motion within the controlled subdomain, but allows the use of a high number of actuators to closely match the distributed self-moment. The second method, denoted as the indirect method, uses the same convolution integral we used in Part I; however, we do not ask the kernel of the integral to vanish, but only the integral itself. This is achieved by using special dummy loading cases instead of arbitrary ones. The advantage of this method is the mechanical interpretation of the resulting bending motion in terms of deflection, slope, mean value of deflection, as well as higher-order moments of the deflection. The disadvantage is that we have to use a high number of dummy loadings in order to use a high number of actuators. The third method combines the direct method and the indirect method, in which the advantages of both methods are reflected.

We present results for the axisymmetric bending motion of a thin circular plate, as we did in Part I. A comparison between the direct method and the indirect method is presented. The combined method is used as a tool to calculate a suboptimal solution, which eliminates characteristic mechanical quantities exactly and which matches the exact solution very closely.

We conclude this introduction by a short discussion on the practical relevance of controlling subdomains of structures. Our motivation for starting to work on this problem was related to the problem of controlling the shape of conformal antennas. These antennas are conformally mounted on vehicles that may undergo deformations large enough to interfere with the antenna's performance. A conformal antenna will deform with the vehicle as it moves. Dynamic shape control of such surfaces and structures is important to preserve integrity of the electromagnetic performance of the antenna. Typically, only that part of the vehicle or structure, to which the antenna is mounted, has to be controlled for the proposed goal; hence, the control of a subdomain of the structure can be applied directly to this problem. First results for the antenna problem can be found in Krommer and Varadan [13]; in this previous work, we also focused on aspects of practicality; however, the underlying plate theory was electromechanically decoupled and we did not use discretely-acting piezoelectric patches. Other applications belong to the field of noise reduction in structures (Gopinathan et al. [8]). In these applications, it is often not necessary to control the whole structure, but a significant noise reduction can be achieved by focusing on critical parts of the structure. An example would be the funnel of a magnetic resonance imaging (MRI) unit, which is responsible for the highly annoying noise experienced by patients (Nader et al. [14]).

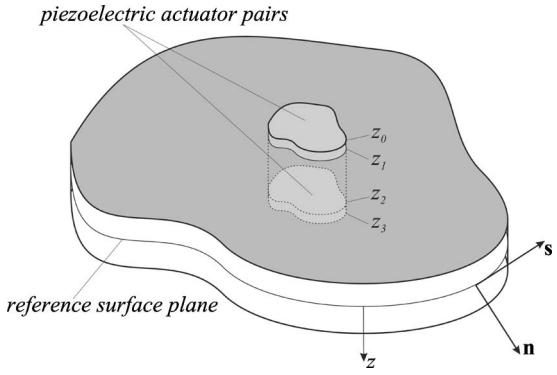


Fig. 1 Substrate plate with a single pair of piezoelectric actuators

2 Piezoelectric Actuators

In the first part of this paper (Krommer and Varadan [9]), we presented an exact solution for the control of bending vibrations within subdomains of thin plates. We assumed any suitable self-stress actuation to be the control agency. In terms of constitutive relations, the actuation was characterized as

$$\mathbf{M} = \mathbf{D}:\boldsymbol{\kappa} - \mathbf{M}^*, \tag{1}$$

wherein $\mathbf{D}$ is the bending stiffness tensor, $\boldsymbol{\kappa}$ is the curvature tensor, and $\mathbf{M}^*(\mathbf{x},t)$ is the self-moment tensor, which was assumed arbitrary and applied within the subdomain to be controlled. $\mathbf{M}$ is the bending moment tensor. In the second part of this paper, we focus on the application of piezoelectric actuators to put the self-stress actuation into practice. Methods that achieve an arbitrary space-wise distribution of $\mathbf{M}^*(\mathbf{x},t)$ have been reported in the literature, see Lee [12] for the method of varying the polarization profile of piezoelectric layers. Their practicality is questionable, because once the polarization profile is fixed, it can no longer be changed in order to adapt the control to changing disturbances. For that reason, we restrict the content of this paper to the use of piezoelectric patch actuators with a constant polarization profile. In such a case, the distributed actuation has to be achieved by the proper placement of individual patches; preferably resulting in a suboptimal solution when compared to the exact solution. To incorporate piezoelectric patches into thin-plate theory we study a single patch attached to a substrate plate at an arbitrary location, see Fig. 1. However, we only use pairs of patches, one of them attached to the upper side of the substrate and the other one attached to the lower side. This ensures pure bending of the plate. The piezoelectric self-moment tensor is given as

$$\mathbf{M}^*(\mathbf{x},t) = \int_{z_0}^{z_1} \bar{\mathbf{e}} E_z^u(\mathbf{x},t) z dz + \int_{z_2}^{z_3} \bar{\mathbf{e}} E_z^l(\mathbf{x},t) z dz, \tag{2}$$

in which z_0 and z_1 characterize the thickness extension of the upper patch, z_2 and z_3 are the thickness extension of the lower one, and z_1 and z_2 are the thickness extension of the substrate, see again Fig. 1, $\bar{\mathbf{e}}$ is the tensor of piezoelectric coefficients, which we assume to be identical for the two patches, and $E_z^{u,l}(\mathbf{x},t)$ is the thickness component of the electric-field vector in the corresponding patch.

Within the piezoelectric patches, the mechanical and the electrical field are coupled to each other. To account for this coupling, we have to consider the charge equation of electrostatics for dielectrics, $\operatorname{div}\mathbf{D}^{u,l}=0$, see e.g., Eringen and Maugin [15]. In thin-plate theory, we assume the piezoelectric patches to be relatively thin, such that an approximation of the three-dimensional formulation by means of $\partial D_z^{u,l}/\partial z=0$ is suitable. This approximation is based on the assumption that in-plane components of the electric displacement vector $\mathbf{D}$ are negligible. The constitutive relation for the nonzero thickness component of the electric displacement vector $D_z^{u,l}$ is

$$D_z^{u,l}=\bar{\mathbf{e}}:z\boldsymbol{\kappa}+\eta E_z^{u,l}, \tag{3}$$

where η is the electric permittivity. To use the patches as actuators, the surfaces are electroded; so that a voltage $V^{u,l}$ can be applied. It is now easy to find a solution for $E_z^{u,l}(\mathbf{x},t)$ from $\partial D_z^{u,l}/\partial z=0$ as follows:

$$\frac{\partial D_z^{u,l}}{\partial z}=0\Rightarrow D_z^{u,l}=\frac{1}{h_p}\int_{z_0,z_2}^{z_1,z_3}(\bar{\mathbf{e}}:z\boldsymbol{\kappa}+\eta E_z^{u,l})dz$$

$$=\frac{\eta V^{u,l}}{h_p}+\frac{1}{h_p}\int_{z_0,z_2}^{z_1,z_3}\bar{\mathbf{e}}:z\boldsymbol{\kappa}dz,$$

$$E_z^{u,l}=\frac{V^{u,l}}{h_p}-\left[\frac{\bar{\mathbf{e}}:\boldsymbol{\kappa}}{\eta}z-\frac{1}{h_p}\int_{z_0,z_2}^{z_1,z_3}\frac{\bar{\mathbf{e}}:\boldsymbol{\kappa}}{\eta}zdz\right]. \tag{4}$$

Equation (4) can be inserted into the definition of the piezoelectric self-moment tensor and the resulting relation can be inserted into the constitutive relation. We start with the constitutive relation, which for the domain with piezoelectric patches is

$$\mathbf{M}=\left(\int_{z_1}^{z_2}\mathbf{C}^{\text{substrate}}z^2+\int_{z_0}^{z_1}\mathbf{C}^{\text{patch}}z^2+\int_{z_2}^{z_3}\mathbf{C}^{\text{patch}}z^2\right):\boldsymbol{\kappa}$$

$$-\int_{z_0}^{z_1}\bar{\mathbf{e}}E_z^u zdz-\int_{z_2}^{z_3}\bar{\mathbf{e}}E_z^l zdz. \tag{5}$$

where $\mathbf{C}^{\text{substrate}}$ and $\mathbf{C}^{\text{patch}}$ are the elasticity tensor of the substrate and of the piezoelectric patches. Inserting Eq. (4) into Eq. (5), we find

$$\mathbf{M}=\mathbf{D}^{\text{eff}}:\boldsymbol{\kappa}-\mathbf{M}^{*,\text{eff}}, \tag{6}$$

in which the bending stiffness tensor $\mathbf{D}^{\text{eff}}=\mathbf{D}_0+\Delta\mathbf{D}$ accounts for the stiffness due to the elasticity tensor $\mathbf{C}^{\text{substrate}}$ of the substrate layer, for the stiffness due to the elasticity tensor $\mathbf{C}^{\text{patch}}$ of the patches and for an additional stiffness due to the influence of the direct piezoelectric effect. The effective self-moment tensor $\mathbf{M}^{*,\text{eff}}$ characterizes the piezoelectric actuation. The definitions of these entities are:

$$\mathbf{M}^{*,\text{eff}}=\int_{z_0}^{z_1}\bar{\mathbf{e}}\frac{V^u}{h_p}zdz+\int_{z_2}^{z_3}\bar{\mathbf{e}}\frac{V^l}{h_p}zdz=\int_{z_0}^{z_1}\bar{\mathbf{e}}\frac{2V^u}{h_p}zdz,$$

$$\mathbf{D}_0=\int_{z_1}^{z_2}\mathbf{C}^{\text{substrate}}z^2dz,$$

$$\Delta\mathbf{D}=\int_{z_0}^{z_1}\left[\mathbf{C}^{\text{patch}}z+\frac{\bar{\mathbf{e}}\otimes\bar{\mathbf{e}}}{\eta}\left(z-\frac{1}{2}(z_0+z_1)\right)\right]zdz$$

$$+\int_{z_2}^{z_3}\left[\mathbf{C}^{\text{patch}}z+\frac{\bar{\mathbf{e}}\otimes\bar{\mathbf{e}}}{\eta}\left(z-\frac{1}{2}(z_2+z_3)\right)\right]zdz. \tag{7}$$

Note that $\mathbf{D}_0$ is not identical to $\mathbf{D}$, because it only accounts for the influence of the substrate layer. Furthermore, we have assumed $z_0=-z_3$, $z_1=-z_2$, and $V^l=-V^u$, to ensure pure bending. With respect to direct problems, this theory can be applied for the analysis of the electromechanically coupled behavior of thin plates. Results have been reported in a paper by Krommer [16], in which the theory is developed in more detail. The dynamic behavior of the plate with piezoelectric patches at arbitrary locations is governed by

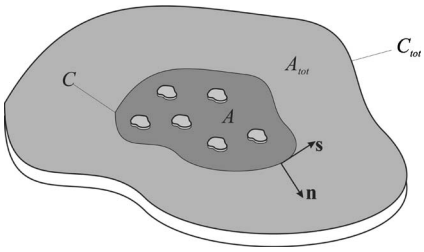


Fig. 2 Plate with subdomain to be controlled

$$A:\ \operatorname{div}(\operatorname{div}\mathbf{M})+p_z=\rho^{(0)}\ddot{\bar{w}}_0,$$

$$C:\ [\operatorname{div}\mathbf{M}\cdot\mathbf{n}+\nabla(\mathbf{Mn}\cdot\mathbf{s})\cdot\mathbf{s}-\bar{q}]\delta w_0=0,$$

$$(\mathbf{Mn}\cdot\mathbf{n}-\bar{m})(\nabla\delta w_0\cdot\mathbf{n})=0,$$

$$P:\ \ (\mathbf{Mn}\cdot\mathbf{s})\delta w_0]_{P^+}^{P^-}=0, \tag{8}$$

with the constitutive relation $\mathbf{M}=\mathbf{D}^{\text{eff}}:\boldsymbol{\kappa}-\mathbf{M}^{*,\text{eff}}$ in areas with piezoelectric patch actuator pairs and $\mathbf{M}=\mathbf{D}_0:\boldsymbol{\kappa}$ in areas without patches. The linear inertia $\rho^{(0)}$ is defined as $\rho^{(0)}=\rho_0^{(0)}+\Delta\rho^{(0)}$, where $\rho_0^{(0)}$ accounts for the substrate and $\Delta\rho^{(0)}$ for the patches; hence, the definitions are

$$\rho_0^{(0)}=\int_{z_1}^{z_2}\rho^{\text{substrate}}dz,\quad \Delta\rho^{(0)}=2\int_{z_0}^{z_1}\rho^{\text{patch}}dz, \tag{9}$$

with ρ being the mass density of the materials.

3 Summary of Methodology

We refer the reader who is interested in the theoretical background of the methodology to Part I of the paper (Krommer and Varadan [9]). In Part II, we only summarize the methodology. We study a thin plate within the domain A_{tot} and with a boundary that is defined by C_{tot}; the kinematical boundary conditions are homogenous. Our aim is to control the bending motion within a subdomain A of A_{tot}; moreover, we apply control only in the subdomain A. The geometry of the plate is shown in Fig. 2, with an arbitrary number of piezoelectric patches attached. We release the subdomain A from the rest of the plate and we account for the effect of the rest of the plate, that is $A_{\text{tot}}-A$, by applying yet unknown boundary conditions at the interfacing curve C. Normally, one would consider dynamical boundary conditions; however, one may as well replace some of the dynamical boundary conditions by kinematical ones. In each point of C, we prescribe dynamical, mixed, or kinematical boundary conditions. The four possible types of boundary conditions are:

$$C:\begin{cases}\operatorname{div}\mathbf{M}\cdot\mathbf{n}+\nabla(\mathbf{Mn}\cdot\mathbf{s})\cdot\mathbf{s}=\bar{q},\ \mathbf{Mn}\cdot\mathbf{n}=\bar{m}\\ w_0=\bar{w}_0,\ \mathbf{Mn}\cdot\mathbf{n}=\bar{m}\\ \operatorname{div}\mathbf{M}\cdot\mathbf{n}+\nabla(\mathbf{Mn}\cdot\mathbf{s})\cdot\mathbf{s}=\bar{q},\ \nabla w_0\cdot\mathbf{n}=\bar{\psi}\\ w_0=\bar{w}_0,\ \nabla w_0\cdot\mathbf{n}=\bar{\psi}\end{cases}, \tag{10}$$

where $\bar{m}$ and $\bar{q}$ stand for moment and transverse force and $\bar{w}_0$ and $\bar{\psi}$ are deflection and normal slope at the boundary C of the released domain. In contrast to Part I, we will also use purely dynamical boundary conditions, if possible. We introduce an additive separation of the deflection $w_0=\tilde{w}_0+\check{w}_0$ within A, in which

the part $\breve{w}_0$ accounts for the nonhomogenous kinematical boundary conditions according to Eq. (10); within the domain A, the deflection of the controlled plate will exactly coincide with this latter deflection $\breve{w}_0$, if we apply a proper control by means of self-stress actuation. In Part I, we showed that the deflection $\tilde{w}_0$ can be represented by a convolution integral of the form:

$$\int_0^t \left[\int_A p_z^d(\mathbf{x},t-\tau)\tilde{w}_0(\mathbf{x},t)dA + \int_{C_q} \bar{q}^d(\mathbf{x},t-\tau)\tilde{w}_0(\mathbf{x},t)dC - \int_{C_m} \bar{m}^d(\mathbf{x},t-\tau)(\nabla\tilde{w}_0(\mathbf{x},t)\cdot\mathbf{n})dC \right] dt = \int_0^t \int_A [\mathbf{M}^p + \mathbf{M}^{*,\mathrm{eff}}](\mathbf{x},t-\tau):\boldsymbol{\kappa}^d(\mathbf{x},t)dAdt. \quad (11)$$

In contrast to Part I, we take nonhomogenous dynamical boundary conditions into account in the dummy loading case. C_q is the part of C on which transverse forces are applied and C_m is the part of C on which moments are applied. We conclude that if $\mathbf{M}^p + \mathbf{M}^{*,\mathrm{eff}} = \mathbf{0}$, then the deflection $\tilde{w}_0(\mathbf{x},t)$ vanishes at every point within the domain A for every instant of time. This follows directly from the arbitrariness of the dummy loading. Therefore, the deflection of the plate is $\breve{w}_0$. This solution represents an exact solution of an inverse problem; given the external loading of the plate, we seek for a distributed actuation, which enforces a subdomain of the plate to perform a desired bending motion. $\mathbf{M}^p$ is a statically admissible moment tensor that can be calculated from

$$A:\ \mathrm{div}(\mathrm{div}\,\mathbf{M}^p) + \tilde{p}_z = 0,$$

$$C:\begin{cases} \mathrm{div}\,\mathbf{M}^p\cdot\mathbf{n} + \nabla(\mathbf{M}^p\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s} = \tilde{q},\ \mathbf{M}^p\mathbf{n}\cdot\mathbf{n} = \tilde{m} \\ \mathbf{M}^p\mathbf{n}\cdot\mathbf{n} = \tilde{m} \\ \mathrm{div}\,\mathbf{M}^p\cdot\mathbf{n} + \nabla(\mathbf{M}^p\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s} = \tilde{q} \\ \text{no boundary conditions to be satisfied} \end{cases}, \quad (12)$$

where a transformed force $\tilde{p}_z$, a transformed moment $\tilde{m}$, and a transformed transverse force $\tilde{q}$ have been introduced as

$$A:\ \tilde{p}_z = p_z - \bar{\rho}^{(0)}\ddot{\breve{w}}_0 + \mathrm{div}(\mathrm{div}\,\breve{\mathbf{M}}),\quad \breve{\mathbf{M}} = \bar{\mathbf{D}}:\breve{\boldsymbol{\kappa}},$$

$$\breve{\boldsymbol{\kappa}} = -\frac{1}{2}[\nabla\nabla\breve{w}_0 + (\nabla\nabla\breve{w}_0)^T],$$

$$C:\ \tilde{q} = \bar{q} - \mathrm{div}\,\breve{\mathbf{M}}\cdot\mathbf{n} - \nabla(\breve{\mathbf{M}}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s},\ \tilde{m} = \bar{m} - \breve{\mathbf{M}}\mathbf{n}\cdot\mathbf{n}. \quad (13)$$

We note that both linear inertia and bending stiffness may vary in the domain A, because in some areas piezoelectric patches are present. We use the parameters $\bar{\rho}^{(0)}$ and $\bar{\mathbf{D}}$, with $\bar{\rho}^{(0)} = \rho_0^{(0)}$ and $\bar{\mathbf{D}} = \mathbf{D}_0$ in case of no patches attached or $\bar{\rho}^{(0)} = \rho^{(0)}$ and $\bar{\mathbf{D}} = \mathbf{D}^{\mathrm{eff}}$ in case of attached patches. From the above equations, we can calculate an exact solution for the plate with nonconstant parameters, which in general will result into a distributed self-moment tensor $\mathbf{M}^{*,\mathrm{eff}}$, which forces the subdomain to have the deflection $\breve{w}_0$. However, this self-moment tensor cannot be realized in practice, because the actuation is restricted to the piezoelectric patches.

3.1 Solution Procedure. In contrast to Part I, we allow curvature free deflections $\breve{w}_0$ within the subdomain A, bounded by C (see Fig. 2); hence, we are able to avoid the introduction of a transition domain, which we used in Part I to bring the deflection back to a zero value. The controlled deflection within A has to satisfy

$$A:\ \nabla\nabla\breve{w}_0 = \mathbf{0},$$

$$C:\begin{cases} \text{no conditions to be satisfied} \\ \breve{w}_0 = \bar{w}_0 \\ \nabla\breve{w}_0\cdot\mathbf{n} = \bar{\psi} \\ \breve{w}_0 = \bar{w}_0,\ \nabla\breve{w}_0\cdot\mathbf{n} = \bar{\psi} \end{cases}. \quad (14)$$

Any deflection satisfying the relations of Eq. (14) is a proper choice for $\breve{w}_0$, resulting in $\nabla\nabla\breve{w}_0 = \mathbf{0}$ inside A. We note the advantage of this procedure is that we do not have to account for a transition domain with a highly complicated distribution of the actuation. We proceed with solving Eq. (12) by using boundary conditions that ensure a solution. Two important aspects in the solution procedure are as follows: first, we have to make sure that the actuation $\mathbf{M}^{*,\mathrm{eff}}$ we calculate satisfies all constraints implied by the piezoelectric material used, e.g., if the piezoelectric material is transversely isotropic with the isotropic plane coinciding with the reference surface of the plate, then the result of our procedure must be a spherical actuation tensor, $\mathbf{M}^{*,\mathrm{eff}} = M^{*,\mathrm{eff}}\mathbf{I}$; second, the choice of boundary conditions for the released domain of the plate has to be proper in the sense that the dummy loading case with arbitrary loading does not result in a rigid body motion of the plate subdomain. For this reason, we are unable to consider dynamical boundary conditions on the whole curve C. In the examples, we will show that indeed kinematical boundary conditions are not avoidable in all situations.

3.2 Approximate Solution. The solutions are approximate because discrete piezoelectric patches cannot ensure the vanishing of the right-hand side of

$$\int_0^t \left[\int_A p_z^d(\mathbf{x},t-\tau)\tilde{w}_0(\mathbf{x},t)dA + \int_{C_q} \bar{q}^d(\mathbf{x},t-\tau)\tilde{w}_0(\mathbf{x},t)dC - \int_{C_m} \bar{m}^d(\mathbf{x},t-\tau)(\nabla\tilde{w}_0(\mathbf{x},t)\cdot\mathbf{n})dC \right] dt = \int_0^t \int_A [\mathbf{M}^p + \mathbf{M}^{*,\mathrm{eff}}](\mathbf{x},t-\tau):\boldsymbol{\kappa}^d(\mathbf{x},t)dAdt. \quad (15)$$

If we have $i = 1 \ldots n$ patches within A, each patch is assumed to be located within a domain A_i^{sub}, which is a subset of A_i such that $A_i^{\mathrm{sub}} \subseteq A_i$, where $\cap_{i=1}^n A_i = A$. The piezoelectric actuation is $\mathbf{M}_i^{*,\mathrm{eff}}(t)$. The right-hand side of Eq. (15) then becomes

$$\int_0^t \int_A [\mathbf{M}^p + \mathbf{M}^{*,\mathrm{eff}}](\mathbf{x},t-\tau):\boldsymbol{\kappa}^d(\mathbf{x},t)dAdt = \int_0^t \left[\sum_{i=1}^n \left(\mathbf{M}_i^{*,\mathrm{eff}}(t-\tau):\int_{A_i^{\mathrm{sub}}} \boldsymbol{\kappa}^d(\mathbf{x},t)dA \right) + \int_A \mathbf{M}^p(\mathbf{x},t-\tau):\boldsymbol{\kappa}^d(\mathbf{x},t)dA \right] dt. \quad (16)$$

3.2.1 Direct Method. A direct method to calculate the functions $\mathbf{M}_i^{*,\mathrm{eff}}(t)$ is to neglect the influence of the curvature tensor in Eq. (16) and simply ask for the satisfaction of

$$\mathbf{M}_i^{*,\mathrm{eff}}(t) = -\frac{1}{A_i^{\mathrm{sub}}}\int_{A_i} \mathbf{M}^p(\mathbf{x},t)dA. \quad (17)$$

We denote this method as the direct method, because the approximate values for the functions $\mathbf{M}_i^{*,\mathrm{eff}}(t)$ are directly calculated from the statically admissible moment tensor. The higher the number of actuators, the closer the approximate solution will be to the exact solution. The advantage of the direct method is that we can use as many actuators as we want to calculate the functions $\mathbf{M}_i^{*,\mathrm{eff}}(t)$, because the exact solution is already known. On the other hand,

the direct method does not permit a mechanical interpretation of the left-hand side of Eq. (15).

3.2.2 Indirect Method. If we use a special dummy loading case, then we can calculate the right hand side of Eq. (16) and force it to vanish. For each choice of the dummy load, we obtain one equation for the unknown functions $\mathbf{M}_i^{*,\mathrm{eff}}(t)$,

$$\int_0^t \sum_{i=1}^{n} \left(\mathbf{M}_i^{*,\mathrm{eff}}(t-\tau) : \int_{A_i^{\mathrm{sub}}} \boldsymbol{\kappa}^{d,j}(\mathbf{x},t)\,dA \right) dt = -\int_0^t \int_A \mathbf{M}^p(\mathbf{x},t-\tau) : \boldsymbol{\kappa}^{d,j}(\mathbf{x},t)\,dA\,dt, \tag{18}$$

in which $\mathbf{M}^p$ and $\boldsymbol{\kappa}^{d,j}$ are known. In the indirect method [the functions $\mathbf{M}_i^{*,\mathrm{eff}}(t)$ are not directly calculated from the statically admissible moment tensor, but indirectly from Eq. (18)] the number of different dummy loading cases $j=1\ldots n$ has to be identical to the number of actuators, which makes it hard to decide what dummy load to use in case of a very high number of piezoelectric patches. The main advantage of the indirect method is the mechanical interpretation of the left-hand side of Eq. (15). Say, $p_z^{d,j}=0$, $\bar{m}^{d,j}=0$, and $\bar{q}^{d,j}(\mathbf{x},t)=f(t)$ or $p_z^{d,j}=0$, $\bar{q}^{d,j}=0$ and $\bar{m}^{d,j}(\mathbf{x},t)=f(t)$ or $\bar{m}^{d,j}=0$, $\bar{q}^{d,j}=0$ and $p_z^{d,j}(\mathbf{x},t)=f(t)$, then we obtain the following mechanical interpretation of the left-hand side of Eq. (15)

$$\int_0^t f(t-\tau) \int_{C_q} \tilde{w}_0(\mathbf{x},t)\,dC\,dt,$$

$$-\int_0^t f(t-\tau) \int_{C_m} \nabla \tilde{w}_0(\mathbf{x},t)\cdot\mathbf{n}\,dC\,dt, \tag{19}$$

$$\int_0^t f(t-\tau) \int_A \tilde{w}_0(\mathbf{x},t)\,dA\,dt.$$

Obviously these relations are related with mean values of the deviation of the deflection from the desired one, either with respect to the boundary or domain. According to Eq. (18), all these quantities can be made zero. Higher-order dummy force loadings $p_z^{d,j}(\mathbf{x},t)$ result in the elimination of higher moments of the deviation.

3.2.3 Combined Method. The two methods we have just discussed can be easily combined, combing the advantages of both. First, we wish to have a high number of piezoelectric patches to closely match the distributed actuation. Second, we want to achieve a mechanical interpretation of the left hand side of Eq. (15) as well. We start by noting that the statically admissible moment tensor $\mathbf{M}^p(\mathbf{x},t)$ and the self-moment tensor $\mathbf{M}^{*,\mathrm{eff}}(t)$ can be separated as

$$\mathbf{M}^p(\mathbf{x},t)=\sum_{j=1}^{m}\mathbf{M}_j^p(\mathbf{x})f_j(t), \quad \mathbf{M}^{*,\mathrm{eff}}(\mathbf{x},t)=\sum_{j=1}^{m}\mathbf{M}_j^{*,\mathrm{eff}}(\mathbf{x})g_j(t), \tag{20}$$

in which the number of terms m depends on the transformed force $\tilde{p}_z$, the transformed moment $\tilde{m}$, and the transformed transverse force $\tilde{q}$; especially with respect to the number of different time variations involved. $\mathbf{M}_j^{*,\mathrm{eff}}(\mathbf{x})$ is domain-wise constant and $g_j(t)$ is its time variation. For each j, we use the direct method to calculate

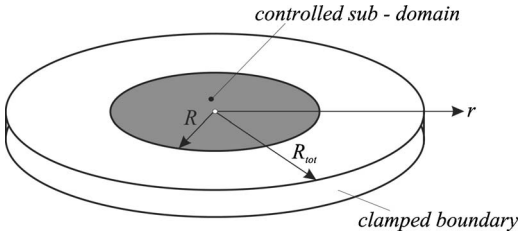


Fig. 3 Circular plate with subdomain to be controlled

$$\mathbf{M}_{ij}^{*,\mathrm{eff}}=-\frac{1}{A_i^{\mathrm{sub}}}\int_{A_i}\mathbf{M}_j^p(\mathbf{x})\,dA, \tag{21}$$

where again $i=1\ldots n$ is the number of patches. The time variations $g_j(t)$ are calculated by using the indirect method with proper dummy loadings. This results in m equations of the form

$$\int_0^t \sum_{i=1}^{n} \left(\sum_{j=1}^{m} g_j(t-\tau)\mathbf{M}_{ij}^{*,\mathrm{eff}} : \int_{A_i^{\mathrm{sub}}} \boldsymbol{\kappa}^{d,k}(\mathbf{x},t)\,dA \right) dt = -\int_0^t \int_A \mathbf{M}^p(\mathbf{x},t-\tau) : \boldsymbol{\kappa}^{d,k}(\mathbf{x},t)\,dA\,dt, \tag{22}$$

if we use $k=1\ldots m$ dummy loadings. It remains to choose the dummy forces $p_z^{d,k}(\mathbf{x},t)$ and the dummy boundary conditions $\bar{m}^{d,k}(\mathbf{x},t)$ and $\bar{q}^{d,j}(\mathbf{x},t)$. We automatically obtain m such dummy loadings by using the transformed force $\tilde{p}_z$, the transformed moment $\tilde{m}$, and the transformed transverse force $\tilde{q}$ for that sake.

After having discussed the solution procedure and methods to find approximate solutions for subdomain control of plates by means of piezoelectric patch actuators, we will present numerical results for the axisymmetric bending of circular plates in the following section.

4 Axisymmetric Bending of Circular Plates

In the previous sections, we have summarized the application of piezoelectric patch actuators for the purpose of subdomain control of a thin plate, the theoretical background of our methodology and we have developed methods to obtain approximate solutions. In this section, we discuss the axisymmetric bending of circular plates. Figure 3 shows the circular plate with the domain we would like to control. We consider a clamped circular plate made of aluminum that has a radius $R_{\mathrm{tot}}=0.2$ m and a thickness $h=3$ mm. The three lowest (axisymmetric bending) natural frequencies of this plate without piezoelectric patches are $f_1=182.0$ Hz, $f_2=708.7$ Hz, and $f_3=1588$ Hz. The domain we want to control is within $r\leqslant R=3$ cm. This example problem is one of the problems that were used in Part I of the paper. Therefore, we refer to the first part of the paper for details on the governing equations and on the formulation of our methodology for axisymmetric bending as well as for the material parameters. In this part, we will allow the domain to perform a space-wise constant motion; hence within $r\leqslant R$, we wish the deflection to be $\breve{w}_0(r,t)=\bar{w}_0 f(t)$, where $\bar{w}_0$ is an arbitrary constant. Within $r\leqslant R$, we calculate

$$r\leqslant R:\ \mathrm{div}(\mathrm{div}\,\mathbf{M}^p)+p_z-\bar{\rho}^{(0)}\bar{w}_0\overset{\circ\circ}{f}(t)=0, \quad r=R:\ \mathbf{M}^p\mathbf{n}\cdot\mathbf{n}=\bar{m}, \tag{23}$$

where Eqs. (12) and (13) have been used. Note that, in this example, we need to apply control only in the subdomain itself. The transverse force loading $p_z(r,t)$ may be considered a general loading in space and time. However, we do not claim that this approach is exhaustive. Therefore, we consider the time variation of the loading to be harmonic; for an arbitrary time variation, we

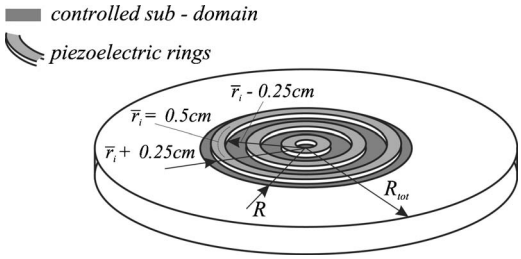


Fig. 4 Circular plate with piezoelectric ring actuators

may find the solution by means of the Fourier integral, see Graff [17]. The space-wise distribution is constant; hence; we have $p_z(r,t)=p_0e^{i\omega t}$.

For the numerical study, we choose $p_0=100N/m^2$ and we consider three forcing frequencies: $\omega=2\pi f=2\pi(100,500,1000)\ s^{-1}$; these are the same as in Part I. Also note that $\breve{w}_0(r,t)=\bar{w}_0e^{i\omega t}$, $\bar{m}=\bar{m}e^{i\omega t}$, and $\mathbf{M}^p=\mathbf{I}M^pe^{i\omega t}$. Equation (23) becomes

$$r\leqslant R:\ \Delta M^p+p_0+\omega^2\bar{\rho}^{(0)}\bar{w}_0=0,\ r=R:\ M^p=\bar{m}. \tag{24}$$

Equation (24) can be easily solved for the axisymmetric case. In contrast to Part I, Eq. (24) accounts for different linear inertia, depending on whether piezoelectric patches are present or not. Therefore, the exact solution is not independent from the choice for the location of piezoelectric patches.

4.1 Comparison Between Direct Method and Indirect Method. Exploiting the circular symmetry of the problem, we attach three concentric piezoelectric PZT-5A ring-shaped patch actuators symmetrically to the upper and lower surfaces of the aluminum plate. The thickness of the patches is $h_p=0.5$ mm and the patches are located within $r\in[\bar{r}_i\pm0.25\ \text{cm}]$, $\bar{r}_i=(0.5,1.5,2.5)$ cm. Figure 4 shows the geometry, detailing the dimensions of the patch with $\bar{r}_i=0.5$ cm. According to Eq. (17), we have to calculate

$$M_i^{*,\text{eff}}\int_{\bar{r}_i-0.0025}^{\bar{r}_i+0.0025}2\pi rdr=-\int_{\bar{r}_i-0.005}^{\bar{r}_i+0.005}M^p(r)2\pi rdr,\ i=1,2,3, \tag{25}$$

if we use the direct method. The three constant values for the self-moment then have to be applied at the three ring patches. Using the indirect method, we find these three values by solving Eq. (18):

$$\int_0^t\sum_{i=1}^{3}\left(M_i^{*,\text{eff}}\int_{\bar{r}_i-0.0025}^{\bar{r}_i+0.0025}\Delta w_0^{d,j}(r)2\pi rdr\right)e^{i\omega(t-\tau)}e^{i\omega t}dt=-\int_0^t\left(\int_0^R M^p(r)\Delta w_0^{d,j}(r)2\pi rdr\right)e^{i\omega(t-\tau)}e^{i\omega t}dt,$$

$$j=1,2,3. \tag{26}$$

Obviously, we need to use three dummy loadings to obtain three equations for the three unknown $M_i^{*,\text{eff}}$ to be applied at the ring patches. We choose $p_z^{d,1}(r,t)=1e^{i\omega t}$ with $\bar{q}^{d,1}=0$ and $\bar{m}^{d,1}=0$, $p_z^{d,2}(r,t)=(\bar{\rho}^{(0)}/\rho_0^{(0)})e^{i\omega t}$ with $\bar{q}^{d,2}=0$ and $\bar{m}^{d,2}=0$, and $p_z^{d,3}=0$ with $\bar{q}^{d,3}=0$ and $\bar{m}^{d,3}=1e^{i\omega t}$. The first type of loading corresponds to the original loading of the plate, the second one to the loading due to the space-wise constant deflection of the subdomain, and the third arises via the dynamical boundary conditions for the subdomain. The left-hand side of Eq. (15) for these three loadings is

$$\int_0^t\left(\int_0^R\tilde{w}_0(r)2\pi rdr\right)e^{i\omega(t-\tau)}e^{i\omega t}dt,\ \int_0^t\left(\int_0^R\frac{\bar{\rho}^{(0)}}{\rho_0^{(0)}}\tilde{w}_0(r)2\pi rdr\right)\times e^{i\omega(t-\tau)}e^{i\omega t}dt,-\int_0^t\left[\frac{\partial\tilde{w}_0(r)}{\partial r}\right]_{r=R}e^{i\omega(t-\tau)}e^{i\omega t}dt. \tag{27}$$

We finally combine Eqs. (26) and (27) by means of Eq. (15), a procedure which directly leads to the conclusion that the quantities

$$\int_0^R\tilde{w}_0(r)2\pi rdr,\ \int_0^R\frac{\bar{\rho}^{(0)}}{\rho_0^{(0)}}\tilde{w}_0(r)2\pi rdr,\left[\frac{\partial\tilde{w}_0(r)}{\partial r}\right]_{r=R} \tag{28}$$

vanish, if

$$\sum_{i=1}^{3}\left(M_i^{*,\text{eff}}\int_{\bar{r}_i-0.0025}^{\bar{r}_i+0.0025}\Delta w_0^{d,j}(r)2\pi rdr\right)+\int_0^R M^p(r)\Delta w_0^{d,j}(r)2\pi rdr=0,\quad j=1,2,3. \tag{29}$$

The three constant values for the self-moment, which are calculated from the three equations in Eq. (29) then have to be applied at the three ring patches. The result will be that the mean value of $\tilde{w}_0$, a weighted mean value (incorporating different linear inertia) of $\tilde{w}_0$, and the slope of $\tilde{w}_0$ at $r=R$ vanish exactly.

Figure 5 shows the results for forcing frequencies $\omega=2\pi(100,500,1000)\ s^{-1}$. The deflection is presented for the uncontrolled plate, the controlled plate with the exact solution for the piezoelectric moment applied, as well as the deflection we obtained by using either the direct method or the indirect method. The exact solution is the optimal solution; the subdomain has a space-wise constant deflection. The suboptimality of the approximate solutions has to be measured with respect to the exact solution. If we use the direct method, the deflection is close to the exact solution everywhere throughout the plate. However, we have no criteria to measure the quality itself. Subjectively, one may say the quality is poor. If we use the indirect method, we are able to measure the quality of the solution. First, the deflection in the uncontrolled domain is identical to the one of the exact solution; second, at $r=R$, the deflection and the slope of the approximate solution coincide exactly with the exact solution; third, the mean value of the deviation vanishes. However, the satisfaction of the third criteria does not result in a deflection close to the exact one. The reason is that deviations from the exact solution at points with small r are not properly accounted for because the area element is $2\pi rdr$ and it vanishes for $r=0$. To find a better solution, we proceed by using the combined method.

4.2 Combined Method. To use the combined method we attach six concentric piezoelectric PZT-5A ring-shaped patch actuators to the upper surface and the lower surface of the aluminum plate. These patches are located within $r\in[\bar{r}_i\pm0.125\ \text{cm}]$, $\bar{r}_i=(0.25,0.75,1.25,1.75,2.25,2.75)$ cm. In a first step, we use Eq. (21) to calculate

$$M_{ij}^{*,\text{eff}}\int_{\bar{r}_i-0.00125}^{\bar{r}_i+0.00125}2\pi rdr=-\int_{\bar{r}_i-0.0025}^{\bar{r}_i+0.0025}M_j^P(r)2\pi rdr,$$

$$i=1,2,3,4,5,6,\ j=1,2,3, \tag{30}$$

where M_1^P is due to the constant force loading p_0, M_2^P is due to the piece-wise constant force loading $\bar{\rho}^{(0)}/\rho_0^{(0)}$, and M_3^P is due to the applied moment $\bar{m}$ at $r=R$. At each piezoelectric ring, we define the self-moment as

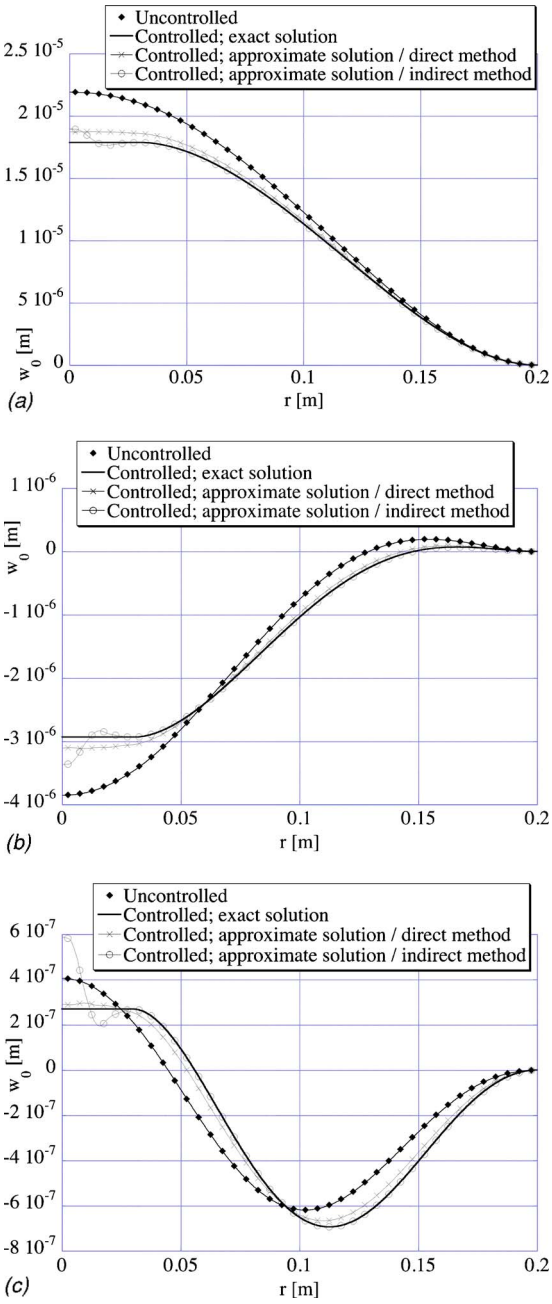


Fig. 5 Deflection of the circular plate: (a) $\omega=2\pi 100\ \mathrm{s}^{-1}$, (b) $\omega=2\pi 500\ \mathrm{s}^{-1}$, and (c) $\omega=2\pi 1000\ \mathrm{s}^{-1}$

$$M_i^{*,\mathrm{eff}}e^{i\omega t}=\sum_{j=1}^{3}M_{ij}^{*,\mathrm{eff}}K_j e^{i\omega t},\ i=1,2,3,4,5,6, \tag{31}$$

in which K_j, $j=1,2,3$ is an amplitude we still need to calculate. Accordingly, we use Eq. (22) in the form

$$\sum_{i=1}^{6}\left(\sum_{j=1}^{3}M_{ij}^{*,\mathrm{eff}}K_j\right)\int_{\bar{r}_i-0.00125}^{\bar{r}_i+0.00125}\Delta w_0^{d,k}(r)2\pi r dr = -\int_0^R M^p(r)\Delta w_0^{d,k}(r)2\pi r dr,\quad k=1,2,3, \tag{32}$$

To avoid the poor result we obtained with the indirect method in the previous section, we use a different dummy loading. For Case 1, $p_z^{d,1}(r,t)=1/(2\pi\bar{r})e^{i\omega t}$ with $\bar{q}^{d,1}=0$ and $\bar{m}^{d,1}=0$ instead of

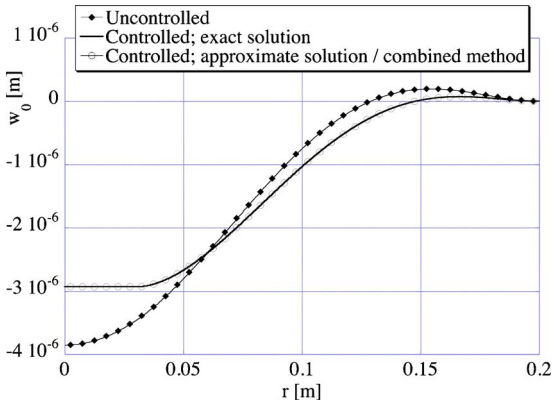


Fig. 6 Deflection of the circular plate; $\omega=2\pi 500\ \mathrm{s}^{-1}$

$p_z^{d,1}(r,t)=1e^{i\omega t}$ with $\bar{q}^{d,1}=0$ and $\bar{m}^{d,1}=0$; for Cases 2 and 3, we use the same loading as before. Note that by $2\pi\bar{r}$, we mean a piece-wise-constant function that approximates $2\pi r$ because we cannot use the latter due to its singular value at $r=0$. The three mechanical quantities we are able to eliminate by using the combined method are

$$\int_0^R \tilde{w}_0(r)\frac{2\pi r}{2\pi\bar{r}}dr,\ \int_0^R \frac{\bar{\rho}^{(0)}}{\rho_0^{(0)}}\tilde{w}_0(r)2\pi r dr,\ \left[\frac{\partial\tilde{w}_0(r)}{\partial r}\right]_{r=R}. \tag{33}$$

For the sake of brevity, we only show results for a forcing frequency $\omega=2\pi 500\ \mathrm{s}^{-1}$. Figure 6 shows the deflection of the circular plate. The deviation of the deflection due to the approximate solution from the deflection due to the exact solution is very small within the whole plate. We conclude that, indeed, the combined method results in a suboptimal solution, which eliminates the three quantities of Eq. (33) exactly, and which is hard to distinguish from the exact, hence optimal, solution. We note that results for $\omega=2\pi(100,1000)\ \mathrm{s}^{-1}$ are of the same quality. In Fig. 7, the required self-moment to achieve the results shown in Fig. 6 is plotted. The exact solution results in a distributed self-moment; contrarily, the approximate solution only allows us to apply a self-moment within the six domains where the piezoelectric rings are located. These latter ones are space-wise constant. From Fig. 7 we see that the approximate values are pretty close to a simple discretization of the distributed self-moment in the sense of the direct method. However, the combined method corrected the values one would obtain for the direct method such that proper mechanical quantities can be exactly eliminated. The resulting solution is much better than the one we would get using the direct method. In addition, the alternative choice of the loading Case 1 resulted in a solution, which prevents the undesired behavior we have experienced by only using the indirect method.

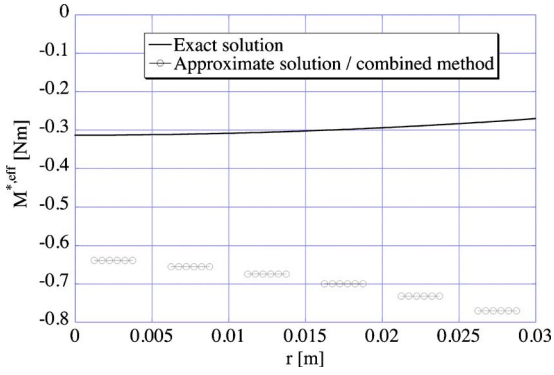


Fig. 7 Self-moment applied in subdomain; $\omega=2\pi 500\ \mathrm{s}^{-1}$

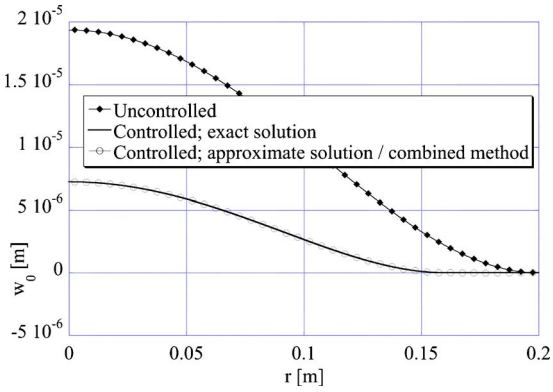


Fig. 8 Deflection of the circular plate; $\omega=2\pi 100\ s^{-1}$

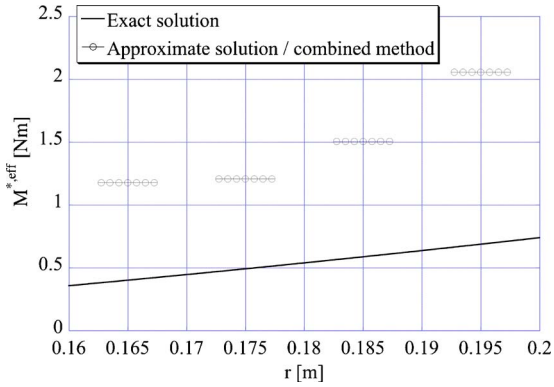


Fig. 9 Self-moment applied in subdomain; $\omega=2\pi 100\ s^{-1}$

4.3 Control of an Alternative Subdomain. In this section, we seek to control the subdomain within $0.16\ \text{m}=R\leqslant r\leqslant R_{\text{tot}}=0.2$ m. We use four concentric piezoelectric PZT-5A ring-shaped patches at the upper surface and the lower surface of the aluminum plate located at $r\in[\bar{r}_i\pm 0.25\ \text{cm}]$, $\bar{r}_i=(0.165,0.175,0.185,0.195)$ m. For the exact solution, we note that without having to introduce a transition domain, we are able to eliminate the deflection in the subdomain. For that sake, we use dynamical boundary conditions at $R=0.16$ m for calculating a statically admissible bending moment from

$$R\leqslant r\leqslant R_{\text{tot}}:\ \Delta M^p+p_0=0,\ r=R:\ M^p=\bar{m}\ \text{and}\ \frac{\partial M^p}{\partial r}=\bar{q}. \tag{34}$$

The exact solution $M^p+M^{*,\text{eff}}=0$ would result in a zero deflection within $R\leqslant r\leqslant R_{\text{tot}}$, because the kinematical boundary conditions at $r=R_{\text{tot}}$ are trivial. We find the approximate solution by using the combined method, in which the first step is to calculate

$$M_{ij}^{*,\text{eff}}\int_{\bar{r}_i-0.0025}^{\bar{r}_i+0.0025}2\pi r dr=-\int_{\bar{r}_i-0.005}^{\bar{r}_i+0.005}M_j^P(r)2\pi r dr,$$

$$i=1,2,3,4,\ j=1,2,3, \tag{35}$$

where M_1^P is due to the constant force loading p_0, M_2^P is due to the applied transverse force $\bar{q}$ at $r=R$, and M_3^P is due to the applied moment $\bar{m}$ at $r=R$. At each piezoelectric ring, we define the self-moment as

$$M_i^{*,\text{eff}}e^{i\omega t}=\sum_{j=1}^{3}M_{ij}^{*,\text{eff}}K_j e^{i\omega t},\ i=1,2,3,4, \tag{36}$$

with the amplitudes K_j, $j=1,2,3$, which are calculated from

$$\sum_{i=1}^{4}\left(\sum_{j=1}^{3}M_{ij}^{*,\text{eff}}K_j\right)\int_{\bar{r}_i-0.0025}^{\bar{r}_i+0.0025}\Delta w_0^{d,k}(r)2\pi r dr$$

$$=-\int_R^{R_{\text{tot}}}M^p(r)\Delta w_0^{d,k}(r)2\pi r dr,\quad k=1,2,3. \tag{37}$$

For the dummy loadings, we use $p_z^{d,1}(r,t)=1/(2\pi\bar{r})e^{i\omega t}$ with $\bar{q}^{d,1}=0$ and $\bar{m}^{d,1}=0$, with the piece-wise constant function $2\pi\bar{r}$, $p_z^{d,2}(r,t)=0$ with $\bar{q}^{d,2}=1e^{i\omega t}$ and $\bar{m}^{d,2}=0$, and $p_z^{d,3}(r,t)=0$ with $\bar{q}^{d,3}=0$ and $\bar{m}^{d,3}=1e^{i\omega t}$. Therefore, we are able to eliminate the mechanical quantities

$$\int_R^{R_{\text{tot}}}\tilde{w}_0(r)\frac{2\pi r}{2\pi\bar{r}}dr,\ [\tilde{w}_0(r)]_{r=R},\ \left[\frac{\partial\tilde{w}_0(r)}{\partial r}\right]_{r=R}. \tag{38}$$

Figure 8 shows the deflection of the circular plate for a forcing frequency $\omega=2\pi 100\ s^{-1}$. As in the previous example, the deviation between the exact solution and approximate solution is very small. Again, the approximate one can be called suboptimal with respect to the exact one and optimal with respect to the elimination of the mechanical quantities of Eq. (38). Results for different forcing frequencies are easily obtained, but not presented for the sake of brevity. The self-moment is shown in Fig. 9, and it can be seen that the space-wise values are related to the distributed ones.

Finally, we calculate the needed voltages to be applied at the piezoelectric patches for both example problems. The results for using the combined method are shown in Table 1. The calculation of the voltages is easy. We simply use the first relation of Eq. (7) with a given self-moment for each patch. The voltages are within a reasonable range, which is of high relevance for the practical applicability of our method.

5 Conclusion

The present part of this paper (Part II) on the control of subdomains of elastic plates focused on the practical aspect of using discrete piezoelectric patch actuators to achieve a desired goal in a suboptimal manner. In contrast to Part I, we did not try to eliminate the bending motion within the subdomain, but rather to make it a very simple one with respect to its space-wise distribution. The main features of this paper were:

- The use of discrete piezoelectric patch actuators incorporating electromechanical coupling, and
- The calculation of a suboptimal approximate solution for the actuation to be applied at the piezoelectric patches.

With respect to both features, we were successful. Electromechanical coupling was incorporated by means of effective piezoelastic stiffness parameters, and we were able to find highly accurate approximate solutions by combining two methods: one allowing the use of a high number of piezoelectric patches, and the other one the mechanical interpretation of the quality of the approximate solution.

We would also like to mention that the use of discrete piezoelectric patch actuators has two main practical advantages. One is that we are able to use the actuation for different goals as well, and the other one is that we do not need to apply the actuation

Table 1 Voltages to be applied at piezoelectric rings

ω[rad s^{-1}]	V_1 [V]	V_2 [V]	V_3 [V]	V_4 [V]	V_5 [V]	V_6 [V]
$2\pi 500$	−11.30	−11.58	−11.93	−12.38	−12.94	−13.63
ωrad [s^{-1}]	V_1 [V]	V_2 [V]		V_3 [V]		V_4 [V]
ωrad [s^{-1}]	20.794	21.333		26.594		36.336

within the whole subdomain we want to control, e.g., in all the examples, we only applied the actuation within half of the domain to be controlled. Concerning the problem of controlling the shape of conformal antennas, we note that our method is of high practical relevance, because the antenna does not undergo a deformation, if it is located within the subdomain. With respect to the first example problem, an antenna of any shape located within $r \leqslant R$ is kept in its nondeformed state. The space-wise constant deflection we allowed the subdomain to perform would only result in a constant phase shift of the antenna-radiation pattern and, hence, would not interfere with its performance. The constant phase shift can be easily corrected electronically, if desired. We also note that the presence of the antenna itself would not load the structure since a conformal antenna may only constitute a deposited thin metal electrode of the desired pattern.

Future research should be targeted toward the subjects of sensor design and automatic control, such that the structure becomes a smart structure. The smart structure should then be able to react to disturbances and to control the motion of a subdomain automatically.

Acknowledgment

Support for one of the authors (M.K.) from the MAX—KADE Foundation, the Austrian Academy of Sciences, and the Linz Center of Competence in Mechatronics, as well as support for another author (V.V.) from the National Science Foundation, is gratefully acknowledged.

References

[1] Haftka, R. T., and Adelman, H. M., 1985, "An Analytical Investigation of Static Shape Control of Large Space Structures by Applied Temperature," AIAA J., **23**, pp. 450–457.

[2] Irschik, H., and Pichler, U., 2001, "Dynamic Shape Control of Solids and Structures by Thermal Expansion Strains," J. Therm. Stresses, **24**, pp. 565–578.

[3] Irschik, H., 2002, "A Review on Static and Dynamic Shape Control of Structures by Piezoelectric Actuation," Eng. Struct., **24**, pp. 5–11.

[4] Reissner, H., 1931, "Selbstspennungen Elastischer Gebilde," Z. Angew. Math. Mech., **11**, pp. 59–70.

[5] Nemenyi, P., 1931, "Eigenspannungen und Eigenspannungsquellen," Z. Angew. Math. Mech., **11**, pp. 1–8.

[6] Mura, T., 1991, *Micromechanics of Defects in Solids*, 2nd ed., Kluwer, Dordrecht, The Netherlands.

[7] 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.

[8] Gopinathan, S. V., Varadan, V. V., and Varadan, V. K., 2001, "Active Noise Control Studies Using the Rayleigh–Ritz Method," *Proc. of IUTAM-Symposium on Smart Structures and Structronic Systems*, U. Gabbert and H. S. Tzou, eds., Kluwer, Dordrecht, The Netherlands, pp. 169–177.

[9] Krommer, M., and Varadan, V. V., 2005, "Control of Bending Vibrations Within Subdomains of Thin Plates. Part I: Theory and Exact Solution," ASME J. Appl. Mech., **72**(3), pp. 432–444.

[10] Tani, J., Takagi, T., and Qiu, J., 1998, "Intelligent Material Systems: Application of Functional Materials," Appl. Mech. Rev., **51**, pp. 505–521.

[11] Tzou, H. S., 1998, "Multifield Transducers, Devices, Mechatronic Systems, and Structronic Systems with Smart Materials," Shock Vib. Dig., **30**, pp. 282–294.

[12] Lee, C. K., 1992, "Theory of Laminated Piezoelectric Plates for the Design of Distributed Sensors/Actuators. Part I: Governing Equations and Reciprocal Relationship," J. Acoust. Soc. Am., **87**, pp. 1144–1155.

[13] Krommer, M., and Varadan, V. V., 2003, "Dynamic Shape Control of Conformal Antennas," Proc. SPIE, **5049**, pp. 622–630.

[14] Nader, M., von Garssen, H.-G., Krommer, M., and Irschik, H., 2003, "Piezoelectric Actuation of Thin Shells with Support Actuation," Proc. SPIE, **5049**, pp. 180–189.

[15] Eringen, A. C., and Maugin, G. A., 1990, *Electrodynamics of Continua I*, Springer, New York.

[16] Krommer, M., 2003, "On the Significance of Nonlocal Constitutive Relations for Composite Thin Plates Including Piezoelastic Layers With Prescribed Electric Charge," Smart Mater. Struct., **12**, pp. 318–330.

[17] Graff, K. F., 1975, *Wave Motion in Elastic Solids*, Clarendon Press, Oxford, UK.

The Viscoelastic Fiber Composite with Nonlinear Interface

Mayue Xie

Alan J. Levy

Department of Mechanical and Aerospace Engineering,
Syracuse University,
Syracuse, NY 13244-1240

*Effective viscoelastic response of a unidirectional fiber composite with interfaces that may separate or slip according to uniform Needleman-type cohesive zones is analyzed. Previous work on the solitary elastic composite cylinder problem leads to a formulation for the mean response consisting of a stress-strain relation depending on the interface separation/slip discontinuity together with an algebraic equation governing its evolution. Results for the fiber composite follow from the composite cylinders representation of a representative volume element (RVE) together with variational bounding. Here, the theory is extended to account for viscoelastic matrix response. For a solitary elastic fiber embedded in a cylindrical matrix which is an n*th*-order generalized Maxwell model in shear relaxation, a pair of nonlinear n*th*-order differential equations is obtained which governs the relaxation response through the time dependent stress and interface separation/slip magnitude. When the matrix is an n*th*-order generalized Kelvin model in shear creep, a pair of nonlinear n*th*-order differential equations is obtained governing the creep response through the time dependent strain and interface separation/slip magnitude. We appeal to the uniqueness of the Laplace transform and its inverse to show that these equations also apply to an RVE with the composite cylinders microstructure. For a matrix, which is a standard linear solid ($n=2$), the governing equations are analyzed in detail paying particular attention to issues of bifurcation of response. Results are obtained for transverse bulk response and antiplane shear response, while axial tension with related lateral Poisson contraction and transverse shear are discussed briefly. The paper concludes with an application of the theory to the analysis of stress relaxation in the pure torsion of a circular cylinder containing unidirectional fibers aligned parallel to the cylinder axis. For this problem, the redistribution of shear stress and interface slip throughout the cross section, and the movement of singular surfaces, are investigated for an interface model that allows for interface failure in shear mode.*
[DOI: 10.1115/1.2083807]

1 Introduction

This paper presents an analysis of the effective response of randomly arrayed unidirectional viscoelastic fiber composites containing interfaces whose separation behavior can be captured by uniform Needleman-type cohesive zones of vanishing thickness [1]. Ultimately, the goal is to extend the results of Hashin's effective property analyses of rigid interface viscoelastic composites [2,3] to the realm of distributed composite damage. Here, only damage by interfacial debonding and decohesion are treated. Other damage mechanisms that may be active in composite materials exhibiting viscoelastic response, such as matrix or fiber cracking, crazing, etc, are not considered in this paper.

To begin, consider the solution of the elastic nonlinear interface problem [4–6]. Because there is no angular dependence of the elastic fields for both axial tension response and transverse equibiaxial tension response, the constitutive relations for a single-composite cylinder are, apart from the form of the coefficients, identical in structure. The constitutive relation for antiplane shear, which does have angular dependence of the elastic fields, can be put in the same form as that of axial tension and transverse equibiaxial tension provided a one-mode approximation to the fiber-matrix interfacial slip is utilized and the interface force is correctly interpreted (see Eq. (28) and the paragraph preceding it). Note that in all cases considered in this paper, we do not allow for axial coordinate dependence of the elastic or viscoelastic fields. In the case of interface failure, this amounts to the highly idealized situation in which failure occurs uniformly at all values of axial coordinate along the interface. In this paper, antiplane shear and transverse equibiaxial tension are given detailed treatment but axial tension is not. This is because the effect on axial tension response (and associated Poisson contraction) of fiber-matrix interface separation, due to differential contraction between fibers and matrix, has been shown in [5] to be negligible. (Because of this, viscoelastic response for this loading case may be obtained by assuming that the interfaces are rigid as has been done in [3].)

The approach taken is to consider both antiplane shear and transverse equibiaxial tension loading together in one general framework, and to consider a specific load case only when detailed results are desired. This unified treatment brings out features common to both cases but at the expense of some additional algebraic complexity. The general form of the elastic composite cylinder constitutive relations, valid for both loading configurations, is,

$$\begin{aligned} 0 &= F_\varepsilon(\sigma,\varepsilon,\omega,\mu^+) = -\sigma + (g_{11}\mu^+ + g_{12})\varepsilon + (g_{13}\mu^+ + g_{14})\omega \\ &\quad + (g_{15}\mu^+ + g_{16})\sigma_{\max} f(\omega), \\ 0 &= G_\varepsilon(\varepsilon,\omega,\mu^+) = (g_{21}\mu^+ + g_{22})\varepsilon + (g_{23}\mu^+ + g_{24})\omega + (g_{25}\mu^+ \\ &\quad + g_{26})\sigma_{\max} f(\omega). \end{aligned} \tag{1}$$

Note that in Eq. (1), the functional dependence on matrix shear modulus μ^+ is made explicit because ultimately, it will be the only time dependent material property in the following development. In Eq. (1_1) σ, ε are corresponding mean stress and strain measures (e.g., axial tension, transverse tension, and antiplane shear). The

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received January 11, 2005; final manuscript received July 22, 2005. Review conducted by 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.

coefficients $\{g_{ij},\ i=1,2,\ j=1,\ldots,6\}$ assume different forms depending on the loading configuration (see Eqs. (20) and (29), and [5]) and are generally functions of fiber transverse bulk modulus κ^- and fiber shear modulus μ^-, three dimensional matrix bulk modulus[1] k^+, as well as the fiber volume concentration c. They may also depend on a nondimensional interface constitutive parameter ρ (characteristic interface force length/fiber radius R_1) which arises in the description of the interface force law f. Relation (1_1) is the effective property relation for a single composite cylinder. The internal variable ω is the uniform interface displacement discontinuity[2] (normalized with respect to characteristic interface force length) and appears in Eq. (1_1) through expressions for the uniform strain in the fiber phase and a component of the displacement discontinuity tensor[3]. Relation (1_2) is essentially the interface boundary condition for the composite cylinder. It depends explicitly on the component of interface traction f (normalized with respect to interface strength $\sigma_{\max}$) and therefore, on the interface traction-displacement discontinuity constitutive relation $[f(\omega)]$. The stress-strain relation (1), for a single-composite cylinder, can be shown to apply (exactly for axial tension and transverse bulk response or, in approximation for antiplane shear) for an entire representative volume element (RVE) composed of the composite cylinders microstructure [4–6]. First, the elastic fields are obtained from the uniform strain boundary value problem and the uniform stress boundary value problem. They are then used to construct the potential and complementary energies (including interfacial energy), which are then shown to bound the exact response of the RVE. It has been demonstrated for axial tension [5] and transverse equibiaxial load [4] that, similar to the rigid and linear interface problems [7], the bounds coincide and therefore an exact solution governed by Eq. (1) exists. For the antiplane shear case, it was demonstrated in [6] that, unlike the rigid or linear interface composite [7], the bounds do not coincide [this is because a uniform traction (strain) applied to the boundary does not result in a uniform boundary strain (traction)]. The bounds do however differ by a term of order $O(c^4)$ so that the effective response may be accurately predicted by either a uniform stress or strain boundary condition provided that c is not excessively large.

The extension to viscoelastic matrix response begins with Eq. (1). In the next section, the general theory is established. The equations for the solitary viscoelastic composite cylinder problem under general boundary conditions are presented focusing on relaxation and creep conditions. Specific forms of the governing equations are obtained for a composite that consists of elastic fibers embedded in a matrix that is elastic in dilatation and an nth-order generalized Maxwell (Kelvin) model in shear relaxation (creep). A potential interface force constitutive model is employed that allows for interface failure in either normal or shear mode. Issues of bifurcation of response are treated qualitatively. This is followed by a discussion on the relationship between the response of an RVE composed of the composite cylinders microstructure and the response of a solitary composite cylinder. Owing to nonlinear interface characterization the elastic-viscoelastic analogy is no longer available as the method to map elastic composite behavior to viscoelastic composite behavior. In order to show that solitary composite cylinder response applies to a composite RVE, we appeal to the fact that in the Laplace transform domain the composite is elastic and therefore energies can provide variational bounds. For a matrix that is a standard solid in shear and an interface characterized by an exponential force law in normal and shear mode [8], model predictions for transverse equibiaxial tension and antiplane shear are obtained. The paper concludes with a section on the application of the antiplane shear constitutive relation to the analysis of stress relaxation in the pure torsion of a circular cylinder containing unidirectional fibers aligned parallel to the cylinder axis. For this problem, the redistribution of shear stress and interface slip throughout the cross section, and the movement of singular surfaces, are investigated for an (exponential) interface model that allows for interface failure in shear mode.

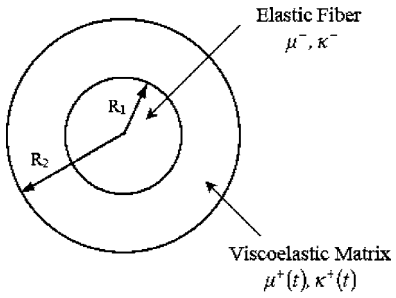


Fig. 1 The viscoelastic composite cylinder

2 General Formulation

The equations governing the viscoelastic response of a composite cylinder are obtained from the equations governing elastic response (1) by use of the Laplace transform. Assume that the composite cylinder (Fig. 1) is composed of a viscoelastic fiber, with time dependent moduli $\mu^-(t), \kappa^-(t)$, embedded in a viscoelastic matrix with time dependent moduli $\mu^+(t), \kappa^+(t)$. The mean stress, strain, and interface separation are now time dependent so that $\sigma(t), \varepsilon(t), \omega(t)$, respectively. Of primary interest for the viscoelastic composite cylinder, is the relaxation response through the time dependent mean stress and interface separation magnitude, and the creep response through the time dependent strain and interface separation magnitude.

2.1 Relaxation Response. Assume the boundary condition on the outer surface $(r=R_2)$ is a uniform time dependent strain $\varepsilon(t)$. The elastic-viscoelastic analogy is not applicable in the present circumstances because the Laplace transformed viscoelastic solution cannot be obtained from the elastic solution [governed by Eq. (1)] by employing the following replacement scheme $\mu^{\pm} \rightarrow s\bar{\mu}^{\pm}(s)$, $\kappa^{\pm} \rightarrow s\bar{\kappa}^{\pm}(s)$, where an overbar indicates Laplace transform and s is the transform variable. The reason is that for the nonlinear interface force-separation relation f, $L[f(\omega(t))] \neq f(L[\omega(t)])$, where L is the Laplace operator. However, the equations governing the relaxation response in the s-domain may be obtained from Eq. (1) by applying the Laplace transformation,

$$\begin{aligned}\bar{\sigma}(s) &= [g_{11}s\bar{\mu}^+(s) + g_{12}]\bar{\varepsilon}(s) + [g_{13}s\bar{\mu}^+(s) + g_{14}]\bar{\omega}(s) + [g_{15}s\bar{\mu}^+(s) \\ &\quad + g_{16}]\sigma_{\max}\bar{f}(s) \\ 0 &= [g_{21}s\bar{\mu}^+(s) + g_{22}]\bar{\varepsilon}(s) + [g_{23}s\bar{\mu}^+(s) + g_{24}]\bar{\omega}(s) + [g_{25}s\bar{\mu}^+(s) \\ &\quad + g_{26}]\sigma_{\max}\bar{f}(s),\end{aligned} \tag{2}$$

where $\bar{f}(s)=L[f(\omega(t))]=L[f_\omega(t)]$. Note that in Eq. (2), the physically reasonable assumption that the fiber is elastic and the matrix is elastic in dilatation has been made so that, $\kappa^+(t)=k^+ + \frac{1}{3}\mu^+(t)$, where k^+ is the (constant) three-dimensional elastic bulk modulus. In this case, the coefficients g_{ij} [in Eq. (2)] are constant and the inversion of the Laplace transformed Eq. (2) leads to an integral form of the governing equations:

$$\sigma(t) = \int_0^t \dot{\mu}^+(t-T)K_1(T)dT + A_1\varepsilon(t) + A_2\omega(t) + A_3\sigma_{\max}f(\omega(t)), \tag{3}$$

[1]Note that $\kappa=\lambda+\mu=k+\mu/3$ where λ, μ are Lame moduli

[2]For antiplane shear, it is the mode multiplier in the single-mode approximation [see Eq. (28_2)].

[3]The difference between the applied boundary strain and the mean strain of the composite cylinder.

$$0 = \int_0^t \dot{\mu}^+(t-T)K_2(T)dT + B_1\varepsilon(t) + B_2\omega(t) + B_3\sigma_{\max}f(\omega(t)).$$

In Eq. (3),

$$A_1 = g_{11}\mu^+(0) + g_{12}; \quad A_2 = g_{13}\mu^+(0) + g_{14}; \quad A_3 = g_{15}\mu^+(0) + g_{16},$$

$$B_1 = g_{21}\mu^+(0) + g_{22}; \quad B_2 = g_{23}\mu^+(0) + g_{24}; \quad B_3 = g_{25}\mu^+(0) + g_{26}, \tag{4}$$

$$K_1(t) = g_{11}\varepsilon(t) + g_{13}\omega(t) + g_{15}\sigma_{\max}f(\omega(t)),$$

$$K_2(t) = g_{21}\varepsilon(t) + g_{23}\omega(t) + g_{25}\sigma_{\max}f(\omega(t)),$$

where K_1, K_2 are linear combinations of functions $\varepsilon(t)$, $\omega(t)$, $f(\omega(t))$, and the A_i, B_i coefficients are constants.

In order to explore some of the features of Eq. (3) assume that the matrix is a generalized nth-order Maxwell model in shear relaxation:

$$\mu^+(t) = \sum_{i=1}^{n} \mu_i^+ e^{-\lambda_i t}, \quad \text{or equivalently,}$$

$$\mu^+(t) = \sum_{i=1}^{n} \mu_i(t), \quad \dot{\mu}_i(t) = -\lambda_i\mu_i(t), \quad \mu_i(0) = \mu_i^+, \tag{5}$$

where the μ_i^+, λ_i are constants. It is shown in the Appendix that assumption (5) allows Eq. (3) to be reduced to a set of two autonomous nonlinear nth order differential equations for the stress $\sigma(t)$ and interface displacement jump $\omega(t)$ as given (in the Appendix) by Eq. (A4).

The stress relaxation behavior may be obtained by applying a strain history of the form $\varepsilon(t)=H(t)\varepsilon_0$ [$H(t)$ is the Heaviside step function, ε_0 a constant]. The initial values $\sigma_0=\sigma(t=0)$, $\omega_0=\omega(t=0)$ are obtained from Eq. (3) and are given by

$$0 = F_\varepsilon(\sigma_0, \varepsilon_0, \omega_0, \mu^+(0))$$

$$0 = G_\varepsilon(\varepsilon_0, \omega_0, \mu^+(0)) \tag{6}$$

where F_ε and G_ε have been defined in Eq. (1). Thus, the initial values (σ_0, ω_0) are just the elastic response (1) with shear modulus $\mu^+=\mu^+(0)$. Equilibrium solutions (σ_e, ω_e), of the differential equations (A4) can be obtained by setting $\sigma^{(k)}=0$, $\omega^{(k)}=0$, $k=1..n$. The result is Eq. (6) with $\sigma_0, \omega_0, \mu^+(0)$ replaced by $\sigma_e, \omega_e, \mu^+(\infty)$, respectively. The long-term relaxation behavior $(\sigma_\infty, \omega_\infty)$ follows from, $\sigma_\infty=\operatorname{Lim}_{t\to\infty}\sigma(t)$, $\omega_\infty=\operatorname{Lim}_{t\to\infty}\omega(t)$ and limiting values, when they exist, can be obtained from the Tauberian or limiting value theorems [9,10], $\operatorname{Lim}_{t\to\infty}\sigma(t)=\operatorname{Lim}_{s\downarrow 0}[s\bar{\sigma}(s)]$, $\operatorname{Lim}_{t\to\infty}\omega(t)=\operatorname{Lim}_{s\downarrow 0}[s\bar{\omega}(s)]$. Application of this theorem to the Laplace transformed governing equations (2) implies that the long-term values satisfy Eq. (6) with $\sigma_0, \omega_0, \mu^+(0)$ replaced by $\sigma_\infty, \omega_\infty, \mu^+(\infty)$, respectively. In obtaining this last result, the following relations have been employed, $\operatorname{Lim}_{s\downarrow 0}[s\bar{f}(s)]=\operatorname{Lim}_{t\to\infty} f_\omega(t)=\operatorname{Lim}_{t\to\infty} f[\omega(t)]=f(\omega_\infty)$. A comparison of the equations governing equilibrium solutions and long-term solutions indicates that both (σ_e, ω_e) and $(\sigma_\infty, \omega_\infty)$ are solutions of the same system of equations and therefore identical. This is to be expected because after a long time has elapsed, the viscoelastic matrix will have essentially relaxed to its equilibrium state. Further examination of these equations as well as Eq. (6) (which governs initial behavior) indicates that they all govern elastic response (1) with shear modulus $\mu^+(\infty)$ (in the case of long-term and equilibrium response) and with shear modulus $\mu^+(0)$ (in the case of initial response).

For a matrix which is a standard solid in shear, $n=2$ and $\lambda_1=0$. In this case, (A4) reduces to two first-order equations of the form

$$\dot{\sigma} + \sigma = a_2(\omega),$$

$$a_0(\omega)\dot{\omega} + a_1(\omega) = 0. \tag{7}$$

In Eq. (7), a superimposed dot indicates derivative with respect to the normalized time $\hat{t}=\lambda_2 t$. The coefficients a_0, a_1, a_2 are related to the functions F_ε and G_ε defined in Eq. (1), i.e.,

$$a_0(\omega) = D_\omega G_\varepsilon(\varepsilon_0, \omega, \mu^+(0)) = B_2 + B_3\sigma_{\max}Df(\omega)$$

$$a_1(\omega) = G_\varepsilon(\varepsilon_0, \omega, \mu^+(\infty)) = (\mu_1^+ g_{21} + g_{22})\varepsilon_0 + (\mu_1^+ g_{23} + g_{24})\omega + (\mu_1^+ g_{25} + \mu_2^+ g_{26})\sigma_{\max}f(\omega)$$

$$a_2(\omega) = -D_\omega F_\varepsilon(\varepsilon_0, \omega, \mu^+(0))\frac{G_\varepsilon(\varepsilon_0, \omega, \mu^+(\infty))}{DG_\varepsilon(\varepsilon_0, \omega, \mu^+(0))} + [F_\varepsilon(\sigma, \varepsilon_0, \omega, \mu^+(\infty)) + \sigma] = (\mu_1^+ g_{11} + g_{12})\varepsilon_0 + (\mu_1^+ g_{13} + g_{14})\omega + (\mu_1^+ g_{15} + \mu_2^+ g_{16})\sigma_{\max}f(\omega) + [A_2 + A_3\sigma_{\max}Df(\omega)]\dot{\omega}, \tag{8}$$

where $Df(\omega)$ represents the derivative of $f(\omega)$ with respect to its argument.

2.2 Bifurcation. The differential equations (7) are of the form $\dot{x}=h(x,\chi)$ with $x=[\sigma,\omega]^T$, $h=[a_2-\sigma, -a_1/a_0]^T$ and $\chi=\varepsilon_0$. The initial conditions for this system are not arbitrary but predetermined to be solutions of the nonlinear algebraic equations (6), i.e., the constitutive characterization of the constituents fixes the initial elastic response, which in turn influences the evolutionary response. Critical behavior occurs when: (i) initial states suffer bifurcation, and (ii) h ceases to exist. For (i), bifurcation of solutions of Eq. (6) means that initial conditions (σ_0, ω_0) suffer a jump discontinuity under increasing values of applied strain ε_0. In the case of (ii), the nonexistence of h means that the $[\sigma(t), \omega(t)]$ have a jump discontinuity at a fixed instant in time. It is clear from the a_i coefficients defined in Eq. (8) that h may indeed fail to exist at certain values of ω. Note that equilibrium or long-term response defined by Eq. (6) [with $\sigma_0, \omega_0, \mu^+(0)$ replaced by $\sigma_e, \omega_e, \mu^+(\infty)$] may suffer a jump discontinuity as well. An interesting feature of the response of this class of composite is that bifurcation of initial condition, bifurcation of equilibria or long-term response, and nonexistence of h at finite values of x, are all governed by elastic equations of the same basic form. Thus, bifurcation of equilibria is governed by

$$0 = G_\varepsilon(\varepsilon_0, \omega_e, \mu^+(\infty)),$$

$$0 = D_{\omega_e}G_\varepsilon(\omega_e, \mu^+(\infty)), \tag{9}$$

$$0 = F_\varepsilon(\sigma_e, \varepsilon_0, \omega_e, \mu^+(\infty)),$$

where the second equation, which is independent of ε_0, determines ω_e, the first equation in (9) yields ε_0, and the third equation gives σ_e. Bifurcation of initial states is governed by the same system (9), provided the terminal value $\mu^+(\infty)$ is replaced by the initial value $\mu^+(0)$ [recall Eq. (5) with $n=2$ and $\lambda_1=0$]. An examination of the differential equations (7) and (8) reveals that h ceases to exist when

$$a_0(\omega) = D_{\omega_0}G_\varepsilon(\omega, \mu^+(0)),$$

$$= B_2 + B_3\sigma_{\max}Df(\omega) = 0. \tag{10}$$

Because a_2 ceases to exist when a_0 vanishes, we have [from Eq. (7)] the fact that the mean stress $\sigma(t)$ and the interface separation

$\omega(t)$ bifurcate simultaneously. Furthermore, note that Eq. (10) is precisely the condition (9_2) [with $\mu^+(\infty)$ replaced by the initial value $\mu^+(0)$] required for bifurcation of equilibrium states. Thus, we have the fact that bifurcation of long-time equilibrium states is governed by elastic equations with terminal shear relaxation $\mu^+(\infty)$, bifurcation of initial states is governed by the same elastic equations except with initial shear relaxation $\mu^+(0)$ [replacing $\mu^+(\infty)$] and nonexistence of h at finite values of ω is governed by an identical elastic condition (10) required for bifurcation of initial states. Physically, the breakdown in h is the most interesting since it gives rise to a discontinuity in the separation and the stress at some stage in their evolution. Because B_2, B_3 are generally positive, Eqs. (8) and (10) indicate that bifurcation occurs when $Df(\omega)<0$, i.e., the interface force is on the descending branch of the interface force-separation curve.

2.3 Creep Response. A parallel development applies for creep response. Thus, for a traction boundary condition applied to the composite cylinder, the equations governing elastic response are:

$$\begin{aligned} 0 &= F_\sigma(\sigma,\varepsilon,\omega,J^+) = -\varepsilon + (q_{11}J^+ + q_{12})\sigma + (q_{13}J^+ + q_{14})\omega + (q_{15}J^+ + q_{16})\sigma_{\max} f(\omega), \\ 0 &= G_\sigma(\sigma,\varepsilon,\omega,J^+) = (q_{21}J^+ + q_{22})\sigma + (q_{23}J^+ + q_{24})\omega + (q_{25}J^+ + q_{26})\sigma_{\max} f(\omega) + (q_{27}J^+ + q_{28})\varepsilon, \end{aligned} \tag{11}$$

which are analogous to the equations governing elastic response of a composite cylinder subject to an applied strain boundary condition (1). Because effective response for this class of composite is self-consistent [4–6], Eqs. (1) and (11) are reciprocal.[4] Apply to the outer surface of a composite cylinder (with unit normal $\mathbf{n}=\mathbf{e}_\mathbf{r}$) the uniform time dependent traction $\sigma(t)=H(t)\sigma_0$, where σ_0 is constant. Assume that matrix response is governed by a generalized nth-order Kelvin model in shear creep so that

$$\begin{aligned} J^+(t) &= \sum_{i=1}^{n} J_i^+(1-e^{-\lambda_i' t}) + \lambda_{n+1}' t, \quad \text{or equivalently,} \\ J^+(t) &= \sum_{i=1}^{n} J_i(t) + \lambda_{n+1}' t, \quad \dot{J}_i(t) = \lambda_i'[J_i^+ - J_i(t)], \end{aligned} \tag{12}$$

where J_i^+, λ_i' are constants. By employing arguments similar to those used to derive relaxation response, it can be shown that two nth-order nonlinear differential equations of form (A4) govern $\varepsilon(t), \omega(t)$.

The initial values $\varepsilon_0=\varepsilon(\hat{t}=0)$, $\omega_0=\omega(\hat{t}=0)$ are governed by

$$\begin{aligned} 0 &= F_\sigma(\sigma_0,\varepsilon_0,\omega_0,J^+(0)), \\ 0 &= G_\sigma(\sigma_0,\varepsilon_0,\omega_0,J_{(0)}^+), \end{aligned} \tag{13}$$

and represent the initial elastic response (11) with creep modulus $J^+(0)$.

Long-term and equilibrium solutions are identical to each other and describe the elastic response with creep modulus $\mu^+(\infty)$, i.e., Eq. (13) with $J^+(0)$ replaced by $J^+(\infty)$ and (ε_0,ω_0) replaced by (ε_e,ω_e) or $(\varepsilon_\infty,\omega_\infty)$. When $n=2$, $\lambda_1'\to\infty$, $\lambda_3'=0$, we recover a matrix that is a standard solid in shear creep function (as well as elastic in bulk response), i.e., $J^+(t)=J_0^+-J_2^+e^{-\lambda' 2t}, J_0^+=J_1^++J_2^+$, where the creep constants (J_0^+,J_2^+,λ'_2) can be expressed in terms of relaxation constants $\mu_1^+,\mu_2^+,\lambda_2$. When this is the case, the equations governing the creep response reduce to two-first order equations of the form,

$$\begin{aligned} \dot{\varepsilon} + \frac{1}{1+\kappa_\mu}\varepsilon &= b_2(\omega), \\ b_0(\omega)\dot{\omega} + b_1(\omega) &= 0, \end{aligned} \tag{14}$$

where, as before, we have assumed that a superimposed dot indicates derivative with respect to the normalized time $\hat{t}=\lambda_2 t$ and $\kappa_\mu=\mu_2^+/\mu_1^+$. The coefficients b_0,b_1,b_2 are related to the functions F_σ and G_σ defined in Eq. (11), i.e.,

$$\begin{aligned} b_0(\omega) &= q_{23}J_0^+ + q_{24}(1+\kappa_\mu) + [q_{27}J_0^+ + q_{28}(1+\kappa_\mu)][q_{13}J^+(0)+q_{14}] + \{q_{25}J_0^+ + q_{26}(1+\kappa_\mu) + [q_{27}J_0^+ + q_{28}(1+\kappa_\mu)][q_{15}J^+(0) + q_{16}]\}\sigma_{\max} Df(\omega) \\ b_1(\omega) &= \{q_{21}J_0^+ + q_{22} + [q_{27}J^+(0)+q_{28}](q_{11}J_0^+ + q_{12})\}\sigma_0 + \{q_{23}J_0^+ + q_{24} + [q_{27}J^+(0)+q_{28}](q_{13}J_0^+ + q_{14})\}\omega + \{q_{25}J_0^+ + q_{26} + [q_{27}J^+(0)+q_{28}](q_{15}J_0^+ + q_{16})\}\sigma_{\max} f(\omega) \\ b_2(\omega) &= \frac{1}{1+\kappa_\mu}[(q_{11}J_0^+ + q_{12})\sigma_0 + (q_{13}J_0^+ + q_{14})\omega + (q_{15}J_0^+ + q_{16})\sigma_{\max} f(\omega)] + [(q_{13}J^+(0)+q_{14}) + (q_{15}J^+(0) + q_{16})\sigma_{\max} Df(\omega)]\dot{\omega} \end{aligned} \tag{15}$$

The differential equations (14) are of the form $\dot{x}=h(x,\chi)$ with $x=[\varepsilon,\omega]^T$, $h=[b_2-\varepsilon/(1+\kappa_\mu),-b_1/b_0]^T$ and $\chi=\sigma_0$. The vector field h ceases to exist when

$$0 = b_0(\omega) = q_{23}J_0^+ + q_{24}(1+\kappa_\mu) + [q_{27}J_0^+ + q_{28}(1+\kappa_\mu)][q_{13}J^+(0) + q_{14}] + \{q_{25}J_0^+ + q_{26}(1+\kappa_\mu) + [q_{27}J_0^+ + q_{28}(1+\kappa_\mu)][q_{15}J^+(0) + q_{16}]\}\sigma_{\max} Df(\omega). \tag{16}$$

From Eq. (15_3) b_2 ceases to exist when b_0 vanishes. Therefore, the mean effective strain $\varepsilon(t)$ and the interface separation $\omega(t)$ bifurcate simultaneously. It can be shown that the bifurcation condition for initial states $D_\omega G_\sigma(\omega,J^+(0))$ is precisely the condition (16) required for the nonexistence of h. Thus, we have the fact that bifurcation of long-time equilibrium states is governed by elastic equations with terminal creep modulus $J^+(\infty)$, bifurcation of initial states is governed by the same elastic equations with initial creep modulus $J^+(0)$, and nonexistence of h at finite values of ω is governed by an identical elastic condition (16) required for bifurcation of initial states. This is entirely analogous to relaxation behavior. Equation (16) is of the form (10), which has been considered previously.

3 Effective Composite Response

Here, we demonstrate that the effective response of a viscoelastic nonlinear interface composite RVE, composed of the composite cylinders microstructure [7], is governed by the equations for the single-composite cylinder. Recall that the composite cylinders microstructure consists of composite cylinders of varying dimensions down to the vanishingly small, such that the fiber volume concentration c remains constant for each composite cylinder.[5] Conceptually, the effective response of such a composite RVE is expected to relate to the response of the individual composite cylinders comprising it. For the elastic case, variational arguments can be used to show that the effective response of a nonlinear interface composite cylinder also applies to an entire composite RVE [4–6]. For the viscoelastic RVE, there is no appropriate energy bounding theorem so a different approach must be adopted. When the composite has rigid or linear interfaces, the standard

[4]For antiplane shear, this is only approximately true [6].

[5]An additional constraint is that the force length ratio ρ remains constant for each composite cylinder [4].

tool to pass from the response of a single viscoelastic composite cylinder to the effective response of a viscoelastic RVE is the elastic-viscoelastic analogy [2,3]. As stated previously, the elastic-viscoelastic analogy does not apply here because, for the nonlinear interface force separation relation, $L[f(\omega(t))] \neq f(L[\omega(t)])$. In this section, another approach is developed based on the uniqueness of Laplace transform and the fact that a viscoelastic composite cylinder corresponds to an elastic composite cylinder in the transform domain.

First, if the force separation law $f(\omega)$ is potential, then there exists a function $p(\bar{\omega}(s))$ such that, $p(\bar{\omega}(s))=p(L[\omega(t)])=\bar{f}(s)=L[f(\omega(t)\ ]$, with $p(\bar{\omega}(s))$ potential in $\bar{\omega}(s)$. To show this, assume that functions $f\circ\omega=f_\omega$ and ω are integrable over a finite interval [a,b], and that they are of exponential order. Then it can be proven that $\bar{f},\bar{\omega}$ are analytic functions of s (see [9]). Now $d\bar{\omega}/ds\neq 0$, $\mathrm{Re}(s)>a$, a a real constant. This follows by differentiating the definition of the Laplace transform of $\bar{\omega}(s)$ with respect to s and noting that $\omega(t)$ does not vanish on $(0,\infty)$. Thus, since $d\bar{\omega}/ds\neq 0, \mathrm{Re}(s)>a$; then, by the inverse function theorem, there is an analytic inverse $s=\bar{\omega}^{-1}(\bar{\omega}(s))$. The composite function $p=\bar{f}\circ\bar{\omega}^{-1}$ is therefore analytic and potential.

Equations (2) governing the relaxation response for a single-composite cylinder in the s domain can now be written as

$$\begin{aligned}\bar{\sigma}(s) &= [g_{11}(s)s\bar{\mu}^+(s)+g_{12}(s)]\bar{\varepsilon}(s)+[g_{13}(s)s\bar{\mu}^+(s)+g_{14}(s)]\bar{\omega}(s)\\ &\quad+[g_{15}(s)s\bar{\mu}^+(s)+g_{16}(s)]\sigma_{\max}p(\bar{\omega}(s))\\ 0 &= [g_{21}(s)s\bar{\mu}^+(s)+g_{22}(s)]\bar{\varepsilon}(s)+[g_{23}(s)s\bar{\mu}^+(s)+g_{24}(s)]\bar{\omega}(s)\\ &\quad+[g_{25}(s)s\bar{\mu}^+(s)+g_{26}(s)]\sigma_{\max}p(\bar{\omega}(s)).\end{aligned} \tag{17}$$

The form of Eq. (17) suggests that the viscoelastic composite cylinder in the t domain can be viewed as a geometrically similar elastic composite cylinder in the s domain, with the interface force separation law $f(\omega)$ being replaced by its Laplace transformed form $p(\bar{\omega}(s))$ and the material properties being modified according to the replacement scheme $\mu^\pm\rightarrow s\bar{\mu}^\pm(s)$, $\kappa^\pm\rightarrow s\bar{\kappa}^\pm(s)$. In other words, if V_i represents the equations for a viscoelastic composite cylinder in the t domain, then there are unique equations E_i for an elastic composite cylinder in the s domain such that, $E_i=L(V_i)$ or $V_i=L^{-1}(E_i)$. Now in the s domain where the composite is elastic, it can be proven[6] [4–6] that the effective response of an RVE is governed by the same equations governing the elastic composite cylinder, i.e., [Eq. (17)]. This suggests that, for the RVE in the s domain, there is an effective homogeneous continuum with unique equations E satisfying, $E=E_i$. Finally, by application of the inverse Laplace transform, there exist unique equations V governing the response of a viscoelastic homogeneous continuum in the t domain, $V=L^{-1}(E)$. Because $E=E_i$, sets E and E_i must govern the same effective response in the s domain. Because of the uniqueness of the Laplace transform and its inverse, V and V_i will predict the same effective response in the t domain, i.e.,

$$V=L^{-1}(E)=L^{-1}(E_i)=L^{-1}(L(V_i))=V_i. \tag{18}$$

Therefore, the response predicted by Eq. (7) is identical to that of Eq. (14), and both predict the response of an RVE of the viscoelastic composite.

The above argument applies to transverse equibiaxial tension and axial tension where energy bounds for the elastic problem coincide. For antiplane shear, where the energy bounds (for the nonlinear interface problem) do not coincide, a modification is required. Let V_i^γ, V_i^τ represent sets of equations for the solitary viscoelastic composite cylinder in the t domain under uniform strain and uniform stress boundary conditions, respectively. Then there are unique sets of equations E_i^γ, E_i^τ in the s domain, in the form governing an elastic composite cylinder, such that, $E_i^\gamma=L(V_i^\gamma)$, $E_i^\tau=L(V_i^\tau)$. Now for the elastic problem, it has been shown in [6] using variational bounding that the effective antiplane shear response of the RVE is bounded by solutions to the equations governing the elastic composite cylinder under uniform strain and uniform stress boundary conditions, respectively. The associated energy bounds can be shown to differ by a term of order $O(c^4)$. Thus, in the s domain where the composite cylinder is governed by equations that are elastic in form this suggests that, for the RVE in the s domain, there is an effective homogeneous continuum, whose effective response is bounded by solutions to unique equations E^γ, E^τ satisfying, $E^\gamma=E_i^\gamma$, $E^\tau=E_i^\tau$. The difficulty arises when we try to apply the inverse Laplace transform to obtain bounds on the response in the t domain. Generally, bounds on the Laplace transform do not convert into bounds on the inverse transformed function. However, because the difference in the "energy" bounds in the s domain is proportional to the product of a function of s and a constant term of order $O(c^4)$ we can assert that if the response of E^γ and E^τ is nearly the same (i.e., for c not too large) in the s domain, then in the t domain, $V^\gamma=L^{-1}(E^\gamma)$ and $V^\tau=L^{-1}(E^\tau)$ will be nearly the same as well. Because of the uniqueness of the Laplace transform and its inverse, V^γ, V^τ should predict the same effective response as V_i^γ, V_i^τ, respectively, in the t domain, i.e.,

$$\begin{aligned}V^\gamma &= L^{-1}(E^\gamma)=L^{-1}(E_i^\gamma)=L^{-1}(L(V_i^\gamma))=V_i^\gamma,\\ V^\tau &= L^{-1}(E^\tau)=L^{-1}(E_i^\tau)=L^{-1}(L(V_i^\tau))=V_i^\tau.\end{aligned} \tag{19}$$

4 Transverse Equibiaxial Tension

In this section, we specialize the preceding results to obtain the time dependent effective transverse bulk response of an RVE composed of the composite cylinders microstructure. Interface separation response is assumed to be uniform at all points along the cylindrical interface (Fig. 1). Unless otherwise noted for the remainder of the paper, assume that the matrix is a standard solid in shear. Because it often happens that a composite system consists of a fiber that is stiffer than the matrix by an order of magnitude, (e.g., glass fiber reinforced epoxy matrix), it is further assumed that the fiber is perfectly rigid. In many of the preceding equations, this amounts to setting $\kappa^-\rightarrow\infty$. For transverse equibiaxial tension (see Fig. 1) with an applied strain boundary condition given by displacement vector $\mathbf{u}=\varepsilon(t)R_2\mathbf{e}_r$ on $r=R_2$, the g_{ij} coefficients are:

$$\begin{aligned}&g_{11}=-2,\quad g_{12}=0,\quad g_{13}=2\rho,\quad g_{14}=0,\quad g_{15}=\frac{1}{\kappa^-},\quad g_{16}=1,\\ &g_{21}=\frac{4}{3},\quad g_{22}=k^+,\quad g_{23}=-\rho\left(1+\frac{c}{3}\right),\quad g_{24}=-c\rho k^+,\\ &g_{25}=-\frac{1+c/3}{2\kappa^-},\quad g_{26}=-\frac{\kappa^-(1-c)+ck^+}{2\kappa^-},\end{aligned} \tag{20}$$

which help to determine the coefficients (8) of the differential equations (7). For an applied traction boundary condition, with traction vector $\mathbf{s}(\mathbf{e}_r)=\sigma(t)\mathbf{e}_r$, the q_{ij} coefficients are:

$$q_{11}=-\frac{1}{2},\quad q_{12}=q_{13}=0,\quad q_{14}=\rho,\quad q_{15}=\frac{1}{2},\quad q_{16}=\frac{1}{2\kappa^-}, \tag{21}$$

[6]This is exactly true for equibiaxial and axial tension load. For antiplane shear, bounds are required (see below).

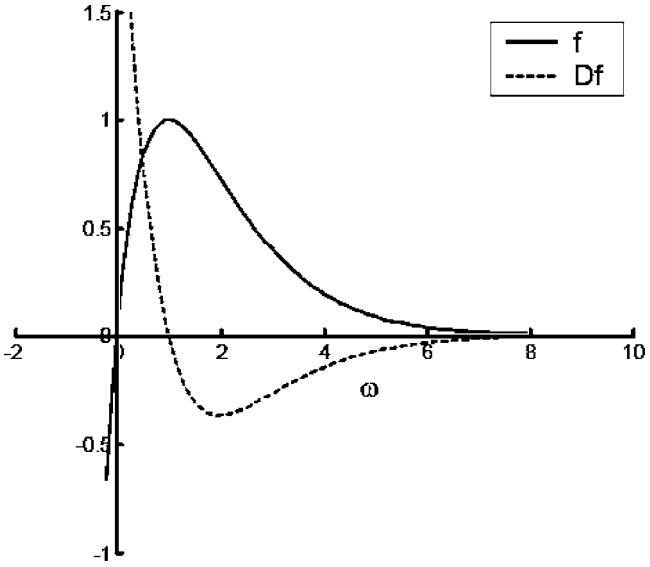


Fig. 2 The normal interface force law *f* and *Df*

$$q_{21}=\frac{1}{2},\quad q_{22}=0,\quad q_{23}=c\rho k^{+},\quad q_{24}=\frac{1}{3}c\rho;\quad q_{25}=\frac{k^{+}-\kappa^{-}}{2\kappa^{-}}c,\quad q_{26}=\frac{c}{6\kappa^{-}},$$

$$q_{27}=-k^{+},\quad q_{28}=-\frac{1}{3},$$

which help to fix the coefficients (15) in the governing equations (14).

In order to proceed, further assume that the interface is characterized by a simple exponential constitutive relation that allows for interface failure in normal mode (Fig. 2),

$$\sigma_{\max}f(\omega)=\sigma_{\max}\omega e^{1-\omega},\quad \sigma_{\max}Df(\omega)=\sigma_{\max}(1-\omega)e^{1-\omega}, \tag{22}$$

where $\omega=w/\rho, w$ is the ratio of the normal interface displacement jump to fiber radius, and ρ is the characteristic interface force length ratio introduced previously. This force law was developed by Ferrante et al. [11] primarily for crystalline interfaces and it, along with its variants, has been utilized in numerous debonding analyses. Here Eq. (22) is regarded as phenomenological with the two parameters $\rho,\sigma_{\max}$ to be determined from experiment. Note that bifurcation occurs when $\omega>1$. This last result follows from Eq. (22) and the fact that bifurcation occurs when $Df(\omega)<0$. Furthermore, Eqs. (9_2) and (10) are of the same form and, when combined with Eq. (22), can be written as

$$\alpha_1(1-\omega)e^{1-\omega}+\alpha_2=0, \tag{23}$$

where α_1,α_2 are positive constants defined in Eqs. (9_2) and (10). Bifurcation condition (23) has been discussed in detail in [12]. Solutions to (23) are given by, $\omega=1-W(-\alpha_2/\alpha_1)$ where $W:x\rightarrow W(x)$ is the multivalued Lambert W function [13] defined to be the solution y to the equation $ye^{y}=x$. Because W is defined on the interval $[-e^{-1},\infty)$, and is such that $W(-e^{-1})=-1$, it follows that a necessary condition for bifurcation is $\alpha_2 e^{1}/\alpha_1\leqslant 1$. A graph of solution ω indicates that there will generally exist two bifurcation points, the exception being when $\alpha_2/\alpha_1=e^{-1}$ when there is only one bifurcation point.

4.1 The Voided Composite. For the voided rigid interface and linear interface composite, the problems are linear and the elastic-viscoelastic correspondence principle may be utilized to obtain the effective relaxation and creep response directly from the effective elastic response. Here, the goal is to recover known results from the general formulation.

For relaxation response of the voided composite set $\sigma_{\max}=0$ in the coefficients (8). The differential equations (7) become linear with solution given by

$$\rho\omega(\hat{t})=\rho\omega_e+(\rho\omega_0-\rho\omega_e)e^{-p_1\hat{t}}$$
$$\hat{\sigma}(\hat{t})=\hat{\sigma}_e-\frac{c_1}{1-p_1}e^{-p_1\hat{t}}+\left[\hat{\sigma}_0-\hat{\sigma}_e+\frac{c_1}{1-p_1}\right]e^{-\hat{t}} \tag{24}$$

where

$$\rho\omega_e=\frac{\kappa_k+4/3}{c\kappa_k+(1+c/3)}\varepsilon_0,\quad \rho\omega_0=\frac{\kappa_k+4(1+\kappa_\mu)/3}{c\kappa_k+(1+c/3)(1+\kappa_\mu)}\varepsilon_0,$$

$$\hat{\sigma}_e=2\varepsilon_0\frac{(1-c)(\kappa_k+1/3)}{c\kappa_k+(1+c/3)},$$

$$\hat{\sigma}_0=2\varepsilon_0\frac{(1-c)[\kappa_k+(1+\kappa_\mu)/3]}{c\kappa_k+(1+c/3)(1+\kappa_\mu)}(1+\kappa_\mu), \tag{25}$$

$$c_1=\frac{c(1-c)\kappa_k^2\kappa_\mu^2\varepsilon_0}{[c\kappa_k+(1+c/3)(1+\kappa_\mu)]^2[c\kappa_k+(1+c/3)]},$$

$$p_1=\frac{c\kappa_k+(1+c/3)}{c\kappa_k+(1+c/3)(1+\kappa_\mu)}.$$

In Eq. (25), the nondimensional constants κ_k,κ_μ, and the normalized stress $\hat{\sigma}$, are defined by

$$\kappa_k=\frac{k^{+}}{\mu_1^{+}},\quad \kappa_\mu=\frac{\mu_2^{+}}{\mu_1^{+}},\quad \hat{\sigma}=\frac{\sigma}{\mu_1^{+}}. \tag{26}$$

Note that $\kappa^{*}(t)$ is the effective time dependent relaxation bulk modulus, $\kappa^{*}(\hat{t})=\hat{\sigma}(\hat{t})\mu_1^{+}/2\varepsilon_0$. The quantities $\hat{\sigma}_0$ and $\hat{\sigma}_e$ are the initial and equilibrium stresses normalized with respect to μ_1^{+}, while $\rho\omega_0$ and $\rho\omega_e$ are the initial and equilibrium interface separations, respectively. It is readily seen from Eq. (25) that $\hat{\sigma}_0>\hat{\sigma}_e$ and $\rho\omega_0<\rho\omega_e$, which means that for the voided composite under relaxation boundary condition, it is always true that the stress relaxes and interface separation increases.

The creep compliance of the voided composite can be obtained from Eqs. (24)–(26), together with the reciprocal relationship between the Laplace transform of the moduli or, by direct integration of Eq. (14), following the same procedure used to obtain the relaxation modulus. It can be shown that $\rho\omega_0<\rho\omega_e$ and $\varepsilon_0<\varepsilon_e$, which means that for the voided composite under creep boundary condition the strain always creeps and interface separation always increases.

4.2 The Rigid Interface Composite. The rigid interface composite corresponds to $\omega=0$ and $\dot{\omega}=0$ so that interface force f is constant and determined by Eqs. (7_2) and (8) in terms of the applied strain ε_0, concentration c, and constitutive parameters. Consider relaxation response. The differential Eqs. (7_1) and (8) together with initial condition (6) have solution for the relaxation bulk modulus,

$$\kappa^{*}(\hat{t})=k^{+}+\frac{1}{3}\mu^{+}(\hat{t})+\frac{c}{1-c}\left[k^{+}+\frac{4}{3}\mu^{+}(\hat{t})\right]. \tag{27}$$

It follows from Eq. (27) that the stress $\sigma(\hat{t})$ will relax as $\mu^{+}(t)$ decreases with time. Equation (27) is consistent with the relaxation modulus obtained in [3].

The bulk creep compliance for the rigid interface composite can be obtained either from the reciprocal relationship between the Laplace transforms of the moduli or, from the governing equations (14) and (15) following the same procedure used to obtain the

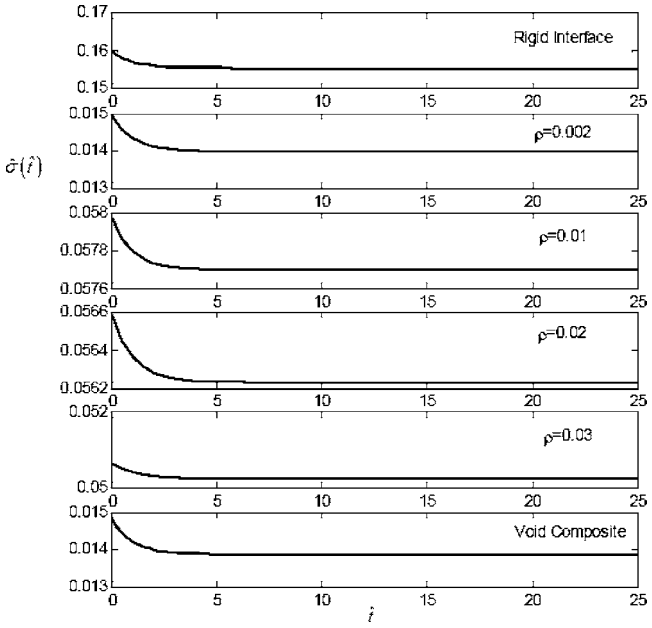


Fig. 3 Stress response under relaxation boundary condition; $\varepsilon_0=0.012$

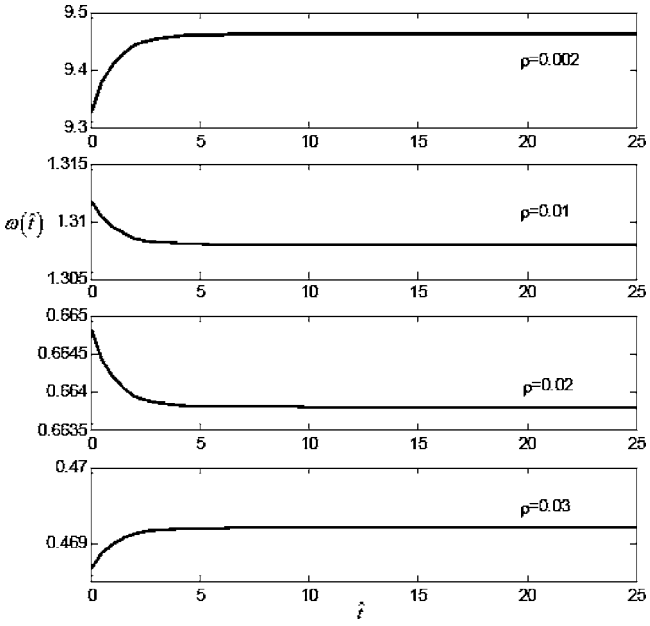


Fig. 4 Interface separation under relaxation boundary condition; $\varepsilon_0=0.012$

relaxation modulus above. Note that while Eq. (27) is generally valid for any $\mu^+(t)$, the effective time dependent bulk creep compliance can only be obtained in closed form for the simplest of viscoelastic models. For the standard solid in shear, it can be shown that the strain $\varepsilon(\hat{t})$ increases with time.

4.3 The Nonlinear Interface Composite. The response of the nonlinear interface composite is obtained by numerical solution of Eq. (7) (relaxation) or Eq. (14) (creep). In the results that follow, assume that the interface can be modeled by the interface force law (22). Let the geometrical and constitutive parameters have the following values: $\kappa_k=2.407$, $\kappa_\mu=9/10$, $c=1/2$, and $\sigma_{\max}=E^+/50$, where E^+ is the matrix elastic modulus.

Consider relaxation response with equibiaxial strain boundary condition such that $\varepsilon(t)=H(t)\varepsilon_0$ and $\varepsilon_0=0.012$. Figure 3 depicts the normalized stress response $\hat{\sigma}(\hat{t})$ for varying characteristic length ratios (ρ). Also included are curves depicting the limiting behavior of the rigid interface composite and the voided composite. Figure 3 indicates that the effect of decreasing ρ is to: (i) decrease the initial stress and (ii) decrease the (long-term) equilibrium stress provided the interface separation is on the descending branch of the interface force-separation curve ($\rho=0.01$, $\rho=0.002$). For the two other cases shown ($\rho=0.03$, $\rho=0.02$), the interface separation is on the ascending branch of the force-separation curve and the opposite trend is observed. Furthermore, for the parameter values shown, the stress relaxes. For extremely small values of ρ however, the stress may creep. In these circumstances, bifurcation can occur (see Fig. 6).

Figures 4 and 5 show interface separation $\omega(t)$ and interface force $f(\omega(t))$, respectively. Several distinct kinds of behavior are observed depending on force length parameter ρ. As indicated in Fig. 4, when $\rho=0.002$ and 0.03, the interface separation $\omega(t)$ does not relax but increases with time. When $\rho=0.03$, this is accompanied by an increasing interface force $f(\omega(t))$ and the process happens prior to the attainment of the interface strength $\sigma_{\max}$ (i.e., on the ascending branch of the interface-separation curve). This kind of behavior gives rise to creep driven loading of the interface. When $\rho=0.002$, increasing $\omega(t)$ is accompanied by a decreasing interface force $f(\omega(t))$ which happens after the interface strength has been attained (i.e., on the descending branch of the interface-separation curve). This kind of behavior leads to creep driven relaxation of the interface force. Note that when $\rho=0.02$, both the interface separation $\omega(t)$ and the interface force $f[\omega(t)]$ decrease. In this case, the interface unloads under a relaxation of interface separation. For $\rho=0.01$, the interface separation $\omega(t)$ decreases accompanied by an increase in interface force $f[\omega(t)]$. This indicates that the increasing interface force occurs *after* the interface strength has been attained, on the descending branch of the interface-separation curve. This kind of self-repairing interface response is unrealistic and ultimately is a consequence of using a potential interface force separation law.

For $\rho=0.002$ and $\varepsilon_0=0.0054$, the evolution of mean stress, interface separation, and interface force are depicted in Fig. 6. As can be seen from the figure, they all experience a jump disconti-

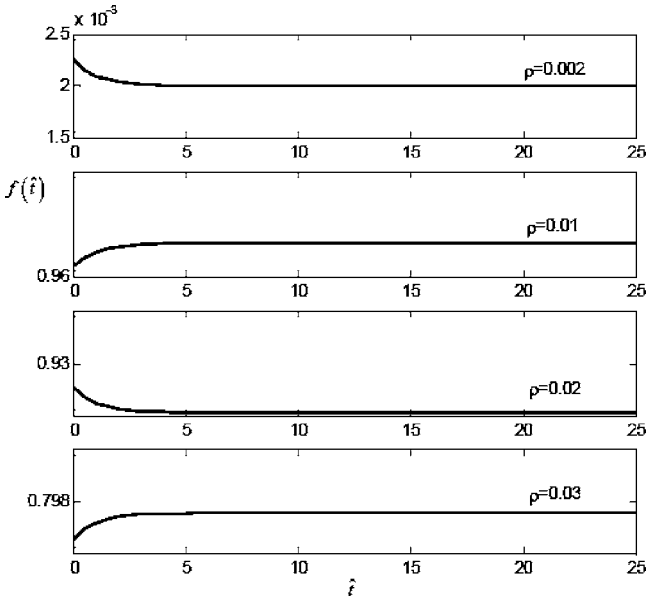


Fig. 5 Interface force under relaxation boundary condition; $\varepsilon_0=0.012$

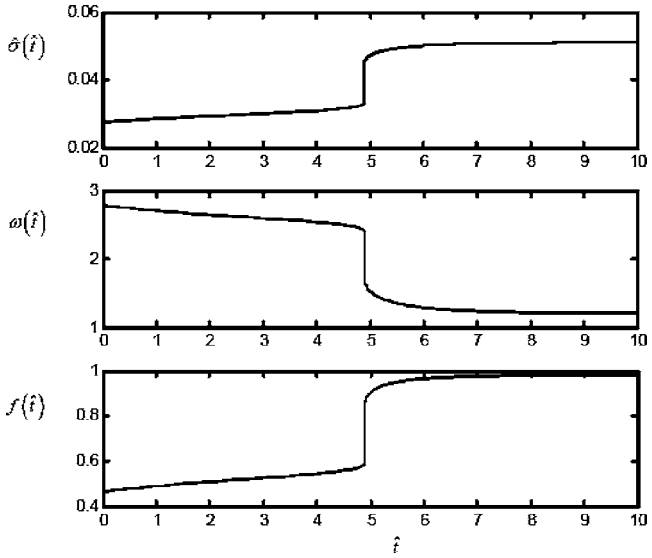


Fig. 6 Bifurcation under relaxation boundary condition; $\rho = 0.002$, $\varepsilon_0 = 0.0054$

nuity at a normalized time of approximately $\hat{t} \cong 4.9$. This jump corresponds to bifurcation of the interface and occurs because $\dot{\omega}$ ceases to exist at finite values of ω. The behavior is governed by elastic condition (10) or (23) required for bifurcation of initial states [see discussion following Eq. (10)]. For large ρ values, e.g., $\rho = 0.01$, 0.02, or 0.03, discontinuous behavior will not occur which is consistent with the results of Levy [4] and also with the results discussed in the previous section, i.e., that small ρ values and large $\sigma_{\max}$ values favor bifurcation.

Consider now the creep response characterized by an applied traction boundary condition with normalized applied stress $\hat{\sigma} = \sigma_0 / \mu_1^+ = 0.03$ (Assume that parameter values used for $c, \kappa_\mu, \kappa_k, E^+, \sigma_{\max}$ are unchanged.) Figure 7 depicts the strain response $\varepsilon(t)$ for varying characteristic length ratios ρ. Also included are curves depicting the limiting behavior of the rigid interface composite and the voided composite. As shown in the figure, the strain creeps for most of the cases. However, for extremely small ρ values, e.g., $\rho = 0.005$, with $\hat{\sigma} = 0.054$, the strain may relax.

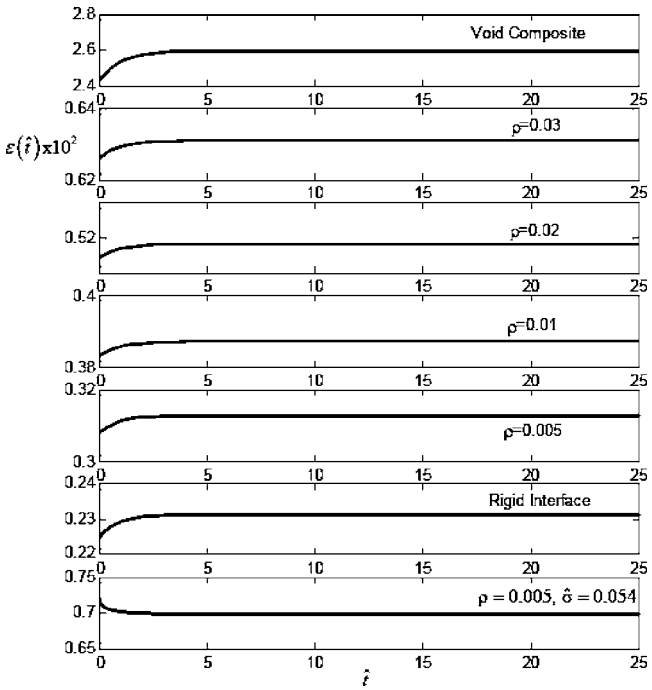


Fig. 7 Strain response under creep boundary condition; $\hat{\sigma} = 0.03$

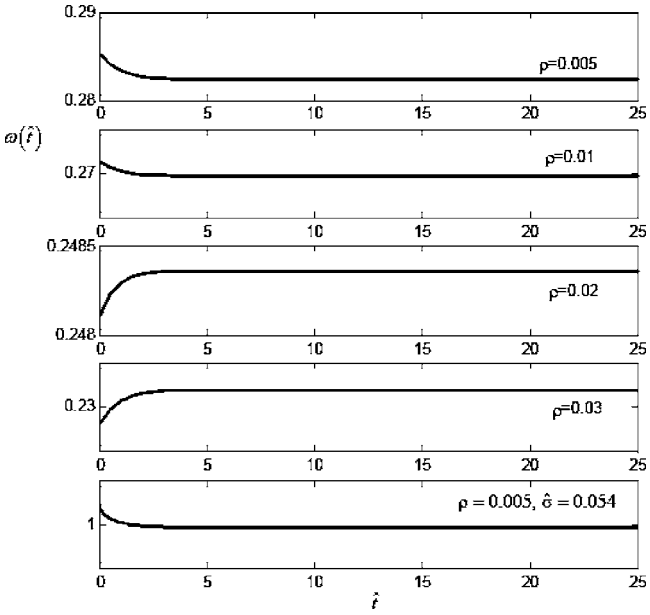


Fig. 8 Interface separation under creep boundary condition; $\hat{\sigma} = 0.03$

Figures 8 and 9 show the response of interface separation $\omega(t)$ and corresponding interface force $f[\omega(t)]$, respectively. For $\rho = 0.02$, 0.03, the interface separation $\omega(t)$ increases and the interface force increases as well. This behavior occurs on the ascending branch of the interface-separation curve, prior to the attainment of the interface strength and gives rise to creep driven loading. When $\rho = 0.005$, 0.01, and $\hat{\sigma} = 0.03$ (Fig. 8), the interface separation $\omega(t)$ decreases. Because this decreasing interface separation

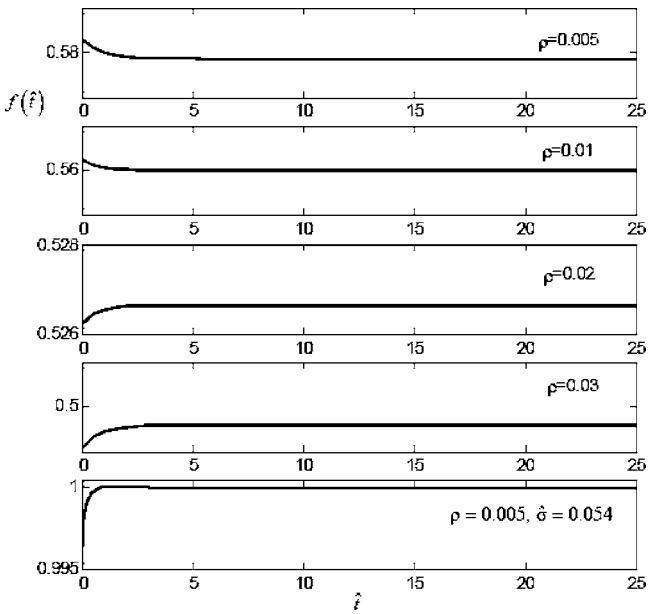


Fig. 9 Interface force under creep boundary condition; $\hat{\sigma} = 0.03$

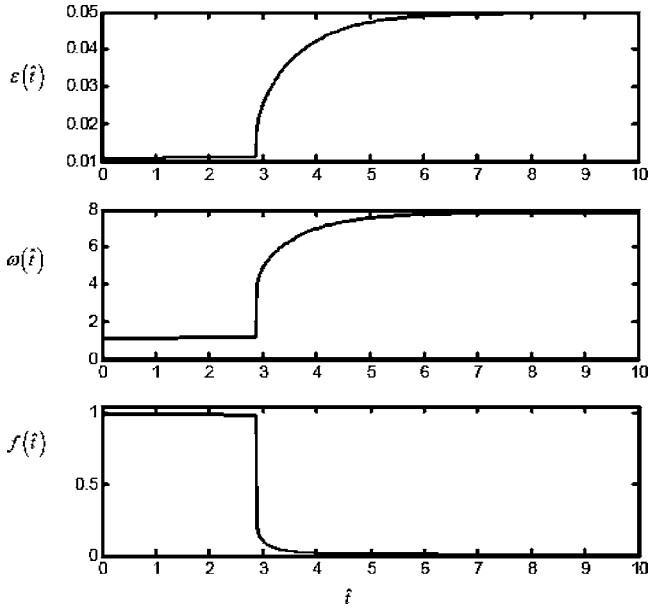


Fig. 10 Bifurcation under creep boundary condition; $\rho = 0.01, \hat{\sigma} = 0.058$

ration occurs on the ascending branch of the interface force-separation law, the interface force $f[\omega(t)]$ decreases as well. In this case, the interface unloads with relaxing interface separation. For $\rho = 0.005$ and $\hat{\sigma} = 0.054$, the interface separation $\omega(t)$ decreases on the descending branch of the interface force-separation law so that the interface force increases. This kind of behavior leads to self-repair of the interface of the kind noted previously.

Finally, for $\rho = 0.01$ and $\hat{\sigma} = 0.058$, the mean strain response, interface separation, and interface force are depicted in Fig. 10. The interface separation $\omega(t)$ increases but the corresponding interface force $f[\omega(t)]$ decreases. Thus, the behavior occurs after the attainment of the interface strength, on the descending branch of the interface-separation curve, and the interface force relaxes under creep. Furthermore, all of the curves experience a jump discontinuity at normalized time $\hat{t} \cong 2.9$. This jump corresponds to bifurcation of the interface and, as stated previously, occurs because $\dot{\omega}$ ceases to exist at finite values of ω. This behavior is governed by elastic condition (16) required for bifurcation of initial states. For larger ρ values, e.g., $\rho = 0.02$ or 0.03, bifurcation will not occur consistent with the conclusion that small values of ρ and large values of $\sigma_{\max}$, favor bifurcation.

5 Antiplane Shear

In what follows, the results obtained in Secs. 2 and 3 are used to obtain the time dependent effective antiplane shear response of an RVE composed of the composite cylinders microstructure. Interface slip response is assumed to be uniform at all values of axial coordinate for a fixed polar angle (Fig. 1). This is because, unlike transverse equibiaxial tension in which the separation is uniform around the interface, antiplane shear gives rise to interface coordinate (polar angle) dependent slip. Here a single-mode approximation for the slip is utilized as outlined in [6]. This assumption gives rise to governing equations of the form (1), provided we take f to be the projection of the normalized shear-slip interface force law g on the first mode, i.e.,

$$\tau_{\max} f(\omega) = \frac{\tau_{\max}}{\pi} \int_0^{2\pi} g(\omega, \theta) \cos\theta \, d\theta, \quad \frac{w}{\rho} = \omega \cos\theta. \tag{28}$$

For an applied strain boundary condition ($\varepsilon_0 = \gamma_0$) the coefficients g_{ij} in Eq. (1) are

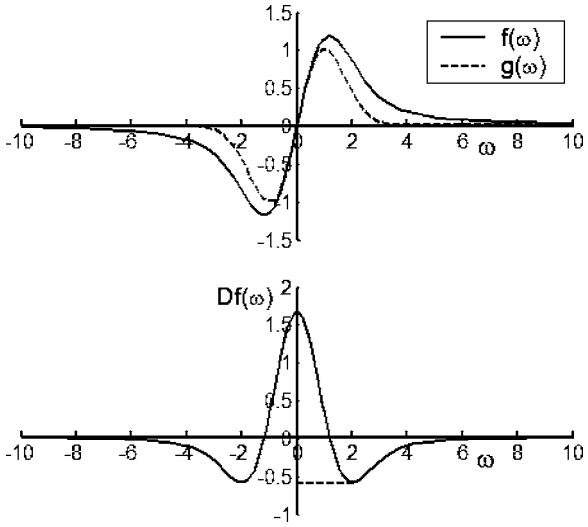


Fig. 11 The shear interface force law *g* and its first mode *f* and *Df*

$$g_{11} = \frac{1-c}{1+c}, \quad g_{12} = 0, \quad g_{13} = 0, \quad g_{14} = 0; \quad g_{15} = 0, \quad g_{16} = \frac{2c}{1+c},$$

$$g_{21} = -2, \quad g_{22} = 0; \quad g_{23} = \rho(1+c), \quad g_{24} = 0; \tag{29}$$

$$g_{25} = \frac{1+c}{\mu^-}, \quad g_{26} = 1-c,$$

and, when a traction boundary condition ($\sigma_0 = \tau_0$) is applied the coefficients q_{ij} are

$$q_{11} = \frac{1+c}{1-c}, \quad q_{12} = 0; \quad q_{13} = 0, \quad q_{14} = 0; \quad q_{15} = -\frac{2c}{1-c},$$

$$q_{16} = 0, \tag{30}$$

$$q_{21} = -2, \quad q_{22} = 0; \quad q_{23} = 0, \quad q_{24} = (1-c)\rho; \quad q_{25} = 1+c,$$

$$q_{26} = \frac{1-c}{\mu^-}; \quad q_{27} = 0, \quad q_{28} = 0.$$

In order to proceed further, assume that the interface is characterized by a simple exponential constitutive relation [8] that allows for interface failure in shear mode given by $g(w/\rho) = e^{1/2}(w/\rho)e^{-(w/\rho)^2/2}$. Then, Eq. (28) implies that

$$\tau_{\max} f(\omega) = \tau_{\max} e^{1/2} \omega e^{-\omega^2/4} \left[I_0\left(\frac{\omega^2}{4}\right) - I_1\left(\frac{\omega^2}{4}\right) \right],$$

$$Df(\omega) = \left(\frac{1}{\omega} - \omega\right) f(\omega) + 2e^{(1/2 - 1/4\omega^2)} I_1\left(\frac{1}{4}\omega^2\right), \tag{31}$$

where I_0, I_1 are modified Bessel functions of the first kind of order zero and one, respectively. The graphs of g, f, and Df are shown in Fig. 11. From Eqs. (10), (26), (29), and (4), and Fig. 11 it follows that a necessary condition for relaxation bifurcation is

$$\frac{\rho(1+c)(1+\kappa_\mu)}{\kappa_\tau[(1-c)+(1+c)(1+\kappa_\mu)/\kappa_f]} < 0.5897, \tag{32}$$

where $\kappa_f = \mu^-/\mu_1^+$, $\kappa_\tau = \tau_{\max}/\mu_1^+$. It is apparent from Eq. (32) that small values of ρ and large values of κ_τ (or $\tau_{\max}$) favor bifurcation.

In what follows, the creep response to an applied boundary traction is discussed. Relaxation response will be discussed in the

following section in the context of pure torsion of a circular cylinder. For the creep boundary condition, a necessary condition for bifurcation is

$$\frac{\rho(1-c)(1+\kappa_\mu)}{\kappa_\tau[(1+c)+(1-c)(1+\kappa_\mu)/\kappa_f]} < 0.5897. \tag{33}$$

A comparison of Eqs. (32) and (33) indicates that, for the same fiber composite, bifurcation under a creep boundary condition occurs more readily than under a relaxation boundary condition.

5.1 The Voided Composite. The creep compliance of the voided composite can be obtained by direct integration of Eq. (14). The result is

$$\rho\omega(\hat{t}) = \rho\omega_0[1+\kappa_\mu(1-e^{-p\hat{t}})],$$

$$\gamma(\hat{t}) = J^*(\hat{t})\tau_0 \tag{34}$$

$$=[1+\kappa_\mu(1-e^{-p\hat{t}})]\gamma_0 = \frac{1+c}{1-c}J^+(\hat{t})\tau_0,$$

where

$$\rho\omega_0 = \frac{2\hat{\tau}}{(1-c)(1+\kappa_\mu)}, \quad \gamma_0 = \frac{1+c}{(1-c)(1+\kappa_\mu)}\hat{\tau}, \quad \hat{\tau} = \frac{\tau_0}{\mu_1^+},$$

$$p = \frac{1}{1+\kappa_\mu} = \frac{\lambda_2'}{\lambda_2}, \tag{35}$$

and $J^*(t)$ is the effective shear creep compliance of the fiber-reinforced composite. It is clear that both the shear strain and the interface slip, creep under a traction boundary condition. By contrast, for the voided composite under a relaxation boundary condition, the shear stress relaxes and interface slip remains constant.

5.2 The Rigid Interface Composite. The shear creep compliance for the rigid interface composite follows by integrating (14) and is given by $\gamma(\hat{t})=J^*(\hat{t})\tau_0=[1+\kappa_\mu(1-e^{-p\hat{t}})]\gamma_0=(1-c)J^+(\hat{t})\tau_0/(1+c)$, where $\gamma_0=(1-c)/[(1+c)(1+\kappa_\mu)]$ is the initial strain. This result is consistent with the results obtained in [3]. Note that for relaxation response, it can be shown that the stress $\tau(\hat{t})$ is proportional to the product of the applied strain $\gamma_0, \mu^+(\hat{t})$ and the term $(1+c)/(1-c)$.

5.3 The Nonlinear Interface Composite. The response of the nonlinear interface composite can be obtained by numerical solution of Eqs. (7) (relaxation) or (14) (creep). Detailed numerical results are not presented here since they are qualitatively identical to the transverse equibiaxial tension case. The basic phenomena that occur for equibiaxial tension also occur for antiplane shear, i.e., creep driven loading of the interface, creep driven relaxation of the interface, interface self-repair, bifurcation, etc. In Sec. 6, antiplane shear relaxation will be considered from the standpoint of torsion of circular rods.

6 Pure Torsion of a Viscoelastic Composite Cylinder

For pure torsion of a solid or annular circular cylinder composed of viscoelastic material modeled by constitutive relation (7), the relevant deformation mode is antiplane shear. Because the geometry and loading are rotationally symmetric about the longitudinal axis, cross sections will rotate about the cylinder axis but will not warp. In the following subsection, the response of an elastic composite cylinder under an applied rate of twist is briefly summarized. For a detailed treatment of this problem see [14]. Note that when specific results are desired the form of the interface force-slip relation employed is given by Eq. (31).

6.1 Elastic Cylinder. Let the angle of twist and the rate of twist at axial coordinate z be $\phi(z)$ and $\alpha=d\phi/dz$ respectively, where α is uniform. The shear strain γ (measured in radians) is $\gamma=r\alpha$. The stress tensor field is $\mathbf{S}=\tau(\mathbf{e}_\theta\otimes\mathbf{e}_z+\mathbf{e}_z\otimes\mathbf{e}_\theta)$ and, from Eqs. (1) and (29), τ is governed by

$$\hat{\tau} = \frac{1-c}{1+c}\alpha r + \frac{2c}{1+c}\kappa_\tau f(\omega),$$

$$0 = (1+c)\rho\omega + (1-c)\kappa_\tau f(\omega) - 2\alpha r, \tag{36}$$

where $\hat{\tau}=\tau/\mu^+$ and $\kappa_\tau=\tau_{\max}/\mu^+$ [7]. This stress field satisfies the equilibrium equation $\operatorname{div}\mathbf{S}=\mathbf{0}$ trivially. Note that the torque, which is the moment resultant of $\hat{\tau}$, is obtained in the usual way by integration over the cross section.

If α is regarded as the bifurcation parameter and a location in the cross section is fixed then a necessary condition for bifurcation follows from Eq. (36_2) and is

$$\rho(1+c)+(1-c)\kappa_\tau Df(\omega)=0 \text{ or } Df(\omega) = -\frac{\rho(1+c)}{\kappa_\tau(1-c)}. \tag{37}$$

Bifurcation points $(\bar{\omega},\bar{\alpha})$ are solutions to Eqs. (36) and (37). Because α and r play an equivalent role in Eq. (36), a jump in the slip field may occur across a singular surface $r=r^*$. When this occurs, a jump in the stress field across the singular surface $r=r^*$ is expected as well. The jump condition $[\mathbf{s}(\mathbf{n})]=\mathbf{0}$, where $\mathbf{s}(\mathbf{n})=\mathbf{S}\mathbf{n}$ is the traction vector on the singular surface with normal $\mathbf{n}=\mathbf{e}_r$, is trivially satisfied for shear stress field $\mathbf{S}=\tau(\mathbf{e}_\theta\otimes\mathbf{e}_z+\mathbf{e}_z\otimes\mathbf{e}_\theta)$. The equation governing the existence of singular surfaces is simply Eq. (37) and, from Fig. 11, a singular surface will exist when

$$\frac{\rho(1+c)}{\kappa_\tau(1-c)} < 0.5897, \tag{38}$$

i.e., there are two distinct solutions to Eqs. (37) (for positive ω). Denote these two solutions by $\bar{\omega}_1,\bar{\omega}_2$ with $\bar{\omega}_1<\bar{\omega}_2$. When Eq. (38) is not satisfied, (37) will have no solutions and a singular surface will not exist. It is apparent from Eq. (38) that small values of ρ and large values of κ_τ (or $\tau_{\max}$) favor the existence of a singular surface.

Denote the two values of ω at the singular surface, and the jump across the surface, as follows, $\omega^-=\operatorname{Lim}\omega(r), r\uparrow r^*$, $\omega^+=\operatorname{Lim}\omega(r), r\downarrow r^*$, and $[\omega]=\omega^+-\omega^-$. Note that we are choosing $\bar{\omega}_1=\omega^-$ as the lower value of the jump, as opposed to choosing $\bar{\omega}_2=\omega^+$ as the higher value of the jump, since $\bar{\omega}_2=\omega^+$ cannot be attained by a continuous increase in strain γ (or radius $r=\gamma/\alpha$) from the reference state (see Fig. 12).

At some value of rate of twist α, assume there is a singular surface at $r=r^*$. To get r^*, first solve Eq. (37) and denote the smaller of the two solutions by $\bar{\omega}_1=\omega^-$. r^*, then follows from Eq. (36_2)

$$0=(1+c)\rho\omega^- + (1-c)\kappa_\tau f(\omega^-) - 2\alpha r^* \text{ or } r^* = \frac{\psi}{\alpha}, \tag{39}$$

where $2\psi=(1+c)\rho\omega^-+(1-c)\kappa_\tau f(\omega^-)$ is constant, fixed by the concentration and the constituent properties but independent of twist rate α. The jump in the interface slip field $[\omega]$ across r^* is obtained after solving Eq. (39) for the second solution $\omega^+(>\omega^-)$.

From Eq. (39), the critical value of angle of twist α^* for the singular surface to initiate at the outer cylindrical surface of the cylinder, i.e., $r^*=b$, is $\alpha^*=\psi/b$. The ultimate value of angle of twist α_u is that value of α for which the singular surface has traversed the entire section. It is obtained, for an annular cylinder of inner radius a, by letting $r^*=a$, or, $\alpha_u=\psi/a$. Note this implies that, for the solid cylinder, an infinite rate of twist α_u is required

[7] Note that for viscoelastic composite response, $\hat{\tau}$ is taken to be τ/μ_1^+.

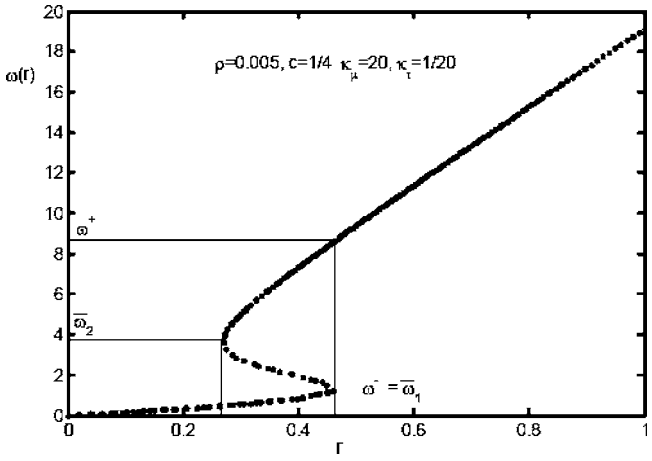


Fig. 12 Notation

for the singular surface to traverse the entire section. Therefore, for the solid cylinder, there will always exist a cylindrical core $r \in (0, r^*)$ of minimally slipped material where r^* is given by Eq. (39).

The velocity of propagation of the singular surface can be obtained by differentiating Eq. (39) with respect to the rate of twist α, i.e., $v^* = dr^*/d\alpha = -\psi/\alpha^2$. Note that both jumps in interface slip and shear stress occur across the same surface $r = r^*$ and that $[\omega] > 0$ while $[\tau] < 0$. The stress values $\tau^- = \text{Lim}\, \tau(r), r \uparrow r^*$, $\tau^+ = \text{Lim}\, \tau(r), r \downarrow r^*$ at a singular surface can be obtained by substituting r^* and $\omega^-(\omega^+)$ into Eq. (36_1), and after rearrangement $\tau_\mu^\pm = (1-c)\rho\omega^\pm/2 + (1+c)\kappa_\tau f(\omega^\pm)/2$, which are constant for fixed values of concentration and constituent properties. The jump in stress is $[\tau] = \tau^+ - \tau^-$.

6.2 Viscoelastic Cylinder. The solution to the elastic torsion problem under a uniform rate of twist α_0 gives the shear stress τ_0, shear strain γ_0, and interface slip magnitude ω_0 at any point throughout the cross section. Now, if the matrix is viscoelastic and the composite constitutive relation is given by Eq. (7), the relaxation response of the circular cylinder can be obtained by holding it under a fixed rate of twist α_0. When this is true, the shear strain will be constant (time independent) and given by $\gamma_0 = r\alpha_0$.

For this problem, the redistribution of shear stress and interface slip magnitude is sought throughout the cross section. Furthermore, if an elastic singular surface forms on initial application of the twist angle, its movement during relaxation can be readily investigated. The analysis is carried out by numerically integrating the first-order differential equations (7) using initial values from governing equation (6). Here, it is assumed that the geometrical and constitutive parameters have the following values: $\kappa_\tau = 1/18$, $\kappa_\mu = 1/9$, $c = 1/4$, and $\alpha_0 = 0.06$.

For $\rho = 0.05$, Fig. 13 indicates that the shear stress and slip magnitude relax for the entire cross section. The distributions of initial and long-term ($\hat{t} = 25$) values of the shear stress and slip magnitude are shown in the figure by dashed and solid lines, respectively. When $\rho = 0.02$, the relaxation curves of shear stress and slip magnitude are shown in Fig. 14. Note that the shear stress creeps within the annular region $r_1 < r < r_2$ (the shear stress does not change at $r = r_{1,2}$). Within this region, $\omega > 1$ so values of interface force are on the descending branch of the interface force slip curve (Fig. 11). Because the interface slip ω relaxes, the interface force $\tau_{\max} g(\omega)$ must increase giving rise to self-repairing interface behavior.

With further decrease in $\rho(\rho = 0.005)$, a singular surface will exist initially $(t = 0)$ at r_1^* as shown by the dashed lines in Fig. 15. (The distributions of shear stress and slip magnitude after relaxation are shown by solid lines). The figure indicates that the singular surface moves outward to r_2^*. This suggests that the initially slipped region $[r_1^*, 1]$ undergoes partial self-repair through relaxation, i.e., $[r_1^*, r_2^*]$ becomes only minimally slipped after relaxation.

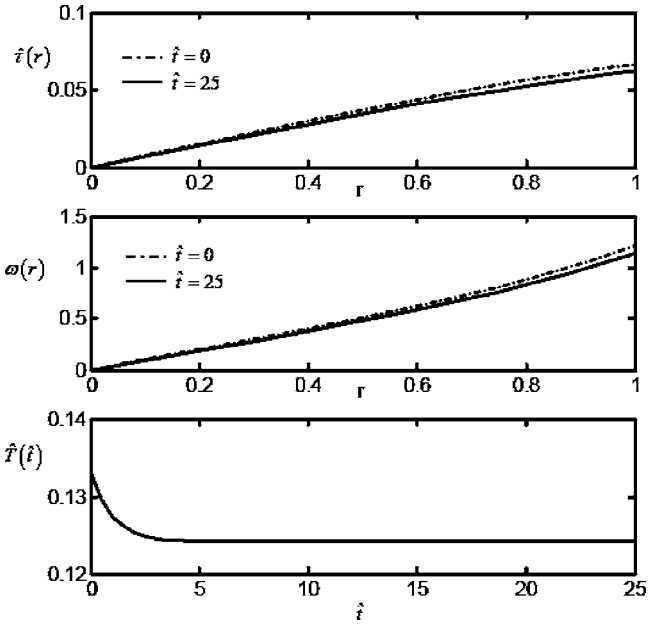


Fig. 13 Stress relaxation of a viscoelastic circular cylinder $\rho = 0.05$, $\hat{T}(\hat{t}) = T(\hat{t})/\mu_1^+$

Time dependent torque T, corresponding to a given value of the twist rate α_0, is the moment resultant of the shear stresses acting on the cross section

$$T(\hat{t}) = \int_a^b 2\pi\tau(r,\hat{t}) r^2 dr, \tag{40}$$

which is the torsion formula in implicit form. Clearly, $T(0) = T_0$ is the initially applied torque required to produce the rate of twist α_0

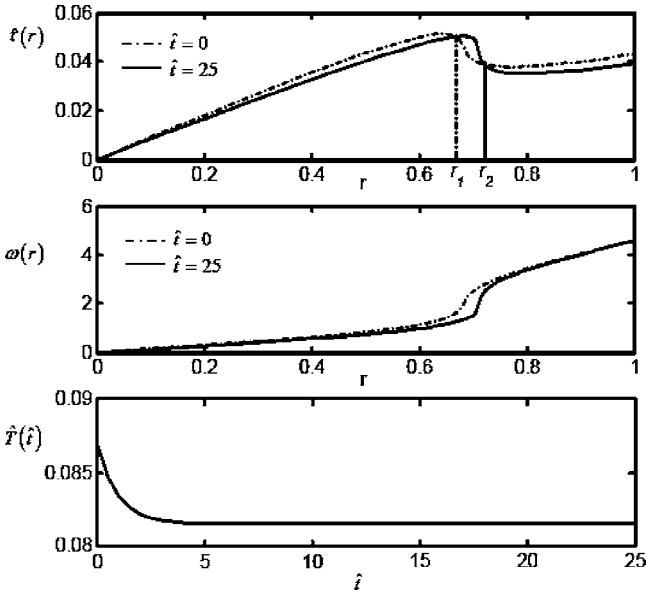


Fig. 14 Stress relaxation of a viscoelastic circular cylinder $\rho = 0.02$, $\hat{T}(\hat{t}) = T(\hat{t})/\mu_1^+$

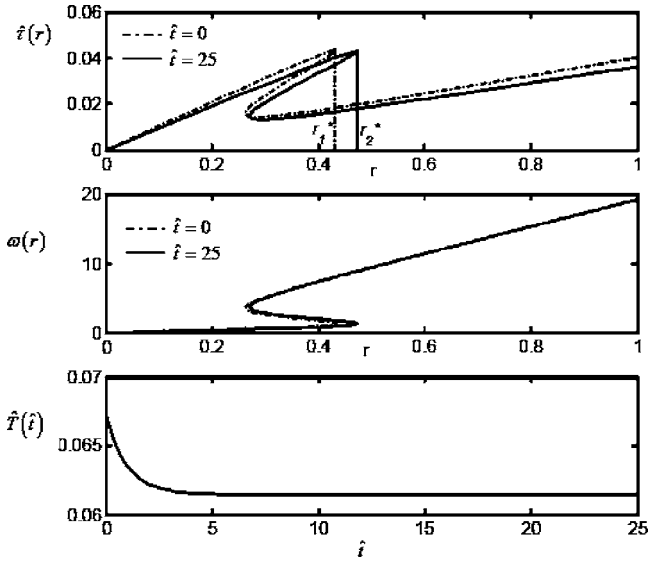


Fig. 15 Stress relaxation of a viscoelastic circular cylinder $\rho = 0.005$, $\hat{T}(\hat{t}) = T(\hat{t})/\mu_1^+$

when the matrix shear modulus is $\mu^+(0) = \mu_1^+ + \mu_2^+$. Furthermore, if $\tau(r,\hat{t})$ is continuous

$$\lim_{t\to\infty} T(\hat{t}) = \lim_{t\to\infty} \int_a^b 2\pi\tau(r,\hat{t})r^2 dr = \int_a^b 2\pi[\lim_{t\to\infty} \tau(r,\hat{t})]r^2 dr$$
$$= \int_a^b 2\pi\tau_e(r)r^2 dr = T_e, \tag{41}$$

where $\tau_e(r) = \lim_{t\to\infty} \tau(r,\hat{t})$ is the long-term (equilibrium) relaxation shear stress (or equilibrium stress). Because $\tau_e(r)$ is also the elastic response with shear modulus $\mu^+ = \mu_1^+, T_e$ is actually the torque for an elastic circular cylinder with matrix shear modulus $\mu = \mu_1^+$ under rate of twist α_0. If $\tau(r,\hat{t})$ is not continuous, the integration of Eq. (41) can be carried out by integrating $\tau(r,\hat{t})$ in intervals of (a,b) where it is continuous. The relaxation of the torque is shown in the bottom of Figs. 13–15 for different values of ρ. The figures indicate that the amount of relaxation (the difference between T_0 and T_e) decreases with decreasing ρ.

7 Conclusions

This paper has been concerned with the mechanical response of viscoelastic unidirectional fiber composites whose interfaces can separate/slip nonlinearly. The focus has been primarily on relaxation and creep response although more complex stress and strain histories can be treated within the framework presented. For a matrix which is a generalized *n*th-order Maxwell or Kelvin model, the development leads to two coupled *n*th-order differential equations. The form of these equations falls within the structure of continuum damage mechanics, i.e., an equation involving the stress, strain, "damage" variable and their rates, and an equation describing the evolution of the damage with stress or strain. For equibiaxial tension (or axial tension), the damage variable is the mean interface separation at the fiber-matrix interfaces, while for antiplane shear the damage variable is the first-mode multiplier in the expansion for the fiber-matrix interfacial slip, i.e., Eq. (28_2).

It has been demonstrated that for both the equibiaxial load and antiplane shear load of the voided or rigid interface composite, the mean stress relaxes under relaxation boundary conditions and the strain creeps under creep boundary conditions. Also, for equibiaxial load of the voided composite, the interface separation creeps under both of these two boundary conditions, while for antiplane shear the interface separation creeps for an applied stress boundary condition but is constant for an applied strain boundary condition. In contrast, the effective response of the nonlinear interface composite depends strongly on the particular values of interface force length parameter (ρ) chosen. At fixed interface strength, ρ is proportional to the interface energy. Small ρ values, corresponding to small interface energy, favor bifurcation. For larger ρ values, the stress will normally relax under relaxation boundary conditions while the strain will creep under creep boundary conditions. The interface separation however, shows relaxation or creep behavior depending on ρ and the initial elastic response. When the interface separation relaxes from the descending branch of the force-separation curve, the decreasing values of interface separation necessitate an increase in interface force. Self-repairing interface behavior of this kind is considered unrealistic. Interface constitutive relations that eliminate this defect typically have an unloading branch for $\omega < \omega_{\max}$ or $\dot{\omega} < 0$, where $\omega_{\max}$ is the maximum value of separation attained [15]. The kind of behavior just described occurs in both the transverse equibiaxial tension case as well as in the antiplane shear case.

An application of the antiplane shear theory has been utilized to study stress relaxation in the problem of pure torsion of a solid or annular circular cylinder. In the elastic problem, singular surfaces have been shown to exist if ρ is small enough. If such a surface forms on initial application of the load, then its time evolution results in interface self-repair under relaxation, which again underscores the difficulty in using a potential interface-force law. Another difficulty is utilizing effective property constitutive relations to study critical phenomena. The *classical approximation*, which allows for the use of effective property constitutive relations (derived for statistically uniform stress and strain fields) in situations where the fields are nonuniform, breaks down owing to the high stress gradients associated with singular surfaces. This issue, in the context of the elastic torsion problem, has been commented on previously [14].

Four of the five response modes of nonlinear interface viscoelastic unidirectional fiber composites have been considered in this paper, i.e., transverse bulk response, antiplane shear response, and axial tension response (and associated Poisson contraction). The remaining response mode, transverse or in-plane shear is the most difficult. The composite cylinders microstructure loses most of its utility here since only bounds on the true response can be obtained. Other approaches, employing geometric composite models, may yield useful results. An analysis of transverse shear of an elastic fiber composite with nonlinear interface, employing the Mori-Tanaka estimate [16], has been carried out in [17].

Appendix

Integral Eqs. (3) can be reduced to a set of two autonomous nonlinear *n*th-order differential equations for the stress $\sigma(t)$ and interface displacement jump $\omega(t)$ as follows. Consider the integrals on the right-hand side of Eq. (3). With the aid of Eq. (5_2), this integral and its *k*th-order time derivative can be written in the form, for $j=1$ or 2,

$$\int_0^t \dot{\mu}^+(t-T)K_j(T)dT = -\sum_{i=1}^n \lambda_i \int_0^t \mu_i(t-T)K_j(T)dT, \tag{A1}$$

$$\frac{d^k}{dt^k}\int_0^t \dot{\mu}^+(t-T)K_j(T)dT = \sum_{i=1}^{n} \mu_i^+[-\lambda_i K_j^{(k-1)}(t) + \lambda_i^2 K_j^{(k-2)}(t) + \ldots + (-1)^k(\lambda_i)^k K_j(t)] + \sum_{i=1}^{n} (-1)^{k+1}(\lambda_i)^{k+1} \times \int_0^t \mu_i(t-T)K_j(T)dT.$$

Then using Eq. (A1), the kth derivative of Eq. (3) is

$$\sigma^{(k)}(t) = \sum_{i=1}^{n} \mu_i^+[-\lambda_i K_1^{(k-1)}(t) + \lambda_i^2 K_1^{(k-2)}(t) + \ldots + (-1)^k(\lambda_i)^k K_1(t)] + \sum_{i=1}^{n} (-1)^{k+1}(\lambda_i)^{k+1}\int_0^t \mu_i(t-T)K_1(T)dT + A_1\varepsilon^{(k)}(t) + A_2\omega^{(k)}(t) + A_3\sigma_{\max}f^{(k)}(\omega(t)), \tag{A2}$$

$$0 = \sum_{i=1}^{n} \mu_i^+[-\lambda_i K_2^{(k-1)}(t) + \lambda_i^2 K_2^{(k-2)}(t) + \ldots + (-1)^k(\lambda_i)^k K_2(t)] + \sum_{i=1}^{n} (-1)^{k+1}(\lambda_i)^{k+1}\int_0^t \mu_i(t-T)K_2(T)dT + B_1\varepsilon^{(k)}(t) + B_2\omega^{(k)}(t) + B_3\sigma_{\max}f^{(k)}(\omega(t)).$$

When $k=1\ldots n$, we have two systems of n equations of the form (A2), i.e., coupled algebraic equations which can be used to solve for the $2n$ integrations of the form,

$$\int_0^t \mu_i(t-T)K_j(T)dT, \quad i=1\ldots n \text{ for } j=1,2 \tag{A3}$$

By Eq. (A2), integrations of the form (A3), can be solved for in terms of higher-order derivatives $\sigma^{(k)}(t), \omega^{(k)}(t), \varepsilon^{(k)}(t), f^{(k)} \times(\omega(t)), k=1\ldots n$ and eliminated in Eq. (3). The result is a system consisting of two nth-order differential equations for $\sigma(t), \omega(t)$ of the form

$$G_1[\sigma^{(k)}(t), \omega^{(k)}(t), \varepsilon^{(k)}(t), f^{(k)}(\omega(t)), c, \mu_k, \lambda_k] = 0,$$
$$G_2[\sigma^{(k)}(t), \omega^{(k)}(t), \varepsilon^{(k)}(t), f^{(k)}(\omega(t)), c, \mu_k, \lambda_k] = 0, \tag{A4}$$

where these equations include derivatives of all orders, $k=0\ldots n$, which occur in Eq. (A4) in linear terms. (Note that $\sigma^{(k)}$ is the kth derivative of σ and $\sigma^{(0)}=\sigma$.)

References

[1] Needleman, A., 1987, "A Continuum Model for Void Nucleation by Inclusion Debonding," ASME J. Appl. Mech., **54**, pp. 525–531.
[2] Hashin, Z., 1965, "Viscoelastic Behavior of Heterogeneous Media," ASME J. Appl. Mech., **32**, pp. 630–636.
[3] Hashin, Z., 1966, "Viscoelastic Fiber Reinforced Materials," AIAA J., **4**(8), pp. 1411–1417.
[4] Levy, A. J., 1996, "The Effective Dilatational Response of Fiber Reinforced Composites with Nonlinear Interface," ASME J. Appl. Mech., **63**, pp. 360–364.
[5] Levy, A. J., 2000a, "The Fiber Composite with Nonlinear Interface—Part I: Axial Tension," ASME J. Appl. Mech., **67**, pp. 727–732.
[6] Levy, A. J., 2000b, "The Fiber Composite with Nonlinear Interface—Part II: Antiplane Shear," ASME J. Appl. Mech., **67**, pp. 733–739.
[7] Hashin, Z., and Rosen, B. W., 1964, "The Elastic Moduli of Fiber Reinforced Materials," ASME J. Appl. Mech., **31**, pp. 223–232.
[8] Needleman, A., 1992, "Micromechanical Modelling of Interfacial Decohesion," Ultramicroscopy, **40**, pp. 203–214.
[9] Sneddon, 1972, *The Use of Integral Transforms*, McGraw-Hill, New York.
[10] Christensen, R. M., 1982, *Theory of Viscoelasticity*, Academic Press, London.
[11] Ferrante, J., Smith, J. R., and Rose, J. H., 1982, "Universal Binding Energy Relations in Metallic Adhesion," *Microscopic Aspects of Adhesion and Lubrication*, Elsevier, Amsterdam, pp. 19–30.
[12] Levy, A. J., 1998, "The Effect of Interfacial Shear on Cavity Formation at an Elastic Inhomogeneity," J. Elast., **50**, pp. 49–85.
[13] Corless, R. M., Gonnet, G. H., Hare, D. E. G., and Jeffrey, D. J., 1993, "Lambert's *W* Function in Maple," Maple Tech. Newsl., **3**, pp. 12–22.
[14] Xie, M., and Levy, A. J., 2004, "On the Torsion of a Class of Nonlinear Fiber Composite Cylinders," J. Elast., **75**, pp. 31–48.
[15] Tvergaard, V., 1990, "Effect of Fiber Debonding in a Whisker-Reinforced Metal," Mater. Sci. Eng., A, **125**, pp. 203–213.
[16] Mori, T., and Tanaka, K., 1973, "Average Stress in Matrix and Average elastic Energy of Materials with Misfitting Inclusions," Acta Metall., **21**, pp. 601–604.
[17] Dong, Z., and Levy, A. J., 2000, "Mean Field Estimates of the Response of Fiber Composites with Nonlinear Interface," Mech. Mater., **32**, pp. 739–767.

Closed-Form Solution for an Antiplane Interface Crack between Two Dissimilar Magnetoelectroelastic Layers

Bao-Lin Wang
e-mail: wangbl2001@hotmail.com

Yiu-Wing Mai
Fellow ASME
e-mail: mai@aeromech.usyd.edu.au

Centre for Advanced Materials Technology (CAMT),
School of Aerospace, Mechanical and Mechatronic Engineering,
Mechanical Engineering Building J07,
The University of Sydney,
Sydney, NSW 2006, Australia

In recent years, ferroelectromagnetic laminated composites have aroused worldwide research interest. In this paper, we developed a closed-form solution for antiplane mechanical and in-plane electric and magnetic fields for a crack between two dissimilar magnetoelectroelastic layers of finite thickness. Explicit expressions for stresses, electric and magnetic fields, together with their intensity factors are obtained for the two extreme cases of an impermeable and a permeable crack. Solutions for some special cases, such as a homogeneous magnetoelectroelastic layer, two magnetoelectroelastic layers with opposite poling directions, and a piezoelectric layer bonded to a piezomagnetic layer are also obtained. Explicit relations between the field intensity factors and the energy release rates for the interface crack are provided. [DOI: 10.1115/1.2083827]

1 Introduction

The effects of electromagnetomechanical coupling have been observed in single-phase materials in which simultaneous electric and magnetic ordering coexists, and in two-phase composites where the participating phases are ferroelectric and ferromagnetic [1–3]. Magnetoelectric composites with a surprisingly large magnetoelectric effect have been made from ferroelectric phases (e.g., $BaTiO_3$) and ferromagnetic phases (e.g., $CoFe_2O_4$). Composites made of ferroelectric/ferromagnetic materials can exhibit a significant magnetoelectric (ME) effect that is not present in single-phase ferroelectric or ferromagnetic materials [4,5]. Laminated composites, such as PZT-ferrite and PZT-Terfernol-D layered structures, have also been shown to be superior to single-phase ferroelectromagnetic materials in that the ME voltage coefficient can reach one or two orders of magnitude higher than that of the single phase.

Magnetoelectroelastic composites can be developed in the form of secondary-phase piezoelectric (or piezomagnetic) inclusions embedded in a piezomagnetic (or piezoelectric) matrix, or in the form of a composite laminate by alternating the ferromagnetic layers and ferroelectric layers during stacking. An area of increasing interest is the fracture mechanics of ferroelectromagnetic composites, which are combinations of the ferromagnetic and ferroelectric phases. This is because: (a) existing cracks may alter the overall effective properties (especially the ME effect) of the materials considerably, and (b) both ferroelectric phase and ferromagnetic phase in ferroelectromagnetic composites are usually brittle and prone to cracking. Unlike in piezoelectricity, limited studies have been focused on the fracture characteristics of ferroelectromagnetic materials. In [6–18], crack problems were investigated in order to obtain the crack-tip stress, electric, and magnetic field intensity factors.

Ferroelectromagnetic laminated composites have been the subject of much research interest in recent years. Exceptionally high ME effects are found in these composite laminates made of piezoelectric and magnetostrictive materials [19,20]. However, there are very few published theoretical results for multilayer magnetoelectric composites [21,22]. Generally, the ME voltage coefficient measured is one order of magnitude lower than that expected from theoretical calculations. The mechanical coupling between the piezoelectric layer and the magnetostrictive layer plays a key role in generating the ME effect in ferroelectromagnetic laminated composites. Therefore, any factors which weaken the mechanical coupling between the interface will naturally reduce the ME effect. One critical factor that can cause poor mechanical coupling between the piezoelectric and magnetostrictive layers is the presence of cracks at the interface, which will alter the integrity of the structure, and hence weaken the mechanical coupling effect.

In this paper, a closed-form solution for a through mode-III crack situated at the interface between two dissimilar magnetoelectroelastic layers is presented. The paper is organized as follows: Sec. 2 gives basic governing equations for the linear magnetoelectroelastic solids, and the solution for the problem of antiplane mechanical and in-plane electromagnetic fields. This solution is expressed in Fourier transform with some unknown variables. These unknown variables are determined in closed-form in Sec. 3 for prescribed stress, electric displacement, and magnetic induction boundary conditions on the medium surfaces, and in Sec. 4 for prescribed displacement, electric potential, and magnetic potential boundary conditions on the medium surfaces. In each of Secs. 3 and 4, solutions for electrically impermeable and magnetically impermeable crack (Subsections 3.1 and 4.1), electrically permeable and magnetically impermeable crack (Subsections 3.2 and 4.2), electrically impermeable and magnetically permeable crack (Subsections 3.3 and 4.3), electrically permeable and magnetically permeable crack (Subsections 3.4 and 4.4), are given. A crack in two magnetoelectroelastic layers with identical properties are given in Subsections 3.5 and 4.5, where the poling direction of the layers are either identical or opposite. Closed-form solution for a crack between a piezoelectric layer and a piezomagnetic layer are given in Subsections 3.6 and 4.6. Also provided are solutions for cases of infinite layer thickness of the layered medium, and semi-infinite crack (Subsections 3.7 and 4.7). Conclusions are finally given in Sec. 5.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received , March 21, 2005; final manuscript received July 17, 2005. Review conducted by 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 of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

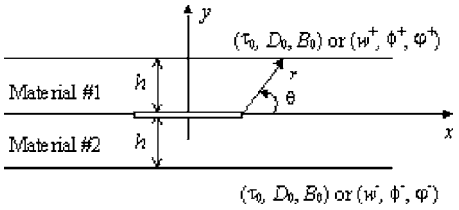


Fig. 1 An interface crack in a magnetoelectroelastic layered medium

2 Description of the Problem

Consider a magnetoelectroelastic layered medium occupying $-h \leqslant y \leqslant h$, $-\infty < x < \infty$, $-\infty < z < \infty$, with a through Griffith mode-III crack of length $2a$ located in the midplane, $y=0$, $-a \leqslant x \leqslant a$, $-\infty < z < \infty$, as shown in Fig. 1. Here, Cartesian coordinates x, y, z are the principal axes of the material symmetry while the z axis is oriented in the poling direction of the magnetoelectroelastic layers. From the viewpoint of applications, antiplane crack problems often provide a useful analogue to the more interesting in-plane fracture problems. The antiplane governing equations for the magnetoelectroelastic medium whose poling direction is perpendicular to the x-y plane are [12]:

- **Constitutive equations**:

$$\sigma_{xz} = c_{44}\frac{\partial w}{\partial x} + e_{15}\frac{\partial \phi}{\partial x} + h_{15}\frac{\partial \varphi}{\partial x}, \quad \sigma_{yz} = c_{44}\frac{\partial w}{\partial y} + e_{15}\frac{\partial \phi}{\partial y} + h_{15}\frac{\partial \varphi}{\partial y}, \tag{1a}$$

$$D_x = e_{15}\frac{\partial w}{\partial x} - \in_{11}\frac{\partial \phi}{\partial x} - \beta_{11}\frac{\partial \varphi}{\partial x}, \quad D_y = e_{15}\frac{\partial w}{\partial y} - \in_{11}\frac{\partial \phi}{\partial y} - \beta_{11}\frac{\partial \varphi}{\partial y}, \tag{1b}$$

$$B_x = h_{15}\frac{\partial w}{\partial x} - \beta_{11}\frac{\partial \phi}{\partial x} - \gamma_{11}\frac{\partial \varphi}{\partial x}, \quad B_y = h_{15}\frac{\partial w}{\partial y} - \beta_{11}\frac{\partial \phi}{\partial y} - \gamma_{11}\frac{\partial \varphi}{\partial y}, \tag{1c}$$

  and

- **Equilibrium equations** (in the absence of body forces, concentrated electric charges, and concentrated magnetic source):

$$c_{44}\nabla^2 w + e_{15}\nabla^2 \phi + h_{15}\nabla^2 \varphi = 0, \tag{2a}$$

$$e_{15}\nabla^2 w - \in_{11}\nabla^2 \phi - \beta_{11}\nabla^2 \varphi = 0, \tag{2b}$$

$$h_{15}\nabla^2 w - \beta_{11}\nabla^2 \phi - \gamma_{11}\nabla^2 \varphi = 0. \tag{2c}$$

In Eqs. (1) and (2), w is antiplane mechanical deformation, ϕ and φ are electric potential and magnetic potential, respectively, σ_{ij}, D_i, and B_i are components of stress, electrical displacement, and magnetic induction, c_{44}, e_{15}, h_{15}, and β_{11} are elastic, piezoelectric, piezomagnetic, and electromagnetic constants, and $\in_{11}$ and γ_{11} are dielectric permeability and magnetic permeability, respectively.

Here, for convenience, we consider the following antiplane mechanical and in-plane electromagnetic loads applied on the surfaces of the layered medium:

$$\sigma_{yz}(x, \pm h) = \tau_0, \quad D_y(x, \pm h) = D_0, \quad B_y(x, \pm h) = B_0, \tag{3a}$$

or

$$w(x, \pm h) = w^{\pm}, \quad \phi(x, \pm h) = \phi^{\pm}, \quad \varphi(x, \pm h) = \varphi^{\pm}, \tag{3b}$$

in which the subscripts 0, +, and − represent the prescribed values. (τ_0, D_0, B_0), (w^+, ϕ^+, φ^+), and (w^-, ϕ^-, φ^-) are known constants. Boundary conditions (3a) indicate that the top and bottom surfaces of the medium, $y=\pm h$, are uniformly loaded by an antiplane mechanical stress, an in-plane electric displacement, and an in-plane magnetic induction. Equations (3b) indicate that the top and bottom surfaces of the layer, $y=\pm h$, are clamped and displaced along the z direction by an amount $w^+ + w^-$, and there is a constant electric potential difference $\phi^+ + \phi^-$ and a constant magnetic potential difference $\varphi^+ + \varphi^-$ between the top and bottom surfaces.

Hereafter, the subscript j will be used to denote the layer number $(j=1,2)$. Due to continuity conditions, there are additional elastic, electric, and magnetic constraints along the interface:

$$w_{(1)}(x,0) = w_{(2)}(x,0), \ \phi_{(1)}(x,0) = \phi_{(2)}(x,0), \ \varphi_{(1)}(x,0) = \varphi_{(2)}(x,0), \quad \|x\| \geqslant a. \tag{4}$$

In solving the crack problem, the crack surfaces are usually stress free. However, since the medium inside the crack (usually air or vacuum) allows some penetrations of the electric and magnetic fields, these fields may not be zero. Suppose the normal components of the electric displacement and the magnetic induction inside the crack are d_0 and b_0, respectively, then along the interface, we have:

$$\sigma_{yz(1)}(x,0) = \sigma_{yz(2)}(x,0) = 0, \ D_{y(1)}(x,0) = D_{y(2)}(x,0) = d_0, \ B_{y(1)}(x,0) = B_{y(2)}(x,0) = b_0, \quad \|x\| < a. \tag{5a}$$

$$\sigma_{yz(1)}(x,0) = \sigma_{yz(2)}(x,0), \quad D_{y(1)}(x,0) = D_{y(2)}(x,0), \quad B_{y(1)}(x,0) = B_{y(2)}(x,0), \quad \|x\| \geqslant 0. \tag{5b}$$

The quantities d_0 and b_0 are unknown and will be determined later.

3 Closed-Form Solution for Prescribed Stress, Electric Displacement, and Magnetic Induction Loads

In this section, the prescribed boundary conditions (3a) are considered. Due to symmetry, it suffices to analyze the right-half portion of the medium, i.e., $x>0$. The magnetoelectroelastic field in the $x<0$ portion can be directly given by symmetry consideration. Using the Fourier transform, an appropriate solution of Eq. (2), in connection with Eq. (1), can be expressed by the following integrals:

$$\begin{Bmatrix} w_{(1)}(x,y) \\ \phi_{(1)}(x,y) \\ \varphi_{(1)}(x,y) \end{Bmatrix} = \int_0^{\infty} \frac{\cosh(sh - sy)}{\cosh(sh)} \cos(sx) \begin{Bmatrix} A_{(1)}(s) \\ B_{(1)}(s) \\ C_{(1)}(s) \end{Bmatrix} ds + \begin{Bmatrix} w_{0(1)} \\ \phi_{0(1)} \\ \varphi_{0(1)} \end{Bmatrix} \frac{y}{h}, \quad 0 \leqslant y \leqslant h, \tag{6a}$$

$$\begin{Bmatrix} w_{(2)}(x,y) \\ \phi_{(2)}(x,y) \\ \varphi_{(2)}(x,y) \end{Bmatrix} = \int_0^{\infty} \frac{\cosh(sh + sy)}{\cosh(sh)} \cos(sx) \begin{Bmatrix} A_{(2)}(s) \\ B_{(2)}(s) \\ C_{(2)}(s) \end{Bmatrix} ds + \begin{Bmatrix} w_{0(2)} \\ \phi_{0(2)} \\ \varphi_{0(2)} \end{Bmatrix} \frac{y}{h}, \quad -h \leqslant y \leqslant 0, \tag{6b}$$

where $x \geqslant 0$, $(j=1,2)$ denotes the layer number, $A_{(j)}(s)$, $B_{(j)}(s)$, and $C_{(j)}(s)$ are unknown functions to be determined from the prescribed conditions (4) and (5) along the cracked interface. The

quantities $w_{0(j)}$, $\phi_{0(j)}$, and $\varphi_{0(j)}$ are determined from the surface conditions (3*a*) as:

$$\begin{Bmatrix} w_{0(j)} \\ \phi_{0(j)} \\ \varphi_{0(j)} \end{Bmatrix} = h \begin{bmatrix} c_{44(j)} & e_{15(j)} & h_{15(j)} \\ e_{15(j)} & -\in_{11(j)} & -\beta_{11(j)} \\ h_{15(j)} & -\beta_{11(j)} & -\gamma_{11(j)} \end{bmatrix}^{-1} \begin{Bmatrix} \tau_0 \\ D_0 \\ B_0 \end{Bmatrix}, \quad j=1,2. \quad (7)$$

From the constitutive equations (1), the stresses, electric displacements, and magnetic inductions can be expressed in terms of $A_{(j)}(s)$, $B_{(j)}(s)$, and $C_{(j)}(s)$. For example, we have

$$\begin{Bmatrix} \sigma_{yz(1)}(x,y) \\ D_{y(1)}(x,y) \\ B_{y(1)}(x,y) \end{Bmatrix} = -\begin{bmatrix} c_{44(1)} & e_{15(1)} & h_{15(1)} \\ e_{15(1)} & -\in_{11(1)} & -\beta_{11(1)} \\ h_{15(1)} & -\beta_{11(1)} & -\gamma_{11(1)} \end{bmatrix} \times \int_0^\infty s \frac{\sinh(sh-sy)}{\cosh(sh)} \cos(sx) \begin{Bmatrix} A_{(1)}(s) \\ B_{(1)}(s) \\ C_{(1)}(s) \end{Bmatrix} ds + \begin{Bmatrix} \tau_0 \\ D_0 \\ B_0 \end{Bmatrix}, \quad (8a)$$

$$\begin{Bmatrix} \sigma_{yz(2)}(x,y) \\ D_{y(2)}(x,y) \\ B_{y(2)}(x,y) \end{Bmatrix} = \begin{bmatrix} c_{44(2)} & e_{15(2)} & h_{15(2)} \\ e_{15(2)} & -\in_{11(2)} & -\beta_{11(2)} \\ h_{15(2)} & -\beta_{11(2)} & -\gamma_{11(2)} \end{bmatrix} \times \int_0^\infty s \frac{\sinh(sh+sy)}{\cosh(sh)} \cos(sx) \begin{Bmatrix} A_{(2)}(s) \\ B_{(2)}(s) \\ C_{(2)}(s) \end{Bmatrix} ds + \begin{Bmatrix} \tau_0 \\ D_0 \\ B_0 \end{Bmatrix}. \quad (8b)$$

Now, the unknowns $A_{(j)}(s)$, $B_{(j)}(s)$, and $C_{(j)}(s)$ can be determined from the mixed mode boundary conditions on the cracked plane. From Eqs. (8) and (5*b*), we know that

$$\begin{bmatrix} c_{44(1)} & e_{15(1)} & h_{15(1)} \\ e_{15(1)} & -\in_{11(1)} & -\beta_{11(1)} \\ h_{15(1)} & -\beta_{11(1)} & -\gamma_{11(1)} \end{bmatrix} \begin{Bmatrix} A_{(1)}(s) \\ B_{(1)}(s) \\ C_{(1)}(s) \end{Bmatrix} = -\begin{bmatrix} c_{44(2)} & e_{15(2)} & h_{15(2)} \\ e_{15(2)} & -\in_{11(2)} & -\beta_{11(2)} \\ h_{15(2)} & -\beta_{11(2)} & -\gamma_{11(2)} \end{bmatrix} \begin{Bmatrix} A_{(2)}(s) \\ B_{(2)}(s) \\ C_{(2)}(s) \end{Bmatrix}. \quad (9)$$

From Eqs. (4) and (6), and by substitution of Eq. (9), it can be shown that:

$$\int_0^\infty \cos(sx) \begin{Bmatrix} A(s) \\ B(s) \\ C(s) \end{Bmatrix} ds = 0, \quad x \geqslant a, \quad (10)$$

where

$$\begin{Bmatrix} A(s) \\ B(s) \\ C(s) \end{Bmatrix} = [E] \begin{Bmatrix} A_{(1)}(s) \\ B_{(1)}(s) \\ C_{(1)}(s) \end{Bmatrix}, \quad (11)$$

and where

$$[E] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} c_{44(2)} & e_{15(2)} & h_{15(2)} \\ e_{15(2)} & -\in_{11(2)} & -\beta_{11(2)} \\ h_{15(2)} & -\beta_{11(2)} & -\gamma_{11(2)} \end{bmatrix}^{-1} \begin{bmatrix} c_{44(1)} & e_{15(1)} & h_{15(1)} \\ e_{15(1)} & -\in_{11(1)} & -\beta_{11(1)} \\ h_{15(1)} & -\beta_{11(1)} & -\gamma_{11(1)} \end{bmatrix}. \quad (12)$$

From Eqs. (8*a*) and (5), and by substitutions of Eqs. (11) and (12), we obtain

$$\int_0^\infty s \tanh(sh) \cos(sx) \begin{Bmatrix} A(s) \\ B(s) \\ C(s) \end{Bmatrix} ds = \begin{Bmatrix} \bar{\gamma}_0 \\ \bar{E}_0 \\ \bar{H}_0 \end{Bmatrix}, \quad 0 \leqslant x < a, \quad (13)$$

where

$$\begin{Bmatrix} \bar{\gamma}_0 \\ \bar{E}_0 \\ \bar{H}_0 \end{Bmatrix} = 2[\bar{E}] \begin{Bmatrix} \tau_0 \\ D_0 - d_0 \\ B_0 - b_0 \end{Bmatrix}, \quad (14)$$

in which $\bar{E}$ is a bimaterial constant matrix:

$$[\bar{E}] = \frac{1}{2} \left(\begin{bmatrix} c_{44(1)} & e_{15(1)} & h_{15(1)} \\ e_{15(1)} & -\in_{11(1)} & -\beta_{11(1)} \\ h_{15(1)} & -\beta_{11(1)} & -\gamma_{11(1)} \end{bmatrix}^{-1} + \begin{bmatrix} c_{44(2)} & e_{15(2)} & h_{15(2)} \\ e_{15(2)} & -\in_{11(2)} & -\beta_{11(2)} \\ h_{15(2)} & -\beta_{11(2)} & -\gamma_{11(2)} \end{bmatrix}^{-1} \right). \quad (15)$$

Thus, $\bar{E}$ is the average of the inverse of two material matrices. For different material combinations (such as two layers poled oppositely, a ferroelectric layer, and a ferromagnetic layer bonded together), $\bar{E}$ has different forms. Detailed discussions are made in Subsections 3.5 and 3.6.

Following the method outlined by Li [23], the functions $A(s)$, $B(s)$, and $C(s)$ can be solved from Eqs. (10) and (13) as follows:

$$\begin{Bmatrix} A(s) \\ B(s) \\ C(s) \end{Bmatrix} = \sqrt{2\pi} s^{-1} \int_0^a \Theta(\xi) \sin(s\xi) d\xi \begin{Bmatrix} \bar{\gamma}_0 \\ \bar{E}_0 \\ \bar{H}_0 \end{Bmatrix}, \quad (16)$$

with

$$\Theta(\xi) = \frac{\sinh(2b\xi)}{\pi^2 \cosh(ba) \sqrt{\sinh^2(ba) - \sinh^2(b\xi)}} [K(\tanh(ba)) - M(\xi)], \quad (17a)$$

$$b = \frac{\pi}{2h}, \quad M(\xi) = \Pi\left(\frac{\pi}{2}, \frac{\sinh^2(ba)}{\sinh^2(ba) - \sinh^2(b\xi)}, \tanh(ba) \right), \quad (17b)$$

where K represents the complete elliptic integral of the first kind, and Π is the elliptic integral of the third kind.

Since Eq. (16) contains unknown quantities d_0 and b_0, which are the normal components of the electric displacement vector and the magnetic induction vector inside the crack (see Eq. (5)), additional assumptions are needed to obtain $A(s)$, $B(s)$, and $C(s)$. This will be discussed in the following Subsections 3.1–3.4. Once the functions $A(s)$, $B(s)$, and $C(s)$ are known, the full field solution can be obtained. In particular, the antiplane shear stress, in-plane electric displacement and magnetic induction on the cracked plane are given by:

$$\begin{Bmatrix} \sigma_{yz}(x,0) \\ D_y(x,0) \\ B_y(x,0) \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ D_0 - d_0 \\ B_0 - b_0 \end{Bmatrix} \frac{\sinh(2bx)[K(\tanh(ba)) - M(x)]}{\pi \cosh(ba) \sqrt{\sinh^2(bx) - \sinh^2(ba)}}, \quad x > a. \quad (18)$$

Define $w(x) = w_{(1)}(x,0) - w_{(2)}(x,0)$, $\phi(x) = \phi_{(1)}(x,0) - \phi_{(2)}(x,0)$ and $\varphi(x) = \varphi_{(1)}(x,0) - \varphi_{(2)}(x,0)$, which are, respectively, the displacement difference, the electric potential difference, and the

magnetic potential difference across the crack. It can be shown that:

$$\begin{Bmatrix} w(x) \\ \phi(x) \\ \varphi(x) \end{Bmatrix} = \frac{1}{\pi \cosh(ba)} \begin{Bmatrix} \bar{\gamma}_0 \\ \bar{E}_0 \\ \bar{H}_0 \end{Bmatrix} \times \int_x^a \frac{\sinh(2b\xi)[K(\tanh(ba)) - M(\xi)]}{\sqrt{\sinh^2(ba) - \sinh^2(b\xi)}} d\xi, \quad 0 \leq x \leq a, \tag{19}$$

Equation (18) indicates that the antiplane shear stress, in-plane electric displacement, and magnetic induction exhibit the usual square-root singularity near the crack tip. For many purposes, it is desirable to determine the intensity factors of magnetoelectroelastic field. From the definitions of the mode III stress intensity factor K_{III}, the electric displacement intensity factor K_D, and the magnetic induction intensity factor K_B:

$$\begin{Bmatrix} K_{\mathrm{III}} \\ K_D \\ K_B \end{Bmatrix} = \lim_{x \to +a} \begin{Bmatrix} \sigma_{yz}(x,0) \\ D_y(x,0) \\ B_y(x,0) \end{Bmatrix} \sqrt{2\pi(x-a)}, \tag{20}$$

we can obtain the closed-form expression for the field intensity factors:

$$\begin{Bmatrix} K_{\mathrm{III}} \\ K_D \\ K_B \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ D_0 - d_0 \\ B_0 - b_0 \end{Bmatrix} \sqrt{4\frac{\tanh(ba)}{\pi b}} K(\tanh(ba)), \quad b = \frac{\pi}{2h}. \tag{21}$$

The angular distributions of the stresses, electric displacements, and magnetic inductions are related to the field intensity factors through [12]:

$$\sigma_{xz} = -\frac{K_{\mathrm{III}}}{\sqrt{2\pi r}} \sin\frac{\theta}{2}, \quad \sigma_{yz} = \frac{K_{\mathrm{III}}}{\sqrt{2\pi r}} \cos\frac{\theta}{2}, \tag{22a}$$

$$D_x = -\frac{K_D}{\sqrt{2\pi r}} \sin\frac{\theta}{2}, \quad D_y = \frac{K_D}{\sqrt{2\pi r}} \cos\frac{\theta}{2}, \tag{22b}$$

$$B_x = -\frac{K_B}{\sqrt{2\pi r}} \sin\frac{\theta}{2}, \quad B_y = \frac{K_B}{\sqrt{2\pi r}} \cos\frac{\theta}{2}, \tag{22c}$$

in which r and θ are shown in Fig. 1. The energy release rate can be obtained from the virtual crack closure integral [12]:

$$G = \lim_{\delta \to 0} \frac{1}{2\delta} \int_0^\delta [\sigma_{yz}(r+a,0) w(r+a-\delta) + D_y(r+a,0)\phi(r+a-\delta) + B_y(r+a,0)\varphi(r+a-\delta)] dr. \tag{23}$$

From Eqs. (19) and (20), an expression relating G to $(K_{\mathrm{III}}, K_D, K_B)$ can be obtained:

$$G = \frac{1}{2}(K_{\mathrm{III}}, K_D, K_B)[\bar{E}] \begin{Bmatrix} K_{\mathrm{III}} \\ K_D \\ K_B \end{Bmatrix}. \tag{24}$$

The 3×3 matrix $[\bar{E}]$ has been defined in Eq. (15). Note that although Eqs. (22) and (24) are derived from the two-layer medium, it is valid for any interface crack problem in any layered magnetoelectroelastic media, subjected to any magnetoelectromechanical loads.

Now, the crack tip field intensity factors, the angular distributions of the stresses, electric displacements, and magnetic inductions, as well as the energy release rate, have been obtained analytically in terms of the applied (τ_0, D_0, B_0) on the surfaces $y = \pm h$ and (d_0, b_0) on the crack faces. Since (d_0, b_0) remains unknown, additional assumptions are needed to obtain the full solution. It is well known that in piezoelectric fracture, the electrically impermeable and permeable crack assumptions are usually adopted. Analogously, the crack can be assumed to be magnetically impermeable or permeable in magnetoelectroelastic fracture. In the following Subsections 3.1–3.4, the electrically impermeable and magnetically impermeable crack, the electrically permeable and magnetically impermeable crack, the electrically impermeable and magnetically permeable crack, the electrically permeable and magnetically permeable crack assumptions will be studied separately. In the remaining part of this paper, the symbols $\bar{E}_{ij}$ will be used to denote the elements of the matrix $[\bar{E}]$.

3.1 The Electrically Impermeable and Magnetically Impermeable Crack Assumption (Fully Impermeable Crack). For this assumption, the crack is absolutely insulated to the electric and magnetic fields. Thus, the normal components of the electric displacement vector and the magnetic induction vector vanish everywhere inside the crack. This is, $d_0=0$ and $b_0=0$. Then the field intensity factors can be obtained directly from Eq. (21) as follows:

$$\begin{Bmatrix} K_{\mathrm{III}} \\ K_D \\ K_B \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ D_0 \\ B_0 \end{Bmatrix} \sqrt{\frac{4\tanh(ba)}{\pi b}} K(\tanh(ba)). \tag{25}$$

Solutions near the crack tip and the energy release rate are obtained from Eqs. (22) and (24) by substituting Eq. (25). It is clear that for electrically and magnetically impermeable crack, the material properties do not enter into the field intensity factors.

3.2 The Electrically Permeable and Magnetically Impermeable Crack Assumption. Here, the upper and lower surfaces of the crack are electrically conducting but magnetically insulated. Hence, the electric potential jump across the crack and the normal component of the magnetic induction vector inside the crack vanish. This is, $\phi(x)=0$ and $b_0=0$. Therefore, the electric displacement inside the crack d_0 can be obtained from Eqs. (14) and (19) so that:

$$d_0 = D_0 + \frac{\bar{E}_{12}}{\bar{E}_{22}} \tau_0 + \frac{\bar{E}_{23}}{\bar{E}_{22}} B_0. \tag{26}$$

Substituting Eq. (26) into Eqs. (14) and (21) yields the field intensity factors:

$$\begin{Bmatrix} K_{\mathrm{III}} \\ K_B \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ B_0 \end{Bmatrix} \sqrt{\frac{4\tanh(ba)}{\pi b}} K(\tanh(ba)), \tag{27}$$

$$K_D = \left(-\frac{\bar{E}_{12}}{\bar{E}_{22}} K_{\mathrm{III}} - \frac{\bar{E}_{23}}{\bar{E}_{22}} K_B \right). \tag{28}$$

Solutions near the crack tip and the energy release rate are obtained from Eqs. (22) and (24) by substituting Eqs. (27) and (28). From Eq. (28) we know that the electric displacement intensity factor and the energy release rate for the electrically permeable and magnetically impermeable crack can be expressed in terms of the applied stress and magnetic induction intensity factors.

3.3 The Electrically Impermeable and Magnetically Permeable Crack Assumption. For this assumption, the upper and lower surfaces of the crack are magnetically in contact but electrically insulated. Hence, the magnetic potential jump across the crack and the normal component of the electric displacement vector inside the crack vanish. This is, $\varphi(x)=0$ and $d_0=0$. The magnetic induction inside the crack b_0 can be obtained from Eqs. (14) and (19). This gives:

$$b_0 = B_0 + \frac{\bar{E}_{13}}{\bar{E}_{33}}\tau_0 + \frac{\bar{E}_{23}}{\bar{E}_{33}}D_0. \tag{29}$$

Substituting Eq. (29) into Eqs. (14) and (21) yields the field intensity factors:

$$\begin{Bmatrix} K_{\mathrm{III}} \\ K_D \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ D_0 \end{Bmatrix} \sqrt{\frac{4\tanh(ba)}{\pi b}} K(\tanh(ba)), \tag{30}$$

$$K_B = \left(-\frac{\bar{E}_{13}}{\bar{E}_{33}} K_{\mathrm{III}} - \frac{\bar{E}_{23}}{\bar{E}_{33}} K_D \right). \tag{31}$$

Solutions near the crack tip and energy release rate are obtained from Eqs. (22) and (24) by substituting Eqs. (30) and (31). It can be shown from Eq. (31) that the magnetic induction intensity factor and the energy release rate for the electrically impermeable and magnetically permeable crack can be expressed in terms of the applied stress and electric displacement intensity factors.

3.4 The Electrically Permeable and Magnetically Permeable Crack Assumption (Fully Permeable Crack). In this case, the upper and lower surfaces of the crack are electrically and magnetically in contact. Hence, the electric potential jump and the magnetic potential jump across the crack vanish. This is, $\phi(x)=0$ and $\varphi(x)=0$. Hence, the electric displacement and magnetic induction inside the crack d_0 and b_0 can be obtained from Eqs. (14) and (19). As a result, we have:

$$\begin{Bmatrix} \bar{\gamma}_0 \\ \bar{E}_0 \\ \bar{H}_0 \end{Bmatrix} = \begin{Bmatrix} \bar{\gamma}_0 \\ 0 \\ 0 \end{Bmatrix} = [\bar{E}] \begin{Bmatrix} \tau_0 \\ D_0 - d_0 \\ B_0 - b_0 \end{Bmatrix}. \tag{32}$$

From the above equation, we know that

$$d_0 = D_0 - \frac{\bar{E}_{13}\bar{E}_{23} - \bar{E}_{33}\bar{E}_{12}}{\bar{E}_{22}\bar{E}_{33} - \bar{E}_{23}^2}\tau_0, \quad b_0 = B_0 - \frac{\bar{E}_{12}\bar{E}_{23} - \bar{E}_{22}\bar{E}_{13}}{\bar{E}_{22}\bar{E}_{33} - \bar{E}_{23}^2}\tau_0, \tag{33}$$

It then follows from Eq. (21) that

$$K_{\mathrm{III}} = \tau_0 \sqrt{\frac{4\tanh(ba)}{\pi b}} K(\tanh(ba)), \tag{34}$$

$$K_D = \frac{\bar{E}_{13}\bar{E}_{23} - \bar{E}_{33}\bar{E}_{12}}{\bar{E}_{22}\bar{E}_{33} - \bar{E}_{23}^2} K_{\mathrm{III}}, \quad K_B = \frac{\bar{E}_{12}\bar{E}_{23} - \bar{E}_{22}\bar{E}_{13}}{\bar{E}_{22}\bar{E}_{33} - \bar{E}_{23}^2} K_{\mathrm{III}}. \tag{35}$$

Solutions near the crack tip and the energy release rate are obtained from Eqs. (22) and (24) by substituting Eqs. (34) and (35). It is clear that for electrically permeable and magnetically permeable crack assumption, the crack-tip field intensity factors and the energy release rate are solely determined by the applied stress intensity factor. An explanation to this fact is that when the crack is electrically and magnetically permeable, it does not obstruct any electric displacement and magnetic induction. Thus, the crack-tip fields will not be affected by the applied electric displacement and magnetic induction loads. Because of the coupling between the electromagnetic fields and mechanical field, there exist electric displacement and magnetic induction intensity factors at the crack tip, which respond only to the applied stress intensity factor.

3.5 Bimaterial Constants. From the solution given above, it is clear that the crack-tip behavior in a magnetoelectroelastic bimaterial is governed by the bimaterial constant matrix $[\bar{E}]$, which is given in Eq. (15). If the lower layer and the upper layer have the same properties and are poled in the same direction (i.e., the crack lies in a homogeneous magnetoelectroelastic medium), then the bimaterial constant matrix is:

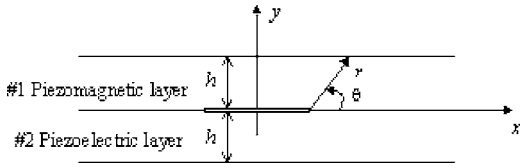


Fig. 2 A center crack between a piezoelectric layer and a piezomagnetic layer

$$[\bar{E}] = \begin{bmatrix} c_{44} & e_{15} & h_{15} \\ e_{15} & -\epsilon_{11} & -\beta_{11} \\ h_{15} & -\beta_{11} & -\gamma_{11} \end{bmatrix}^{-1}. \tag{36}$$

where because of homogeneity, the subscripts $j=1$ and $j=2$ have been omitted. The field intensity factors for this problem can be readily obtained from Subsections 3.1–3.4 by substitution of Eq. (36). The energy release rate and angular distribution of the magnetoelectroelastic fields are given by Eqs. (24) and (22).

Meanwhile, if the lower layer and the upper layer have the same properties but are poled in opposite directions (e.g., the lower layer is poled in the $-z$ direction, and the upper layer is poled in the $+z$ direction), then $c_{44(1)}=c_{44(2)}=c_{44}$, $\in_{11(1)}=\in_{11(2)}=\in_{11}$, $\gamma_{11(1)}=\gamma_{11(2)}=\gamma_{11}$, $e_{15(1)}=-e_{15(2)}=e_{15}$, $h_{15(1)}=-h_{15(2)}=h_{15}$, $\beta_{11(1)}=\beta_{11(2)}=\beta_{11}$. The bimaterial constant matrix can be obtained from Eq. (15) and it can be shown that

$$\bar{E}_{12} = \bar{E}_{13} = \bar{E}_{21} = \bar{E}_{31} = 0. \tag{37}$$

Therefore, the crack-tip field intensity factors can be obtained from related expressions in Subsections 3.1–3.4 with the known bimaterial constant matrix obtained from Eq. (15). In particular, for a fully permeable crack between two identical magnetoelectroelastic layers polarized in opposite directions, we have (from Eq. (35)):

$$K_D = K_B = 0. \tag{38}$$

Once the crack-tip magnetoelectroelastic field intensity factors have been obtained, the energy release rate and the angular distribution of the magnetoelectroelastic fields are given by Eqs. (24) and (22).

3.6 A Crack between a Piezoelectric Layer and a Piezomagnetic Layer. Magnetoelectroelastic materials usually comprise alternating piezoelectric layers and piezomagnetic layers. In this section, we consider a special case. This is, the lower medium is a piezoelectric layer and the upper layer is a piezomagnetic layer as shown in Fig. 2. The material properties for the piezoelectric layer (No. 2) and the piezomagnetic layer (No. 1) have the following form:

$$\begin{bmatrix} c_{44(2)} & e_{15(2)} & h_{15(2)} \\ e_{15(2)} & -\epsilon_{11(2)} & -\beta_{11(2)} \\ h_{15(2)} & -\beta_{11(2)} & -\gamma_{11(2)} \end{bmatrix} = \begin{bmatrix} c_{44(2)} & e_{15(2)} & 0 \\ e_{15(2)} & -\epsilon_{11(2)} & 0 \\ 0 & 0 & -\infty \end{bmatrix}, \tag{39a}$$

$$\begin{bmatrix} c_{44(1)} & e_{15(1)} & h_{15(1)} \\ e_{15(1)} & -\epsilon_{11(1)} & -\beta_{11(1)} \\ h_{15(1)} & -\beta_{11(1)} & -\gamma_{11(1)} \end{bmatrix} = \begin{bmatrix} c_{44(1)} & 0 & h_{15(1)} \\ 0 & -\infty & 0 \\ h_{15(1)} & 0 & -\gamma_{11(1)} \end{bmatrix}. \tag{39b}$$

As a result, the bimaterial constant matrix $[\bar{E}]$ defined in Eq. (15) has the form:

$$[\bar{E}] = \frac{1}{2}\frac{\begin{bmatrix} \epsilon_{11(2)} & e_{15(2)} & 0 \\ e_{15(2)} & -c_{44(2)} & 0 \\ 0 & 0 & 0 \end{bmatrix}}{c_{44(2)}\epsilon_{11(2)} + e_{15(2)}^2} + \frac{1}{2}\frac{\begin{bmatrix} \gamma_{11(1)} & 0 & h_{15(1)} \\ 0 & 0 & 0 \\ h_{15(1)} & 0 & -c_{44(1)} \end{bmatrix}}{c_{44(1)}\gamma_{11(1)} + h_{15(1)}^2}. \quad (40)$$

The stress intensity factor for this problem remains the same as that given in Eq. (21):

$$K_{\mathrm{III}} = \tau_0 \sqrt{4\frac{\tanh(ba)}{\pi b}} K(\tanh(ba)), \quad (41)$$

which are valid for both the piezoelectric layer and the piezomagnetic layer, either for the permeable crack assumption or the impermeable crack assumption.

The crack-tip electric displacement intensity factor K_D and the electric displacement inside the crack d_0 exist only in the piezoelectric layer (layer No. 2). They have the following form:

- For electrically impermeable crack,

$$d_0 = 0, \quad K_D = D_0 \sqrt{4\frac{\tanh(ba)}{\pi b}} K(\tanh(ba)). \quad (42)$$

- For electrically permeable crack,

$$d_0 = D_0 - \frac{e_{15(2)}}{c_{44(2)}}\tau_0, \quad K_D = \frac{e_{15(2)}}{c_{44(2)}} K_{\mathrm{III}}. \quad (43)$$

The crack-tip magnetic induction intensity factor K_B and the magnetic induction inside the crack b_0 exist only in the piezomagnetic layer (layer No. 1). They have the following forms:

- For magnetically impermeable crack,

$$b_0 = 0, \quad K_B = B_0 \sqrt{4\frac{\tanh(ba)}{\pi b}} K(\tanh(ba)). \quad (44)$$

- For magnetically permeable crack,

$$b_0 = B_0 - \frac{h_{15(1)}}{c_{44(1)}}\tau_0, \quad K_B = \frac{h_{15(1)}}{c_{44(1)}} K_{\mathrm{III}}. \quad (45)$$

Once the crack-tip field intensity factors are obtained, the angular distributions of stresses, the electric displacements, and magnetic inductions can be obtained from Eqs. (22), and the energy release rate G can be determined from Eq. (24), in which the matrix $[\bar{E}]$ is defined in Eq. (40). It is clear that the values of G are different for the piezoelectric layer and the piezomagnetic layer. G also depends on the crack-face electric and magnetic boundary condition assumptions.

3.7 Solutions for an Infinitesimal Crack and a Semi-Infinite Crack. In Subsections 3.1–3.6, the thickness of the magnetoelectroelastic layers h is finite. From the solutions given in these subsections, it is possible to obtain the solutions for an infinitesimal crack (i.e., for $a \ll h$). Since

$$\lim_{h\to\infty} \sqrt{\frac{4\tanh(ba)}{\pi b}} K(\tanh(ba)) = \sqrt{\pi a}, \quad (46)$$

solutions for the infinitesimal crack are derived by replacing the quantity $\sqrt{4\tanh(ba)/\pi b}K(\tanh(ba))$ in Subsections 3.1–3.6 with $\sqrt{\pi a}$.

Closed-form expressions are also available for a semi-infinite crack in magnetoelectroelastic layers. In this case, the crack length a is considerably larger than the magnetoelectroelastic layers (i.e., for $a \gg h$). Since

$$\lim_{a\to\infty} \sqrt{\frac{4\tanh(ba)}{\pi b}} K(\tanh(ba)) = \sqrt{2h}, \quad (47)$$

solutions for the semi-infinite crack are obtained by replacing the quantity $\sqrt{4\tanh(ba)/\pi b}K(\tanh(ba))$ in Subsections 3.1–3.6 with $\sqrt{2h}$.

After obtaining the crack-tip field intensity factors, the angular distributions of the magnetoelectroelastic fields can be obtained from Eqs. (22), and the energy release rate G can be determined from Eq. (24).

4 Closed-Form Solution for Prescribed Displacement, Electric Potential, and Magnetic Potential on the Medium Surfaces

Here, the prescribed boundary conditions (3*b*) are considered. Since the normal components of stress, electric displacement, and magnetic induction must be continuous, the quantities (w^+, ϕ^+, φ^+) and (w^-, ϕ^-, φ^-) cannot be independent and they must satisfy the following additional constraints:

$$\begin{bmatrix} c_{44(1)} & e_{15(1)} & h_{15(1)} \\ e_{15(1)} & -\epsilon_{11(1)} & -\beta_{11(1)} \\ h_{15(1)} & -\beta_{11(1)} & -\gamma_{11(1)} \end{bmatrix} \begin{Bmatrix} w^+ \\ \phi^+ \\ \varphi^+ \end{Bmatrix} = -\begin{bmatrix} c_{44(2)} & e_{15(2)} & h_{15(2)} \\ e_{15(2)} & -\epsilon_{11(2)} & -\beta_{11(2)} \\ h_{15(2)} & -\beta_{11(2)} & -\gamma_{11(2)} \end{bmatrix} \times \begin{Bmatrix} w^- \\ \phi^- \\ \varphi^- \end{Bmatrix}. \quad (48)$$

Once again, the right-half portion of the medium (i.e. $x>0$) is investigated because of symmetry. In the present case, an appropriate solution of Eq. (2), in connection with Eq. (1), can be expressed as the following integrals:

$$\begin{Bmatrix} w_{(1)}(x,y) \\ \phi_{(1)}(x,y) \\ \varphi_{(1)}(x,y) \end{Bmatrix} = \int_0^\infty \frac{\sinh(sh-sy)}{\sinh(sh)} \cos(sx) \begin{Bmatrix} A_{(1)}(s) \\ B_{(1)}(s) \\ C_{(1)}(s) \end{Bmatrix} ds + \begin{Bmatrix} w^+ \\ \phi^+ \\ \varphi^+ \end{Bmatrix} \frac{y}{h}, \quad 0 \leqslant y \leqslant h, \quad (49a)$$

$$\begin{Bmatrix} w_{(2)}(x,y) \\ \phi_{(2)}(x,y) \\ \varphi_{(2)}(x,y) \end{Bmatrix} = \int_0^\infty \frac{\sinh(sh+sy)}{\sinh(sh)} \cos(sx) \begin{Bmatrix} A_{(2)}(s) \\ B_{(2)}(s) \\ C_{(2)}(s) \end{Bmatrix} ds - \begin{Bmatrix} w^- \\ \phi^- \\ \varphi^- \end{Bmatrix} \frac{y}{h}, \quad -h \leqslant y \leqslant 0, \quad (49b)$$

where $x \geqslant 0$, $(j=1,2)$ denotes the layer number, $A_{(j)}(s)$, $B_{(j)}(s)$, and $C_{(j)}(s)$ are unknown functions to be determined from the prescribed conditions (4) and (5) on the cracked interface. The stresses, electric displacements, and magnetic inductions associated with Eqs. (49) are obtained from the constitutive equations (1). Hence,

$$\begin{Bmatrix} \sigma_{yz(1)}(x,y) \\ D_{y(1)}(x,y) \\ B_{y(1)}(x,y) \end{Bmatrix} = -\begin{bmatrix} c_{44(1)} & e_{15(1)} & h_{15(1)} \\ e_{15(1)} & -\epsilon_{11(1)} & -\beta_{11(1)} \\ h_{15(1)} & -\beta_{11(1)} & -\gamma_{11(1)} \end{bmatrix} \times \int_0^\infty s\frac{\cosh(sh-sy)}{\sinh(sh)}\cos(sx)\begin{Bmatrix} A_{(1)}(s) \\ B_{(1)}(s) \\ C_{(1)}(s) \end{Bmatrix} ds + \begin{Bmatrix} \tau_0 \\ D_0 \\ B_0 \end{Bmatrix}, \quad (50a)$$

$$\begin{Bmatrix} \sigma_{yz(2)}(x,y) \\ D_{y(2)}(x,y) \\ B_{y(2)}(x,y) \end{Bmatrix} = \begin{bmatrix} c_{44(2)} & e_{15(2)} & h_{15(2)} \\ e_{15(2)} & -\epsilon_{11(2)} & -\beta_{11(2)} \\ h_{15(2)} & -\beta_{11(2)} & -\gamma_{11(2)} \end{bmatrix} \times \left(\int_0^\infty s\frac{\cosh(sh+sy)}{\sinh(sh)}\cos(sx)\begin{Bmatrix} A_{(2)}(s) \\ B_{(2)}(s) \\ C_{(2)}(s) \end{Bmatrix} ds \right) + \begin{Bmatrix} \tau_0 \\ D_0 \\ B_0 \end{Bmatrix}, \quad (50b)$$

where

$$\begin{Bmatrix} \tau_0 \\ D_0 \\ B_0 \end{Bmatrix} = \frac{1}{h}\begin{bmatrix} c_{44(1)} & e_{15(1)} & h_{15(1)} \\ e_{15(1)} & -\epsilon_{11(1)} & -\beta_{11(1)} \\ h_{15(1)} & -\beta_{11(1)} & -\gamma_{11(1)} \end{bmatrix}\begin{Bmatrix} w^+ \\ \phi^+ \\ \varphi^+ \end{Bmatrix} = -\frac{1}{h}\begin{bmatrix} c_{44(2)} & e_{15(2)} & h_{15(2)} \\ e_{15(2)} & -\epsilon_{11(2)} & -\beta_{11(2)} \\ h_{15(2)} & -\beta_{11(2)} & -\gamma_{11(2)} \end{bmatrix}\begin{Bmatrix} w^- \\ \phi^- \\ \varphi^- \end{Bmatrix}. \quad (51)$$

The unknowns $A_{(j)}(s)$, $B_{(j)}(s)$, and $C_{(j)}(s)$ will be determined from the mixed-mode boundary conditions on the crack interface. From Eqs. (50) and (5*b*) it can be shown that $A_{(j)}(s)$, $B_{(j)}(s)$, and $C_{(j)}(s)$ also satisfy Eq. (9) for the present problem. From Eqs. (4) and (49), and by substitution of Eq. (9), we have

$$\int_0^\infty \cos(sx)\begin{Bmatrix} A(s) \\ B(s) \\ C(s) \end{Bmatrix} ds = 0, \quad x \geq a, \quad (52)$$

where $(A(s),B(s),C(s))$ are related to $(A_1(s),B_1(s),C_1(s))$ through Eq. (11), in which the matrix $[E]$ is given in Eq. (12). From Eqs. (50*a*) and (5), and by substitution of Eqs. (11) and (12), we obtain

$$\int_0^\infty s\coth(sh)\cos(sx)\begin{Bmatrix} A(s) \\ B(s) \\ C(s) \end{Bmatrix} ds = \begin{Bmatrix} \bar{\gamma}_0 \\ \bar{E}_0 \\ \bar{H}_0 \end{Bmatrix}, \quad 0 \leq x < a, \quad (53)$$

where $(\bar{\gamma}_0,\bar{E}_0,\bar{H}_0)$ is related to (w^+,ϕ^+,φ^+) and (w^-,ϕ^-,φ^-) through Eqs. (14) and (15) and the bimaterial constant matrix $\bar{E}$ is defined in Eq. (15).

Define $w(x)=w_{(1)}(x,0)-w_{(2)}(x,0)$, $\phi(x)=\phi(x,0)-\phi_{(2)}(x,0)$, and $\varphi(x)=\varphi_{(1)}(x,0)-\varphi_{(2)}(x,0)$, which are, respectively, the displacement difference, the electric potential difference, and the magnetic potential difference across the crack. It can be shown from Eqs. (52) and (53) that

$$\begin{Bmatrix} w(x) \\ \phi(x) \\ \varphi(x) \end{Bmatrix} = \frac{1}{b}\cos^{-1}\left(\frac{\cosh(bx)}{\cosh(ba)}\right)\begin{Bmatrix} \bar{\gamma}_0 \\ \bar{E}_0 \\ \bar{H}_0 \end{Bmatrix}, \quad 0 \leq x \leq a, \quad (54)$$

where $b=\pi/2$ h. Since Eq. (54) contains unknown quantities d_0 and b_0, which are the normal components of the electric displacement vector and the magnetic induction vector inside the crack (see Eqs. (5) and (14)), additional assumptions are needed to obtain $A(s)$, $B(s)$, and $C(s)$. This will be discussed in Subsections 4.1–4.4 If the functions $A(s)$, $B(s)$, and $C(s)$ are known, the full field solution can be obtained. In particular, the antiplane shear stress, in-plane electric displacement, and magnetic induction on the cracked plane are as follows:

$$\begin{Bmatrix} \sigma_{yz}(x,0) \\ D_y(x,0) \\ B_y(x,0) \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ D_0-d_0 \\ B_0-b_0 \end{Bmatrix}\frac{\sinh(bx)}{\sqrt{\sinh^2(bx)-\sinh^2(ba)}}, \quad x > a. \quad (55)$$

In Eq. (55), the parameters (τ_0,D_0,B_0) are determined from Eq. (51). Equation (55) indicates that the antiplane shear stress, in-plane electric displacement, and magnetic induction exhibit the usual square-root singularity near the crack tip. From the definitions of Eq. (20), the closed-form expression for the field intensity factors can be obtained as:

$$\begin{Bmatrix} K_{\mathrm{III}} \\ K_D \\ K_B \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ D_0-d_0 \\ B_0-b_0 \end{Bmatrix}\sqrt{2h\tanh\left(\frac{\pi a}{2h}\right)}. \quad (56)$$

The angular distributions of the stresses, electric displacements, and magnetic inductions are related to the field intensity factors through Eq. (22) and the energy release rates are given in Eq. (24).

The closed-form crack-tip field solutions have now been obtained for the prescribed displacement, electric potential, and magnetic potential loads. Since the quantities (d_0,b_0) inside the crack remain unknown, additional assumptions are needed to obtain the full solutions. In Subsections 4.1–4.4, the electrically impermeable and magnetically impermeable crack, the electrically permeable and magnetically impermeable crack, the electrically impermeable and magnetically permeable crack, the electrically permeable and magnetically permeable crack assumptions will be separately studied. In all cases, the loading parameters (τ_0,D_0,B_0) are related to the applied mechanical displacement, electric potential, and magnetic potential through Eq. (51).

4.1 The Electrically Impermeable and Magnetically Impermeable Crack Assumption (Fully Impermeable Crack). For such assumption, the crack is absolutely insulated to electric and magnetic fields. Hence, the normal components of the electric displacement vector and the magnetic induction vector vanish everywhere inside the crack. This is, $d_0=0$ and $b_0=0$. Then, the field intensity factors can be obtained directly from Eq. (56) as follows:

$$\begin{Bmatrix} K_{\mathrm{III}} \\ K_D \\ K_B \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ D_0 \\ B_0 \end{Bmatrix}\sqrt{2h\tanh\left(\frac{\pi a}{2h}\right)}. \quad (57)$$

Solutions near the crack tip and the energy release rate are obtained from Eqs. (22) and (24) with the substitution of Eq. (57). It is clear that for electrically and magnetically impermeable crack, the material properties do not enter into the field intensity factors.

4.2 The Electrically Permeable and Magnetically Impermeable Crack Assumption. Under this assumption, the upper and lower surfaces of the crack are electrically conducting but magnetically insulated. Hence, the electric potential jump across the crack and the normal component of the magnetic induction

vector inside the crack vanish. This is, $\phi(x)=0$ and $b_0=0$. Therefore, the electric displacement inside the crack d_0 can be obtained from Eqs. (14) and (54) so that

$$d_0 = D_0 + \frac{\bar{E}_{12}}{\bar{E}_{22}}\tau_0 + \frac{\bar{E}_{23}}{\bar{E}_{22}}B_0. \tag{58}$$

Substituting Eq. (58) into Eqs. (14) and (56) yields the field intensity factors:

$$\begin{Bmatrix} K_{\mathrm{III}} \\ K_B \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ B_0 \end{Bmatrix} \sqrt{2h \tanh\left(\frac{\pi a}{2h}\right)}, \quad K_D = \left(-\frac{\bar{E}_{12}}{\bar{E}_{22}}K_{\mathrm{III}} - \frac{\bar{E}_{23}}{\bar{E}_{22}}K_B \right). \tag{59}$$

Solutions near the crack tip and energy release rate are obtained from Eqs. (22) and (24) by substituting Eqs. (58) and (59). From Eq. (59), we know that the electric displacement intensity factor and the energy release rate for the electrically permeable and magnetically impermeable crack can be expressed in terms of the applied stress and magnetic induction intensity factors.

4.3 The Electrically Impermeable and Magnetically Permeable Crack Assumption. Here, the upper and lower surfaces of the crack are magnetically in contact but electrically insulated. Thus, the magnetic potential jump across the crack and the normal component of the electric displacement vector inside the crack vanish. This is, $\varphi(x)=0$ and $d_0=0$. Hence, the magnetic induction inside the crack b_0 can be obtained from Eqs. (14) and (54). We have

$$b_0 = B_0 + \frac{\bar{E}_{13}}{\bar{E}_{33}}\tau_0 + \frac{\bar{E}_{23}}{\bar{E}_{33}}D_0. \tag{60}$$

Substituting Eq. (29) into Eqs. (14) and (56) yields the field intensity factors:

$$\begin{Bmatrix} K_{\mathrm{III}} \\ K_D \end{Bmatrix} = \begin{Bmatrix} \tau_0 \\ D_0 \end{Bmatrix} \sqrt{2h \tanh\left(\frac{\pi a}{2h}\right)}, \quad K_B = \left(-\frac{\bar{E}_{13}}{\bar{E}_{33}}K_{\mathrm{III}} - \frac{\bar{E}_{23}}{\bar{E}_{33}}K_D \right). \tag{61}$$

Solutions near the crack tip and the energy release rate are obtained from Eqs. (22) and (24) by substituting Eq. (61). It can be shown that the magnetic induction intensity factor and the energy release rate for the electrically impermeable and magnetically permeable crack can be expressed in terms of the applied stress and electric displacement intensity factors.

4.4 The Electrically Permeable and Magnetically Permeable Crack Assumption (Fully Permeable Crack). For this assumption, the upper and lower surfaces of the crack are electrically and magnetically in contact. Hence, the electric potential and the magnetic potential jumps across the crack vanish. This is, $\phi(x)=0$ and $\varphi(x)=0$. Therefore, the electric displacement and the magnetic induction inside the crack d_0 and b_0 can be obtained from Eqs. (14) and (54). Thus, we have

$$\begin{Bmatrix} \bar{\gamma}_0 \\ \bar{E}_0 \\ \bar{H}_0 \end{Bmatrix} = \begin{Bmatrix} \bar{\gamma}_0 \\ 0 \\ 0 \end{Bmatrix} = [\bar{E}] \begin{Bmatrix} \tau_0 \\ D_0 - d_0 \\ B_0 - b_0 \end{Bmatrix}. \tag{62}$$

From the above equation, we know that

$$d_0 = D_0 - \frac{\bar{E}_{13}\bar{E}_{23} - \bar{E}_{33}\bar{E}_{12}}{\bar{E}_{22}\bar{E}_{33} - \bar{E}_{23}^2}\tau_0, \quad b_0 = B_0 - \frac{\bar{E}_{12}\bar{E}_{23} - \bar{E}_{22}\bar{E}_{13}}{\bar{E}_{22}\bar{E}_{33} - \bar{E}_{23}^2}\tau_0, \tag{63}$$

It then follows from Eq. (56) that

$$K_{\mathrm{III}} = \tau_0 \sqrt{2h \tanh\left(\frac{\pi a}{2h}\right)}, \tag{64a}$$

$$K_D = \frac{\bar{E}_{13}\bar{E}_{23} - \bar{E}_{33}\bar{E}_{12}}{\bar{E}_{22}\bar{E}_{33} - \bar{E}_{23}^2}K_{\mathrm{III}}, \quad K_B = \frac{\bar{E}_{12}\bar{E}_{23} - \bar{E}_{22}\bar{E}_{13}}{\bar{E}_{22}\bar{E}_{33} - \bar{E}_{23}^2}K_{\mathrm{III}}. \tag{64b}$$

Solutions near the crack tip and the energy release rate are obtained from Eqs. (22) and (24) by substituting of Eqs. (64). It is clear that for the electrically permeable and magnetically permeable crack assumption, the crack-tip field intensity factors and the energy release rate are solely determined by the applied stress intensity factor.

4.5 Two Magnetoelectroelastic Layers of the Same Properties Bonded Together. Similar to the case of stress, electric displacement, and magnetic induction loads applied on the medium surfaces, the solution for the case of mechanical displacement, electric potential and magnetic potential can be readily obtained if the upper layer and the bottom layer have the same properties:

- If the lower layer and the upper layer have the same properties and are poled in the same direction (i.e., the crack is located in a homogeneous magnetoelectroelastic medium), then the solution can be obtained from Subsections 4.1–4.4 by substitution of the bimaterial constant matrix $[\bar{E}]$ given by Eq. (36).
- If the lower layer and the upper layer have the same properties but are poled in opposite directions (e.g., the lower layer is poled in the $-z$ direction, and the upper layer is poled in the $+z$ direction), then $c_{44(1)}=c_{44(2)}=c_{44}$, $\in_{11(1)}=\in_{11(2)}=\in_{11}$, $\gamma_{11(1)}=\gamma_{11(2)}=\gamma_{11}$, $e_{15(1)}=-e_{15(2)}=e_{15}$, $h_{15(1)}=-h_{15(2)}=h_{15}$, $\beta_{11(1)}=\beta_{11(2)}=\beta_{11}$. The bimaterial constant matrix can be obtained from Eq. (15). The non-zero elements in $[\bar{E}]$ are given in Eq. (37). In particular, for a fully permeable crack between two same magnetoelectroelastic layers polarized in opposite directions, we also have: $K_D=K_B=0$.

Once the crack-tip magnetoelectroelastic field intensity factors have been obtained, the energy release rate and the angular distribution of the magnetoelectroelastic fields are given by Eqs. (24) and (22), respectively.

4.6 A Crack between a Piezoelectric Layer and a Piezomagnetic Layer. Now consider the case of a piezoelectric layer bonded with a piezomagnetic layer shown in Fig. 2. In this problem, the magnetic potential in the piezoelectric layer and the electric potential in the piezomagnetic layer should vanish, and the bimaterial constant matrix $[\bar{E}]$ defined in Eq. (15) has the form of Eq. (40). Because of the continuity of the stress σ_{yz}, the mechanical displacement, the electric potential, and the magnetic potential must satisfy the following relations (deduced from Eq. (51)):

$$c_{44(1)}w^+ + h_{15(1)}\varphi^+ = c_{44(2)}w^- + e_{15(2)}\phi^-. \tag{65}$$

The equivalent mechanical, electrical, and magnetical loads τ_0, D_0, and B_0 are expressed as follows:

$$\begin{Bmatrix} \tau_0 \\ B_0 \end{Bmatrix} = \frac{1}{h}\begin{bmatrix} c_{44(1)} & h_{15(1)} \\ h_{15(1)} & -\gamma_{11(1)} \end{bmatrix}\begin{Bmatrix} w^+ \\ \varphi^+ \end{Bmatrix}, \quad \begin{Bmatrix} \tau_0 \\ D_0 \end{Bmatrix} = -\frac{1}{h}\begin{bmatrix} c_{44(2)} & e_{15(2)} \\ e_{15(2)} & -\in_{11(2)} \end{bmatrix}\begin{Bmatrix} w^- \\ \varphi^- \end{Bmatrix}. \tag{66}$$

In this case, the stress intensity factor remains the same as that given in Eq. (56):

$$K_{\mathrm{III}} = \tau_0 \sqrt{2h \tanh\left(\frac{\pi a}{2h}\right)}, \tag{67}$$

which are valid for both the piezoelectric layer and the piezomagnetic layer, either for the permeable crack assumption or for the impermeable crack assumption.

The crack-tip electric displacement intensity factor K_D and the electric displacement inside the crack d_0 exist only in the piezoelectric layer (layer No. 2). They have the following forms:

- For electrically impermeable crack,

$$d_0 = 0, \quad K_D = D_0 \sqrt{2h \tanh\left(\frac{\pi a}{2h}\right)}. \tag{68}$$

- For electrically permeable crack,

$$d_0 = D_0 - \frac{e_{15(2)}}{c_{44(2)}}\tau_0, \quad K_D = \frac{e_{15(2)}}{c_{44(2)}} K_{\mathrm{III}}. \tag{69}$$

The crack-tip magnetic induction intensity factor K_B and the magnetic induction inside the crack b_0 exist only in the piezomagnetic layer (layer No. 1). They have the following forms:

- For magnetically impermeable crack,

$$b_0 = 0, \quad K_B = B_0 \sqrt{2h \tanh\left(\frac{\pi a}{2h}\right)}. \tag{70}$$

- For magnetically permeable crack,

$$b_0 = B_0 - \frac{h_{15(1)}}{c_{44(1)}}\tau_0, \quad K_B = \frac{h_{15(1)}}{c_{44(1)}} K_{\mathrm{III}}. \tag{71}$$

When the crack-tip field intensity factors are obtained, the angular distributions of stresses, electric displacements, and magnetic inductions can be obtained from Eqs. (22), and the energy release rate G determined from Eq. (24), in which the matrix $[\bar{E}]$ is defined by Eq. (40). It is clear that the values of G are different for the piezoelectric layer and the piezomagnetic layer. G also depends on the crack-face electric and magnetic boundary condition assumptions.

4.7 Solution for an Infinitesimal Crack and a Semi-Infinite Crack. In Subsections 4.1–4.6, the thickness of the magnetoelectroelastic layers h is finite. From the solutions given in these subsections, it is possible to obtain the solutions for an infinitesimal crack (i.e., for $a \ll h$). Since

$$\lim_{h\to\infty} \sqrt{2h \tanh\left(\frac{\pi a}{2h}\right)} = \sqrt{\pi a}, \tag{72}$$

solutions for the infinitesimal crack are obtained by replacing the quantity $\sqrt{2h \tanh(\pi a/2h)}$ in Subsections 4.1–4.6 with $\sqrt{\pi a}$.

Closed-form expressions are also available for a semi-infinite crack in magnetoelectroelastic layers. In this case, the crack length a is considerably larger than the magnetoelectroelastic layers (i.e., for $a \gg h$). Since

$$\lim_{a\to\infty} \sqrt{2h \tanh\left(\frac{\pi a}{2h}\right)} = \sqrt{2h}, \tag{73}$$

and solutions for the semi-infinite crack are derived by replacing the quantity $\sqrt{2h \tanh(\pi a/2h)}$ in Subsections 4.1–4.6 with $\sqrt{2h}$.

After obtaining the crack-tip field intensity factors, the angular distributions of the magnetoelectroelastic fields can be obtained from Eqs. (22) and the energy release rate G from Eq. (24).

5 Conclusions

An interface crack in a two-layered magnetoelectroelastic medium subjected to mechanical, electrical, and magnetical loads on its surfaces are studied in this paper and the following points are noted.

- Closed-form solution has been obtained for a crack between two dissimilar magnetoelectroelastic layers of finite thickness. Expressions for the crack-tip field intensity factors, the electromagnetic fields inside the crack are given. The electrically and magnetically impermeable and permeable crack assumptions are investigated.
- The energy release rate can be explicitly expressed in terms of the field intensity factors for the interface cracks in layered magnetoelectroelastic media (by Eq. (24), in which the material constant matrix $[\bar{E}]$ is given by Eq. (15)). The expression is valid not only for interface cracks in two-layer media, but also valid for interface cracks in multilayered media.
- Closed-form solutions for a center crack in a homogeneous magnetoelectroelastic layer, a crack between two similar magnetoelectroelastic layers polarized in opposite directions, and a crack between a piezoelectric layer and a piezomagnetic layer are obtained.
- Applications of electric and magnetic fields do not alter the stress intensity factors. The values of the stress intensity factor are identical for any kind of crack-face electric and magnetic boundary condition assumptions (i.e., the crack-face electric and magnetic boundary conditions have no effect on the stress intensity factor.). The result of the stress intensity factor is the same as the solution for the elastic materials given by Li [23].

Acknowledgment

The authors would like to thank the Australian Research Council (ARC) for the financial support of this project. The authors (B.L.W. and Y.W.M.) are, respectively, Australian Research Fellow and Australian Federation Fellow, supported by the ARC and tenable at the University of Sydney.

References

[1] Alshits, I., Darinskii, A. N., and Lothe, J., 1992, "On the Existence of Surface Waves in Half-Anisotropic Elastic Media With Piezoelectric and Piezomagnetic Properties," Wave Motion, **16**, pp. 265–283.
[2] Van Run, A. M. J. G., Terrell, D. R., and Scholing, J. H., 1974, "An *In Situ* Grown Eutectic Magnetoelectric Composite Material," J. Mater. Sci., **9**, pp. 1710–1714.
[3] Bracke, L. P. M., and Van Vliet, R. G., 1981, "A Broadband Magnetoelectric Transducer Using a Composite Material," Int. J. Electron., **51**, pp. 255–262.
[4] Benveniste, Y., 1995, "Magnetoelectric Effect in Fibrous Composites With Piezoelectric and Piezomagnetic Phases," Phys. Rev. B, **51**, pp. 16,424–16,427.
[5] Pan, E., and Heyliger, P. R., 2002, "Free Vibrations of Simply Supported and Multilayered Magnetoelectroelastic Plates," J. Sound Vib., **252**, pp. 429–442.
[6] Liu, J.-X., Liu, X., and Zhao, Y., 2001, "Green's Functions for Anisotropic Magnetoelectroelastic Solids With an Elliptical Cavity or a Crack," Int. J. Eng. Sci., **39**(12), pp. 1405–1418.
[7] Gao, C. F., Kessler, H., and Balke, H., 2003a, "Crack Problems in Magnetoelectroelastic Solids. Part I: Exact Solution of a Crack," Int. J. Eng. Sci., **41**, pp. 969–981.
[8] Gao, C. F., Kessler, H., and Balke, H., 2003b, "Crack Problems in Magnetoelectroelastic Solids. Part II: General Solution of Collinear Cracks," Int. J. Eng. Sci., **41**, pp. 983–994.
[9] Gao, C. F., Tong, P., and Zhang, T. Y., 2004, "Fracture Mechanics for a Mode III Crack in a Magnetoelectroelastic Solid," Int. J. Solids Struct., **41**, pp. 6613–6629.
[10] Zhou, Z. G., Wang, B., and Sun, Y. G., 2004, "Two Collinear Interface Cracks in Magnetoelectroelastic Composites," Int. J. Eng. Sci., **42**, pp. 1155–1167.
[11] Wang, B. L., and Mai, Y.-W., 2003, "Crack Tip Field in Piezoelectric/Piezomagnetic Media," Eur. J. Mech. A/Solids, **22**, pp. 591–602.
[12] Wang, B. L., and Mai, Y.-W., 2004, "Fracture of Piezoelectromagnetic Materials," Mech. Res. Commun., **31**, pp. 65–73.
[13] Gao, C. F., Tong, P., and Zhang, T. Y., 2004, "Fracture Mechanics for a Mode III Crack in a Magnetoelectroelastic Solid," Int. J. Solids Struct., **41**, pp. 6613–6629.

[14] Liu, J. X., Liu, X. L., and Zhao, Y. B., 2001, "Green's Functions for Anisotropic Magnetoelectroelastic Solids With an Elliptical Cavity or a Crack," Int. J. Eng. Sci., **39**, pp. 1405–1418.
[15] Tian, W. Y., and Gabbert, U., 2005a, "Macrocrack-Microcrack Interaction Problem in Magnetoelectroelastic Solids," Mech. Mater., **37**(5), pp. 565–592.
[16] Tian, W. Y., and Gabbert, U., 2005b, "Parallel Crack Near the Interface of Magnetoelectroelastic Bimaterials," Comput. Mater. Sci., **32**(3–4), pp. 562–567.
[17] Chung, M. Y., and Ting, T. C. T., 1995, "The Green Function for a Piezoelectric Piezomagnetic Anisotropic Elastic Medium With an Elliptic Hole or Rigid Inclusion," Philos. Mag. Lett., **72**, pp. 405–410.
[18] Hu, K. Q., and Li, G. Q., 2005, "Constant Moving Crack in a Magnetoelectroelastic Material under Antiplane Shear Loading," Int. J. Solids Struct., **42**(9–10), pp. 2823–2835.
[19] Ryu, J., Priya, S., Carazo, A. V., Uchino, K., and Kim, H. E., 2001, "Effect of Magnetostricitive Layer on Magnetoelectric Properties in Lead Zirconate Titanate/Terfenol-D Laminate Composites," J. Am. Ceram. Soc., **84**, pp. 2905–2908.
[20] Ryu, J., Priya, S., Uchino, K., and Kim, H. E., 2002, "Magnetoelectric Effect in Composites of Magnetostrictive and Piezoelectric Materials," J. Electroceram., **8**, pp. 107–119.
[21] Harshe, G., Dougherty, J. P., and Newnham, R. E., 1993, "Theoretical Modeling of Multiplayer Magnetoelectric Composites," Int. J. Appl. Electromagn. Mater., **4**, pp. 145–159.
[22] Chen, Z. R., Yu, S. W., Lu, M., and Ye, L., 2002, "Effective Properties of Layered Magnetoelectroelastic Composites," Compos. Struct., **57**, pp. 177–182.
[23] Li, X. F., 2001, "Closed-Form Solution for a Mode III Interface Crack Between Two Bonded Dissimilar Elastic Layers," Int. J. Fract., **109**, pp. L3–L8.

Calendering Pseudoplastic and Viscoplastic Fluids With Slip at the Roll Surface

E. Mitsoulis[1]
e-mail: mitsouli@metal.ntua.gr

S. Sofou

School of Mining Engineering and Metallurgy,
National Technical University of Athens,
Zografou 15780,
Athens, Greece

The lubrication approximation theory (LAT) is used to provide numerical results for calendering a sheet from an infinite reservoir. The Herschel–Bulkley model of viscoplasticity is employed, which reduces with appropriate modifications to the Bingham, power-law, and Newtonian models. A dimensionless slip coefficient is introduced to account for the case of slip at the roll surfaces. The results give the final sheet thickness as a function of the dimensionless power-law index (in the case of pseudoplasticity), the Bingham number or the dimensionless yield stress calculated at the nip (in the case of viscoplasticity), and the dimensionless slip coefficient in both cases. Integrated quantities of engineering interest are also calculated. These include the maximum pressure, the roll-separating force, and the power input to the rolls. Decreasing the power-law index or increasing the dimensionless yield stress lead to excess sheet thickness over the thickness at the nip. All engineering quantities calculated in dimensionless form increase substantially with the departure from the Newtonian values. The presence of slip decreases pressure and the engineering quantities and increases the domain in all cases. [DOI: 10.1115/1.2083847]

1 Introduction

The calendering process is used in a variety of industries, including those of paper, plastics and rubber, for the production of rolled sheets or films of specific thickness and final appearance. The procedure has been theoretically introduced by Gaskell [1] and involves the feed from an infinite reservoir of a pair of corotating and equal sized heated rolls (calenders) with a material to form a sheet (Fig. 1(*a*)) [2]. A variation of the process is to feed the calenders with a finite sheet (Fig. 1(*b*)).

With regard to the case of calendering from an infinite reservoir, both the acquired sheet thickness and the values of the relevant engineering quantities (such as maximum pressure values) are primarily dependent on the rheological properties of the material. Since many materials used in calendering are frequently non-Newtonian, exhibiting *pseudoplastic* (shear thinning or thickening) [2] or *viscoplastic* (presence of a yield stress) behavior (see, e.g., Bird et al. [3]), a model that covers both cases is employed. The Herschel–Bulkley model has the advantage of reducing—with an appropriate choice of parameters—to the Bingham, power-law, or Newtonian model. In simple shear flow, it takes the form [3]:

$$\tau = K|\dot{\gamma}|^{n-1}\dot{\gamma} \pm \tau_y, \quad \text{for } |\tau| > \tau_y, \tag{1a}$$

$$\dot{\gamma} = 0, \quad \text{for } |\tau| \leqslant \tau_y, \tag{1b}$$

where τ is the shear stress, $\dot{\gamma}=du/dy$ is the shear rate (or velocity gradient in the y direction), u is the axial velocity, τ_y is the apparent yield stress, K is the consistency index, and n is the power-law index. Note that when $n=1$ and $K=\mu$ (a constant), the Herschel–Bulkley model reduces to the Bingham model. When $\tau_y=0$, the power-law model is recovered, and when $\tau_y=0$ and $n=1$, the Newtonian model is obtained.

It should be noted that in viscoplastic models, when the shear stress τ falls below τ_y, a solid structure is formed (unyielded). The two regions (yielded/unyielded) are separated by a yield line, related to the pressure gradient in the flow as will be shown later. A previous work by the authors [4] has studied the process of calendering by assuming no slip at the wall. Since many materials, with or without yield, are known to exhibit slip at the wall [5], it is the purpose of the present work to study in a parametric manner the process of calendering for pseudoplastic and viscoplastic materials obeying the Herschel–Bulkley model with slip at the wall.

2 Mathematical Modeling

2.1 Governing Equations. Gaskell [1] carried out an early analysis of calendering, while Middleman [2] gave in his textbook the theoretical treatment and known results up to 1977. As explained by Middleman [2] and with regard to Fig. 1, the lubrication approximation theory (LAT) regards locally fully developed shear flow between the rolls. The conservation of momentum then gives

$$\frac{dP}{dx} = \frac{d\tau_{xy}}{dy}, \tag{2}$$

where $\tau_{xy}=\tau$ is the shear stress in the transverse direction. In the present work, we use the Herschel–Bulkley model of viscoplasticity to express the shear stress (Eq. (1)). It then follows from Eq. (2) that the yield line, y_0, which separates yielded from unyielded regions, is obtained when the shear stress τ is equal to the yield stress τ_y, i.e.,

$$\left|\frac{dP}{dx}\right| = \frac{\tau_y}{y_0}. \tag{3}$$

For the integration of Eq. (2), boundary conditions (BCs) should be applied. For symmetric calendering we have [2],

$$\text{(BC-1)} \quad \tau_{xy} = 0 \text{ at } y = 0. \tag{3a}$$

Also, for the case of calendering with slip at the roll surface [5],

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received September 10, 2004; final manuscript received July 25, 2005. Review conducted by A. Maniatty. 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.

$$\text{(BC-2)}\quad u = U - \beta\tau_w^a, \quad \alpha \neq 0, \text{ at } y = h(x), \tag{3b}$$

where U is the roll speed, τ_w is the wall shear stress, β is the slip coefficient, and α is a slip parameter. Note that the limit $\beta \to 0$ corresponds to the case of no slip, while $\beta \to \infty$ gives perfect slip, thus vanishing wall shear stress. The case of the slip exponent $\alpha = 1$ corresponds to a linear slip law, while $\alpha = 2$ corresponds to a quadratic slip law.

After the appropriate manipulations, we obtain:

$$u = \begin{cases} U - \beta\tau_w^a + \dfrac{1}{\dfrac{(n+1)}{n}\dfrac{dP}{dx}K^{1/n}}\left[\left(y\left(\dfrac{dP}{dx}\right) \mp \tau_y\right)^{n+1/n} - \left(h\left(\dfrac{dP}{dx}\right) \mp \tau_y\right)^{n+1/n}\right], & y > y_0, \\ U - \beta\tau_w^a + \dfrac{1}{\dfrac{(n+1)}{n}\dfrac{dP}{dx}K^{1/n}}\left[\left(y_0\left(\dfrac{dP}{dx}\right) \mp \tau_y\right)^{n+1/n} - \left(h\left(\dfrac{dP}{dx}\right) \mp \tau_y\right)^{n+1/n}\right], & y \leqslant y_0. \end{cases} \tag{4}$$

Equation (4) incorporates both cases: The minus sign is valid in that part of the domain where the pressure gradient is positive ($dP/dx > 0$), while the plus sign is valid where the pressure gradient is negative ($dP/dx < 0$). These regions can only be determined after an expression for the pressure gradient is found.

Integration of the velocity profile gives the volumetric flow rate Q (noting that $\tau_y = y_0|dP/dx|$). In the case of the Bingham plastic model ($n = 1$), the result is

$$Q = 2\int_0^h u\,dy = 2(U - \beta\tau_w^\alpha) - \frac{(2h + y_0)}{3}\left(\frac{1}{\mu}\frac{dP}{dx}\right)(h - y_0)^2, \tag{5}$$

while for the general case of the Herschel–Bulkley fluids, the result is

$$Q = 2h(U - \beta\tau_w^a) - \frac{2n[(n+1)h + ny_0]}{(n+1)(2n+1)}\left(\frac{1}{K}\frac{dP}{dx}\right)^{1/n}(h - y_0)^{n+1/n}, \quad \text{for } dP/dx > 0, \tag{6a}$$

$$Q = 2h(U - \beta\tau_w^\alpha) + \frac{2n[(n+1)h + ny_0]}{(n+1)(2n+1)}\left(-\frac{1}{K}\frac{dP}{dx}\right)^{1/n}(h - y_0)^{n+1/n}, \quad \text{for } dP/dx < 0. \tag{6b}$$

The pressure gradient dP/dx is obtained from Eq. (6).

2.2 Dimensionless Variables. In calendering, the following dimensionless parameters are introduced [2,4]:

$$x' = \frac{x}{\sqrt{2RH_0}}, \quad y' = \frac{y}{H_0}, \quad P' = \frac{P}{K}\left(\frac{H_0}{U}\right)^n,$$

$$\lambda^2 = \frac{Q}{2UH_0} - 1, \quad h = H_0\left(1 + \frac{x^2}{2RH_0}\right), \quad A = \sqrt{\frac{2R}{H_0}}, \tag{7}$$

where λ is a dimensionless flow rate (or *leave-off distance*) and the rest of the symbols are defined in Fig. 1. Note that A is a geometric parameter.

Because of viscoplasticity, some extra dimensionless variables are needed. A dimensionless apparent yield stress can be defined by:

$$\tau_y^* = \frac{\tau_y}{\left|\dfrac{dP}{dx}\right|_{H_0}} = \frac{y_0}{H_0} = y_0', \tag{8}$$

Note that because dP/dx is a function of x, so is the yield line y_0. So, in order to have an unequivocal definition of τ_y^*, its value is taken at the nip ($x = 0$), i.e.,

$$\tau_y^*|_{x=0} = \frac{\tau_y}{\left|\dfrac{dP}{dx}\right|_{x=0} H_0} = \frac{y_{0,0}}{H_0} = y_{0,0}', \tag{9}$$

which means that the unyielded line goes from 0 (viscous fluid) to 1 (unyielded solid) *at the nip*.

Also, in viscoplasticity, the well-known Bingham number is defined as

$$\text{Bn} = \frac{\tau_y}{K}\left(\frac{H_0}{U}\right)^n. \tag{10}$$

It then follows that:

$$y_0' = \frac{\text{Bn}}{\left|\dfrac{dP'}{dx'}\right|}A. \tag{11}$$

In all cases, the purely viscous fluid corresponds to $\tau_y^* = 0$ and Bn = 0. However, at the other extreme of an unyielded solid,

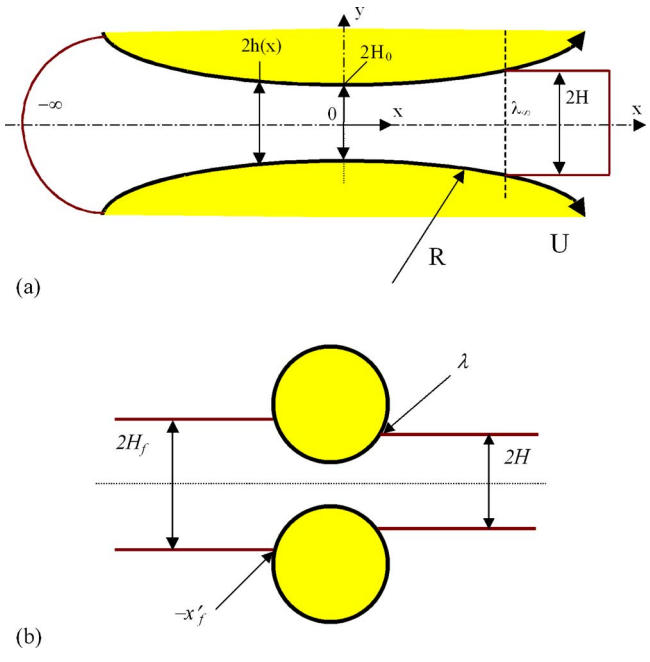


Fig. 1 Schematic representation of the calendering process and definition of variables: (a) Feed from an infinite reservoir and (b) feed with a finite sheet

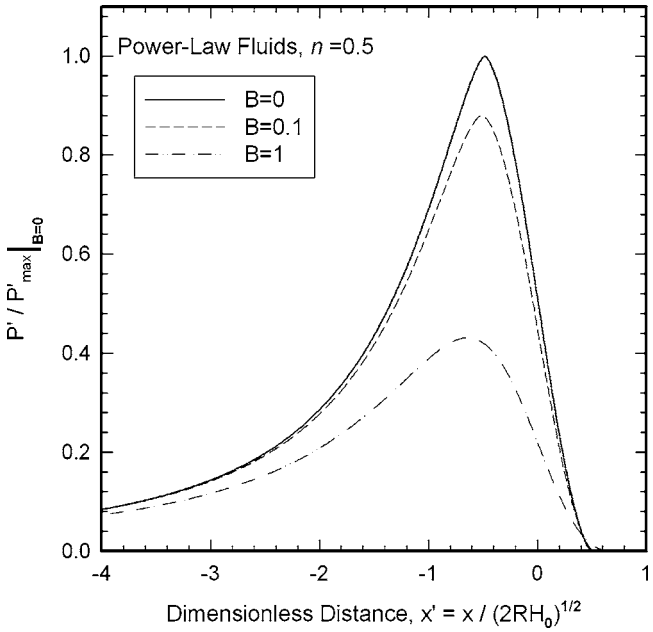


Fig. 2 Pressure distribution for power-law fluids with n=0.5, for various values of the slip coefficient, B

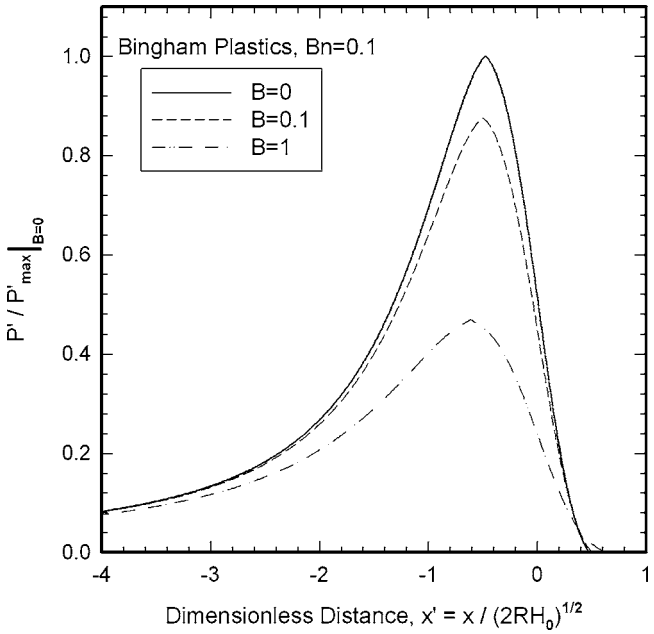


Fig. 4 Pressure distribution for Bingham plastics with Bn =0.1, for various values of the slip coefficient, B

$\tau_y^*|_{x=0} \to 1$ and $\mathrm{Bn} \to \infty$.

Because of the slip law, it is also necessary to consider a dimensionless expression for the slip coefficient, β [6]:

$$B = \frac{\beta}{U}\left(K\left(\frac{U}{H_0}\right)^n\right)^\alpha. \tag{12}$$

After the appropriate manipulations, we obtain the following equation for the dimensionless pressure gradient:

$$(\lambda^2 - x'^2) = \frac{dP'}{dx'}\left\{-\frac{B(1+x'^2)^{(\alpha+1)}}{A^\alpha}\left|\frac{dP'}{dx'}\right|^{\alpha-1} - \frac{n}{2n+1}\frac{1}{A^{1/n}}\left|\frac{dP'}{dx'}\right|^{1/n-1}\left(1 + x'^2 + \frac{n}{n+1}y_0'\right)(1+x'^2 - y_0')^{n+1/n}\right\}. \tag{13}$$

The above equation can be written in the following form:

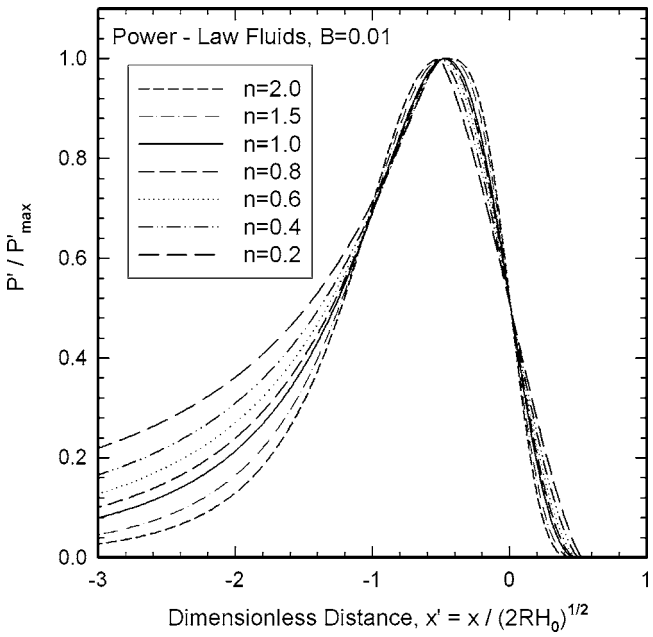


Fig. 3 Pressure distribution for power-law fluids with a slip coefficient B=0.01, for various values of the power-law index, n

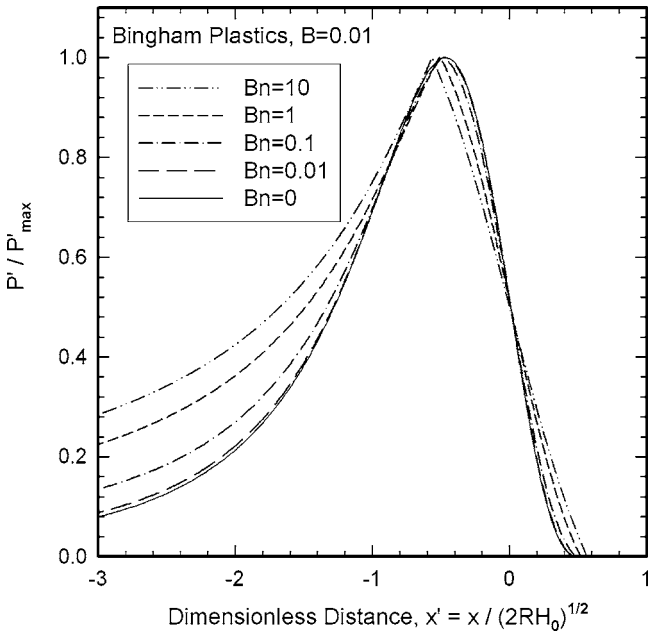


Fig. 5 Pressure distribution for Bingham plastics with a slip coefficient B=0.01, for various values of the Bingham number, Bn

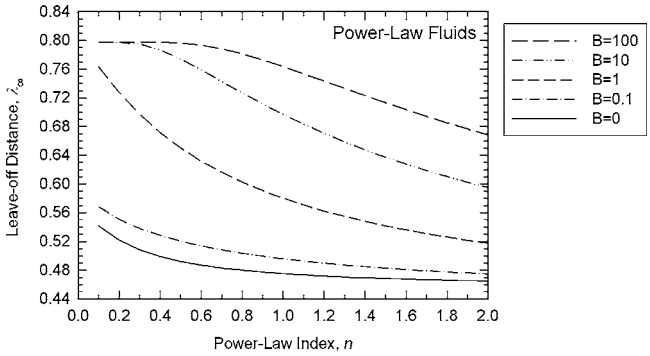


Fig. 6 Dimensionless leave-off distance, λ_∞, for pseudoplastic fluids as a function of the power-law index, *n*, for various values of the slip coefficient, *B*

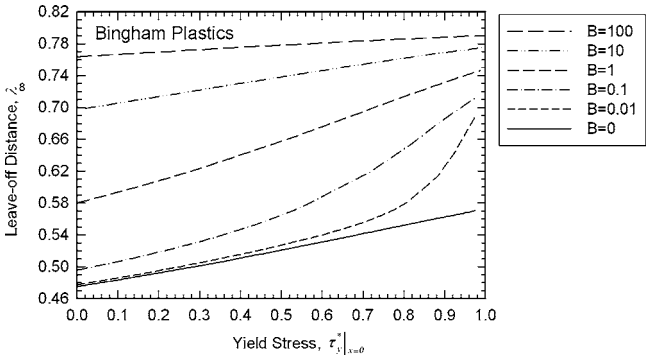


Fig. 7 Dimensionless leave-off distance, λ_∞, for viscoplastic fluids as a function of the dimensionless yield stress calculated at the nip, $\tau_y^*|_{x=0}$, for various values of the slip coefficient, *B*

$$\frac{dP'}{dx'} = A\left(\frac{2n+1}{n}\right)^n \frac{|(x'^2-\lambda^2)-B'|^{n-1}}{\left[(1+x'^2)+\dfrac{n}{n+1}y_0'\right]^n} \frac{[(x'^2-\lambda^2)-B']}{(1+x'^2-y_0')^{n+1}}, \tag{14}$$

$$\text{where} \quad B' = B(1+x'^2)^{(\alpha+1)}\left(\frac{1}{A}\frac{dP'}{dx'}\right)^{\alpha}. \tag{15}$$

From the analysis conducted for Newtonian and power-law fluids [2], we know that for $-\infty < x' < -\lambda$, the pressure gradient is positive and the minus sign in Eq. (4) is valid, while for $-\lambda < x' < \lambda$ the pressure gradient is negative and the plus sign in Eq. (4) is valid. Equation (14) suggests that these distinct regions could be altered in the presence of slip, as more roots are feasible due to the complexity of the equation.

For the no-slip case, for which $B=B'=0$, the following remarks can be made. When $y_0'=0$, the above equation is simplified to the one given by Middleman [2] for the power-law fluids, while for $y_0'=0$ and $n=1$, it is reduced to the one for Newtonian fluids. For $n=1$ and $y_0' \neq 0$, the above equation can be used for the Bingham plastics.

Integration of Eq. (14) will provide the pressure distribution in the domain. For this, boundary conditions are necessary for the pressure and its gradient. For the case of calendering from an infinite reservoir these are [2]:

$$P' = \frac{dP'}{dx'} = 0, \quad \text{at } x' = \lambda_\infty. \tag{16a}$$

$$P' = 0, \quad \text{at } x' \to -\infty. \tag{16b}$$

Then, the pressure is obtained from the integral:

$$P' = A\left(\frac{2n+1}{n}\right)^n \int_{x'}^{\lambda_\infty} \frac{|(x'^2-\lambda^2)-B'|^{n-1}}{\left[(1+x'^2)+\dfrac{n}{n+1}y_0'\right]^n} \frac{[B'-(x'^2-\lambda^2)]}{(1+x'^2-y_0')^{n+1}} dx'. \tag{17}$$

The dimensionless leave-off distance, λ_∞, corresponding to an infinite reservoir, can be found from the above equation, noting that $P'(x' \to -\infty)=0$. Therefore, λ_∞ can be found from the relation:

$$0 = \int_{-\infty}^{\lambda_\infty} \frac{|(x'^2-\lambda^2)-B'|^{n-1}}{\left[(1+x'^2)+\dfrac{n}{n+1}y_0'\right]^n} \frac{[B'-(x'^2-\lambda^2)]}{(1+x'^2-y_0')^{n+1}} dx'. \tag{18}$$

The yield line, y_0', appearing in the above equations is a variable that can be expressed either as a function of kinematic parameters (i.e., roll speed) or dynamic parameters (i.e., pressure gradient). The expressions for the yield line are thus either a function of the Bingham number, Bn, Eq. (11), or the dimensionless yield stress, $\tau_y^*|_{x=0}$, by combining Eqs. (8) and (9):

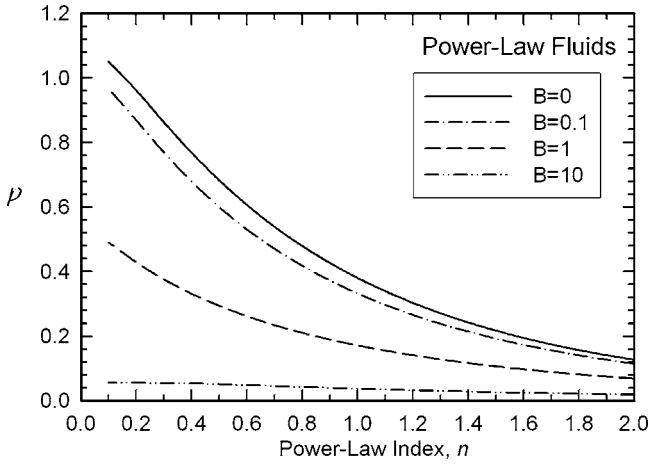
$$y_0' = y_{0,0}' \frac{\left|\dfrac{dP'}{dx'}\right|_{x=0}}{\left|\dfrac{dP'}{dx'}\right|} = \tau_y^*|_{x=0} \frac{\left|\dfrac{dP'}{dx'}\right|_{x=0}}{\left|\dfrac{dP'}{dx'}\right|}. \tag{19}$$

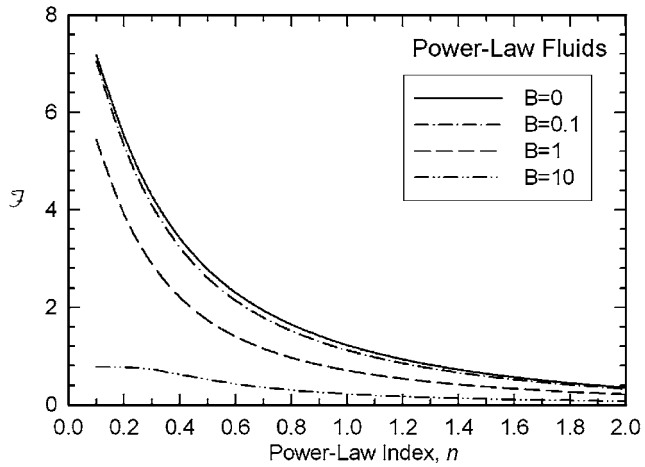
Table 1 Dimensionless leave-off distance, λ_∞, for power-law fluids, and for various values of the slip coefficient, *B*

n	$B=0$	$B=10^{-2}$	$B=10^{-1}$	$B=10^{0}$	$B=10^{1}$	$B=10^{2}$
0.1	0.5414	0.5440	0.5680	0.7632	0.7973	0.7973
0.2	0.5217	0.5246	0.5507	0.7276	0.7973	0.7973
0.3	0.5085	0.5115	0.5380	0.6972	0.7949	0.7973
0.4	0.4991	0.5022	0.5283	0.6717	0.7867	0.7972
0.5	0.4923	0.4953	0.5204	0.6502	0.7740	0.7964
0.6	0.4871	0.4900	0.5138	0.6320	0.7589	0.7935
0.7	0.4831	0.4858	0.5083	0.6163	0.7429	0.7884
0.8	0.4799	0.4825	0.5036	0.6027	0.7270	0.7813
0.9	0.4773	0.4798	0.4995	0.5907	0.7117	0.7728
1	0.4751	0.4775	0.4958	0.5802	0.6971	0.7634
1.1	0.4733	0.4755	0.4926	0.5708	0.6834	0.7535
1.2	0.4718	0.4739	0.4897	0.5624	0.6706	0.7434
1.3	0.4705	0.4724	0.4872	0.5548	0.6587	0.7332
1.4	0.4694	0.4712	0.4849	0.5480	0.6476	0.7231
1.5	0.4685	0.4701	0.4828	0.5417	0.6373	0.7132
1.6	0.4676	0.4691	0.4809	0.5360	0.6277	0.7036
1.7	0.4668	0.4682	0.4791	0.5308	0.6187	0.6943
1.8	0.4662	0.4675	0.4775	0.5260	0.6103	0.6853
1.9	0.4656	0.4667	0.4761	0.5215	0.6025	0.6766
2	0.4650	0.4661	0.4748	0.5174	0.5952	0.6683

Table 2 Dimensionless leave-off distance, λ_∞, for Bingham plastics with Bn=0.005, and for various values of the slip coefficient, *B*

B	λ_∞	$\tau_y^*\|_{x=0}$
0	0.4757	0.0073
10^{-2}	0.4781	0.0074
10^{-1}	0.4967	0.0087
10^{0}	0.5827	0.0195
10^{1}	0.7056	0.1033
10^{2}	0.7861	0.8092





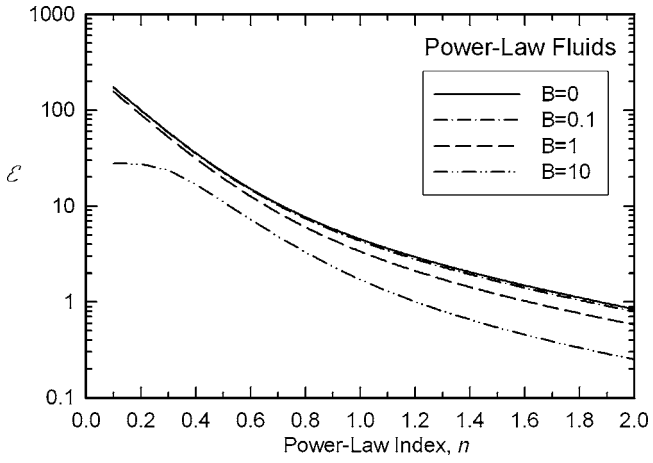
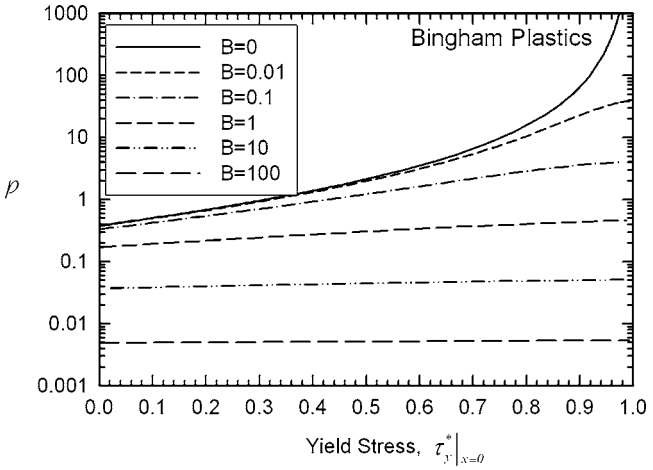
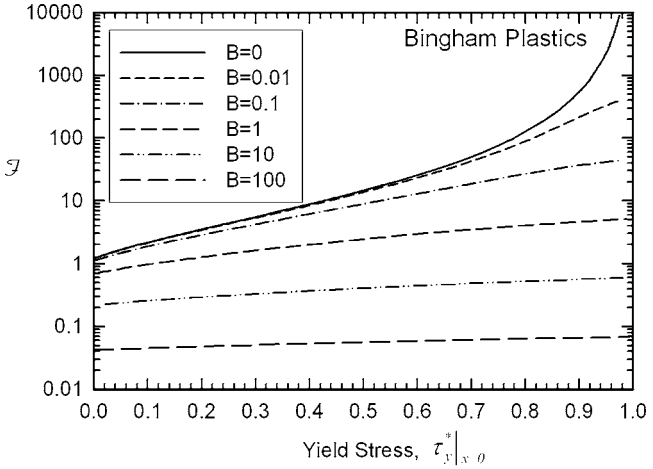


Fig. 8 Operating variables for pseudoplastic fluids as a function of the power-law index, *n*, for various values of the slip coefficient, *B*: (a) maximum pressure, *P*, (b) force factor, *F*, and (c) power factor, *E*





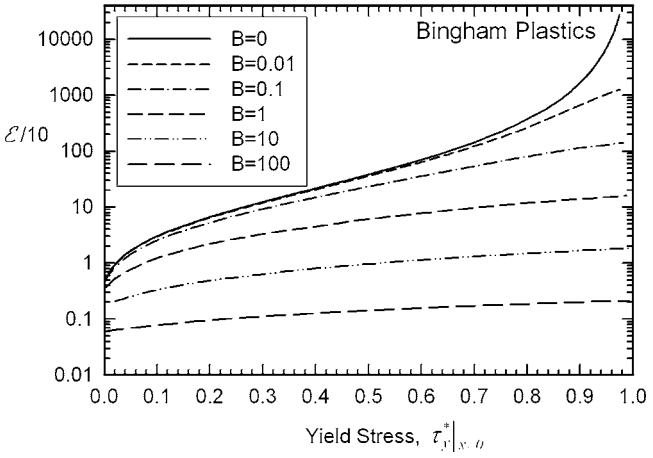


Fig. 9 Operating variables for Bingham plastics as a function of the dimensionless yield stress calculated at the nip, $\tau_y^*|_{x=0}$, for various values of the slip coefficient *B*: (a) maximum pressure, *P*, (b) force factor, *F*, and (c) power factor, *E*

3 Method of Solution

Considering Eq. (11), Eq. (18) can be written as

$$0 = \int_{-\infty}^{\lambda_\infty} I(n, \mathrm{Bn}, B)\, dx' . \tag{20}$$

The above integral has no analytical solution for the general case of Herschel–Bulkley fluids. Therefore, a numerical solution must be found based on some numerical algorithm.

The input data are the entry point, x_f' (which is set to -100 for the infinite case), the material rheological data, i.e., n, Bn, the slip parameters, α, B, and the number of x intervals, $N=100{,}000$. The large number of intervals is justified due to the long domain ($x_f' = -100$) and the highly nonlinear behavior of Eq. (13) at the limits

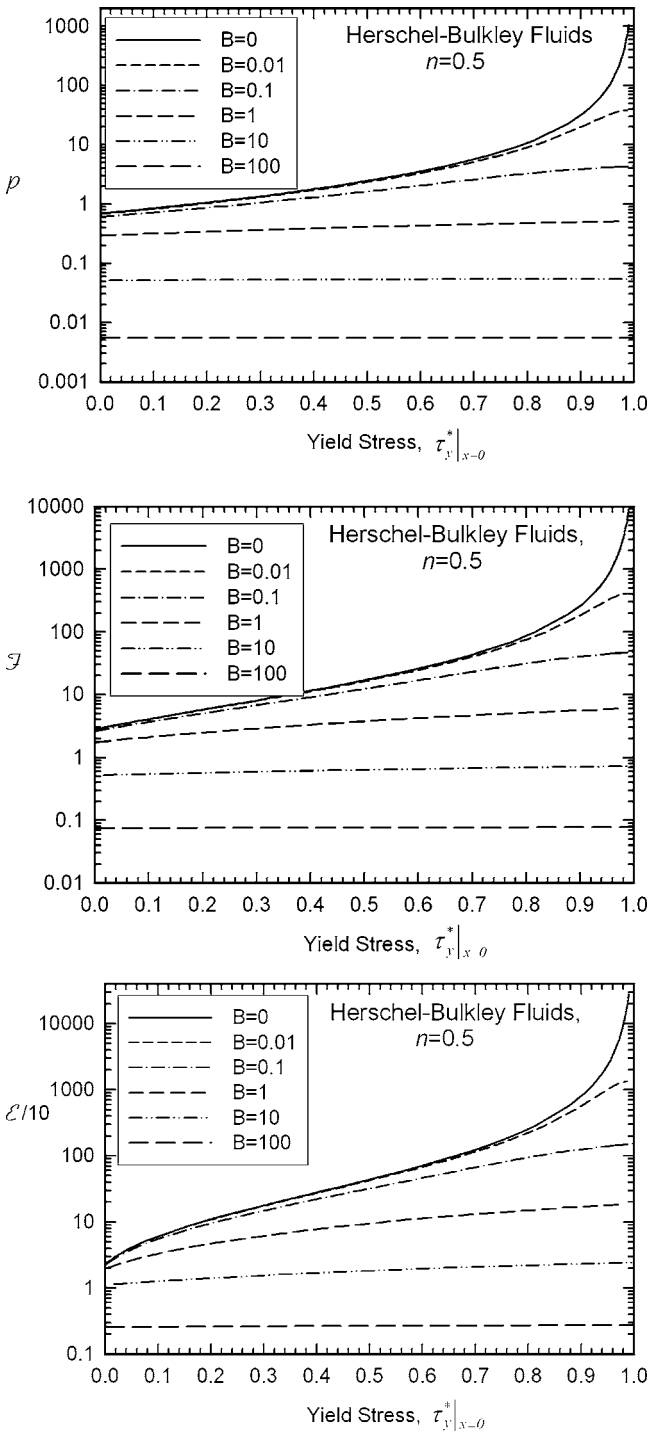


Fig. 10 Operating variables for Herschel–Bulkley fluids with $n=0.5$ as a function of the dimensionless yield stress calculated at the nip, $\tau_y^*|_{x=0}$, for various values of the slip coefficient, B: (a) maximum pressure, P, (b) force factor, F, and (c) power factor, E

of non-Newtonian behavior. The *modified regula falsi* numerical method (or *linear interpolation method of false position*, a variant of the bisection method) [7] is used to solve the governing equation and acquire the pressure-gradient distribution. Simpson's rule is used to compute the above integral, while the *modified regula falsi* method is used again in order to find the value of λ_∞ for which the pressure vanishes. Note that the yield line can be computed from Eq. (19), using the acquired pressure-gradient values.

Before proceeding with the calculations, the following observation should be made. By combining the expressions for the Bingham number (Bn) and the slip parameter (B)—Eqs. (10) and (12), respectively—we find:

$$B = \frac{\beta}{U}\left(\frac{\tau_y}{\mathrm{Bn}}\right)^{\alpha}. \tag{21}$$

The above equation suggests that the Bingham number and the dimensionless slip coefficient are in fact correlated. In order to determine which combinations of values for the two parameters correspond to feasible solutions, we examine the location of the yield line at the nip, which should be $\leqslant 1$. For given values of the slip parameter, B, the calculations start with Bn=0, for which the yield line $\tau_y^*|_{x=0}=0$, and end at some value of Bn, for which $\tau_y^*|_{x=0}\approx 1$. In this way, the acceptable (Bn, B) values are the ones for which $0\leqslant \tau_y^*|_{x=0}<1$.

Once λ_∞ is found as a function of n, Bn, and B, then all other quantities of interest are readily available. The exiting sheet thickness H is given by:

$$\frac{H}{H_0} = 1 + \lambda_\infty^2. \tag{22}$$

4 Operating Variables

The operating variables used in engineering calculations are also of interest [2]:

(1) The maximum pressure, $\mathbf{P}\,(n,\mathrm{Bn},B)$, defined by:

$$\mathbf{P}(n,\mathrm{Bn},B) = \frac{P'_{\max}}{A} = \left(\frac{2n+1}{n}\right)^n \int_{-\lambda_\infty}^{\lambda_\infty} I(n,\mathrm{Bn},B)\,dx', \tag{23}$$

(2) The roll-separating force per unit width W, $F/W(n,\mathrm{Bn},B)$, defined by:

$$\frac{F}{W} = \int_{-\infty}^{\lambda_\infty} P(x)\,dx = K\left(\frac{U}{H_0}\right)^n R\mathbf{F}(n,\mathrm{Bn},B), \tag{24a}$$

with $\mathbf{F}(n,\mathrm{Bn},B)$

$$= 2\left(\frac{2n+1}{n}\right)^n \int_{-\infty}^{\lambda_\infty}\left[\int_{x'}^{\lambda_\infty} I(n,\mathrm{Bn},B)\,dx'\right]dx', \tag{24b}$$

(3) The power input for both rolls, $\dot{\mathbf{W}}(n,\mathrm{Bn},B)$, defined by:

$$\dot{\mathbf{W}}(n,\mathrm{Bn},B) = 2WU\int_{-\infty}^{\lambda_\infty} \tau_{xy}|_{y=h(x)}\,dx$$

$$= WU^2K\left(\frac{U}{H_0}\right)^{n-1}\sqrt{\frac{R}{H_0}}\,\mathbf{E}(n,\mathrm{Bn},B), \tag{25a}$$

$$\text{with } \mathbf{E}(n,\mathrm{Bn},B) = -2\sqrt{2}\left(\frac{2n+1}{n}\right)^n \int_{-\infty}^{\lambda_\infty} I(n,\mathrm{Bn},B)(1 + x'^2)\,dx'. \tag{25b}$$

5 Results and Discussion

First, the calculations are pursued for pseudoplastic fluids, both shear thinning ($0<n\leqslant 1$) and shear thickening ($1\leqslant n\leqslant 2$). Next, the calculations are pursued for viscoplastic fluids, including Bingham plastics and Herschel–Bulkley fluids. In both cases, the

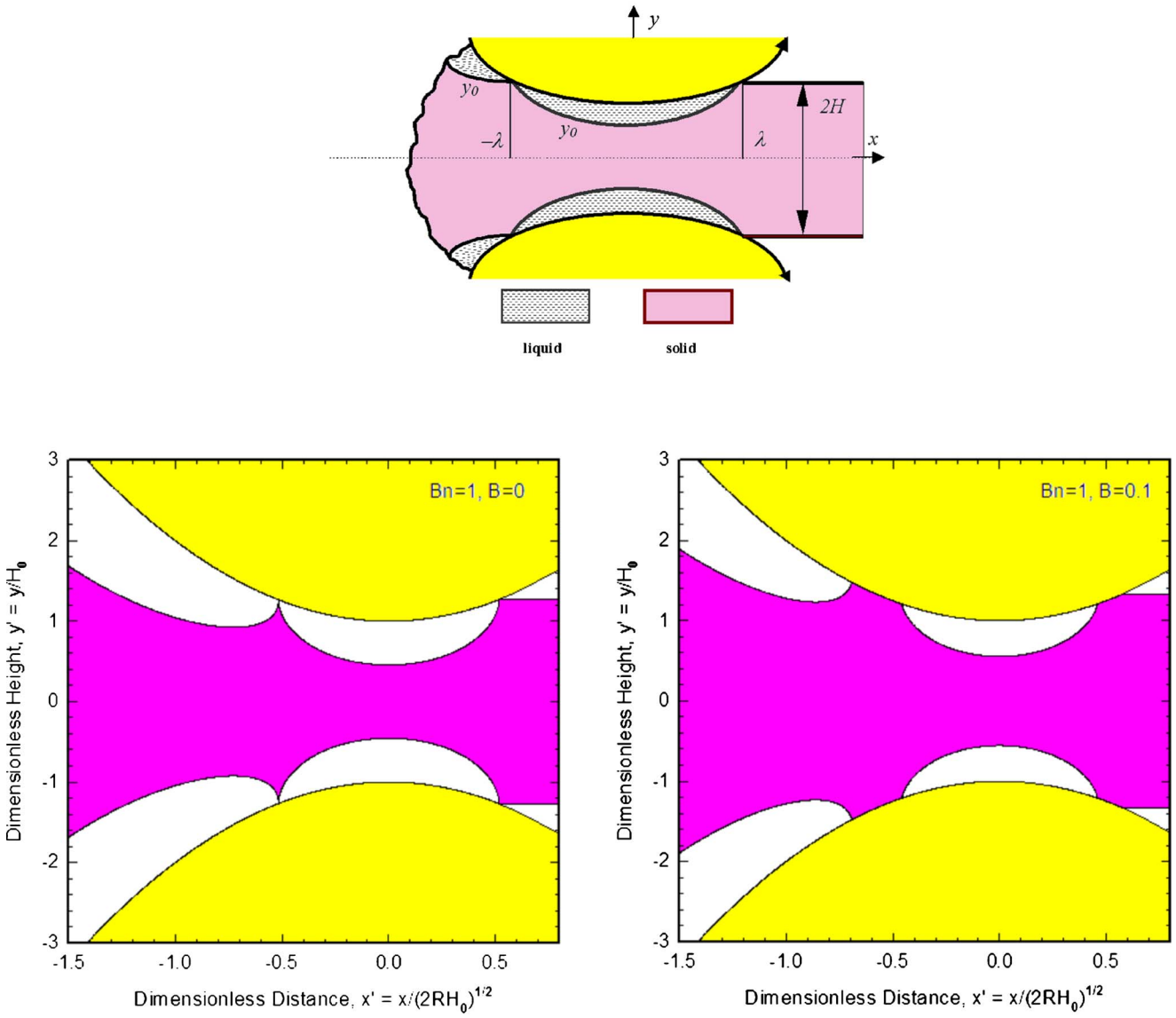


Fig. 11 Yielded and unyielded regions in calendering according to LAT: (a) qualitatively predicted by Gaskell [1], (b) calculated for a Bingham plastic without slip (Bn=1 and B=0), (c) calculated for a Bingham plastic with slip (Bn=1 and B=0.1)

impact of slip conditions at the wall is examined. All results presented here are for a linear slip law (α=1) for ease of presentation.

5.1 Pressure Distribution. We begin presenting our results by showing the pressure distributions obtained for pseudoplastic power-law fluids for different values of the power-law index n and the slip coefficient B.

In Fig. 2, we plot the dimensionless pressure distribution along the domain for a shear-thinning fluid with n=0.5 and for different values of B. The case of B=0 corresponds to no-slip conditions. We observe that slip reduces the pressure curves, as expected, but increases the domain (higher sheet thickness at exit). For the case of B=1 (macroscopically obvious slip), the maximum pressure has been reduced to about 40% of the pressure for no-slip conditions.

In Fig. 3, we plot the dimensionless pressure distribution along the domain for a given slip coefficient B=0.01 and various power-law fluids. Decreasing the power-law index n results in a bigger domain, thus a higher sheet thickness at the exit, and in a more pointed shape of the pressure curve at the maximum.

Figure 4 shows the dimensionless pressure distribution along the domain for a Bingham plastic with Bn=0.1 and for different values of B. The presence of slip has the same impact as in the case of pseudoplastic fluids; i.e., it reduces the pressure curves and increases the domain (and, consequently, the exit sheet thickness). For the case of B=1 (macroscopically obvious slip), the maximum pressure is reduced by 41%, compared to the no-slip condition.

In Fig. 5, we plot the dimensionless pressure distribution for a given slip coefficient B=0.01, and for various Bingham plastics. The bigger the Bingham number Bn, the bigger the domain, and the more pointed the shape of the pressure curve at the maximum.

5.2 Leave-Off Distance and Final Sheet Thickness. The results are shown in Fig. 6 for the dimensionless leave-off distance λ_∞, while numerical values are given in Table 1. With regard to the no-slip case, the well-known Newtonian value (for n=1) of λ_∞=0.475 is a starting point, after which it is noted that shear-thinning increases this value (for n=0.1 by 14%), while shear-thickening decreases it (for n=2 by 2%). Moreover, it is noted that the impact of slip on the leave-off distance is smaller for bigger values of the power-law index. For the case of macroscopically obvious slip (B=1), the domain increases up to 41% for extreme

shear-thinning conditions ($n=0.1$), up to 22% for the Newtonian case, and up to 11% for extreme shear-thickening conditions ($n=2$).

Calculations were carried out for the whole range of values of the dimensionless yield stress at the nip, $\tau_y^*|_{x=0}$, i.e., $0 \leq \tau_y^*|_{x=0} < 1$. The results for the Bingham plastic model are shown in Fig. 7 as a function of $\tau_y^*|_{x=0}$, while Table 2 lists the numerical results for Bingham plastics with Bn=0.005. The well-known Newtonian value (for Bn= $\tau_y^*|_{x=0}=0$) of $\lambda_\infty=0.475$ is again a starting point for the no-slip case, after which it is noted that viscoplasticity increases the values, finally reaching the limit for fully plastic flow of no deformation, i.e., for $\tau_y^*|_{x=0} \approx 1$, $\lambda_\infty=0.571$. This corresponds to a sheet produced 1.33 times thicker than the nip. Increasing the slip coefficient B leads to bigger leave-off distance values. This effect is weaker for bigger values of the Bingham number, and the initial differentiation among the various slip cases is reduced for intense viscoplastic conditions.

5.3 Operating Variables. The results from the current calculations are shown in Fig. 8 for the power-law pseudoplastic fluids, in Fig. 9 for the Bingham plastics, and in Fig. 10 for Herschel–Bulkley fluids with $n=0.5$. The Newtonian values for a no-slip case are $P=0.3802$, $F=1.2234$, and $E=4.5162$, in excellent agreement with the values given by Middleman [2]. Further decreasing the power-law index, n, or increasing the dimensionless yield stress at the nip, $\tau_y^*|_{x=0}$, increases these dimensionless quantities for the pseudoplastic and viscoplastic materials. However, the maximum values acquired for the Bingham and the Herschel–Bulkley materials are much bigger.

The presence of slip decreases these engineering quantities, calculated in a dimensionless form, for both pseudoplastic and viscoplastic materials. With regard to pseudoplastic fluids, and for the case of macroscopically obvious slip ($B=1$), the Newtonian values mentioned above for P, F, and E, decrease by 55%, 43%, and 26%, respectively. For the case of viscoplastic materials, the effect of slip is a little more pronounced for bigger values of the dimensionless yield stress calculated at the nip, $\tau_y^*|_{x=0}$. However, as the slip coefficient increases, it becomes dominant, and the effect of viscoplasticity is much smaller. The same conclusions can be drawn for the Herschel–Bulkley fluids, noting that the values acquired for Herschel–Bulkley fluids with $n=0.5$ are bigger, compared to those for Bingham plastics with $n=1$, due to the increase of the pseudoplastic character.

5.4 Yield-Line Location and Stress/Velocity Profiles. The yield lines separate the yielded areas from the unyielded ones, and serve as a contour of the stress, which is equal to the yield stress along these lines. The first *qualitative* approach of the shape of the yielded and unyielded regions was given by Gaskell [1], and is depicted in Fig. 11(*a*). After obtaining the pressure-gradient values for the domain, Eq. (11) or Eq. (19) can be used to draw these regions *quantitatively*. The calculated yield lines are given for a Bingham plastic with no slip (Bn=1 and $B=0$) in Fig. 11(*b*). Previous work carried out by the authors [8] showed the progressive growth of the unyielded regions, with increasing viscoplasticity. The present approach confirms Gaskell's suggestion that a plug profile should be expected at the points where $x'=\pm\lambda$. This, however, is only valid for the no-slip case. Figure 11(*c*) depicts the shape of the yielded and unyielded areas for a case where both viscoplasticity and slip conditions are encountered (Bn=1 and $B=0.1$). As discussed previously, the final sheet thickness is increased, and so is the dimensionless yield stress at the nip, $\tau_y^*|_{x=0}$. Furthermore, a plug profile is observed for a *region* around $x'=\pm\lambda$, as suggested by Eq. (14).

Readily available from the pressure-gradient results are the axial distributions of the shear stresses along the roll surface, according to Eq. (2). The effect of slip on the axial shear stress profile of a Bingham plastic is examined in Fig. 12. The shear stress is calculated at the wall (roll surface) and divided by the absolute value at the nip, for a Bingham plastic without slip (Bn=1 and $B=0$) and with slip (Bn=1 and $B=0.1$). The sign of the shear stress follows the pressure-gradient sign, being positive only for $-\infty < x' < -\lambda$. The presence of slip broadens the domain, as discussed earlier, adds more curvature, and results in a smooth transition from the positive to the negative values of the wall shear stress.

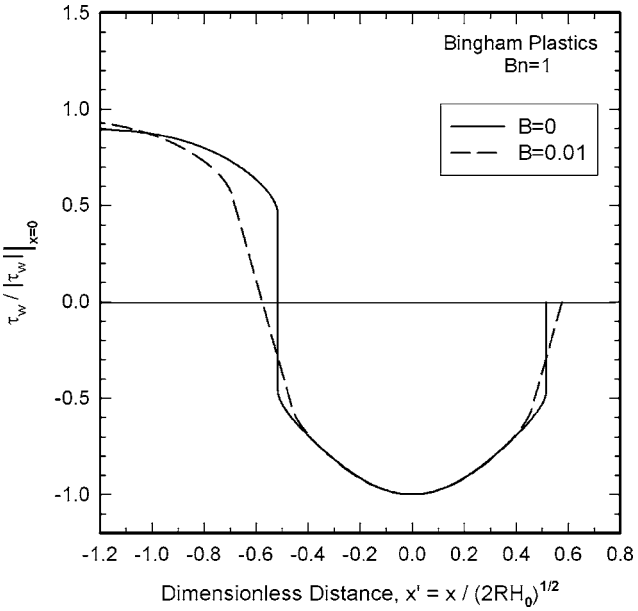


Fig. 12 Axial shear stress distributions at the roll surface for a Bingham plastic without slip (Bn=1 and $B=0$) and with slip (Bn=1 and $B=0.1$)

The increase in λ, the dimensionless flow rate or leave-off distance, may also be observed in Fig. 13, where the axial dimensionless velocity is depicted along the centerline from $-\lambda < x' < +\lambda$, for the same conditions. The maximum velocity value is

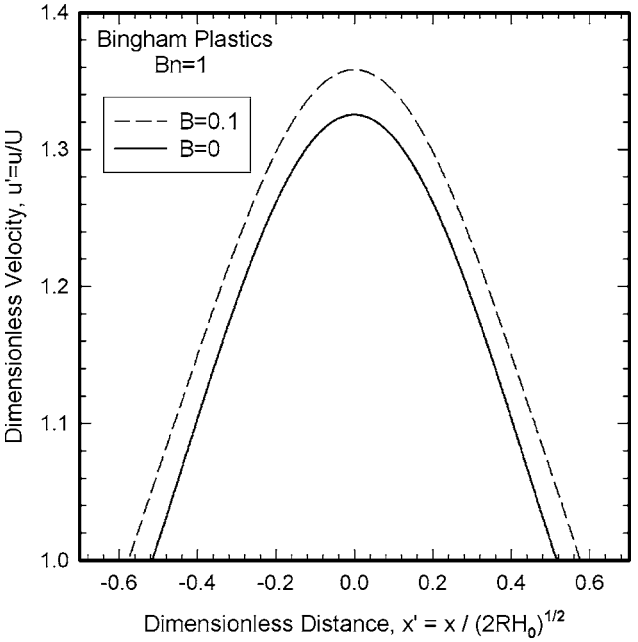


Fig. 13 Axial velocity distribution along the centerline for a Bingham plastic without slip (Bn=1 and $B=0$) and with slip (Bn=1 and $B=0.1$) in the region $-\lambda < x' < +\lambda$

observed at the nip. At that point, the velocity of the Bingham plastic without slip is 1.33 times bigger than the roll speed. The presence of slip has no impact on the bell shape of the curves, it does however increase the material velocity uniformly along the domain, which explains the increase in λ, the dimensionless flow rate. With the presence of slip, the material velocity value at the center is 1.36 times bigger than the roll speed. It should be noted that the velocity beyond $-\lambda$ is further reduced and is not shown here due to the length of the domain (up to -100).

6 Critique on the Lubrication Approximation Theory

Both the mathematical analysis and results presented in this work are carried out by making use of the LAT. An important question is therefore raised, regarding the impact of this approximation on the quality of the results.

LAT regards locally fully developed flow and simplifies the conservation equations. Furthermore, LAT becomes even more questionable in the case where the presence of plasticity is induced by a curved boundary and it is interesting to investigate the "paradox" pointed out by Lipscomb and Denn [9], by which yielded regions appear where none should have. A closer look at the physics of the problem and at Fig. 11 reveals that the shaded areas cannot be rigid plugs, since the speeds at the entry and exit are different, and it is not therefore reasonable to accept that a constant speed occurs along the centerline. The use of LAT is therefore inadequate with regard to shaping the yielded and unyielded regions. A sample test-case calculation of the full two-dimensional (2D) problem showed a continuing paraboliclike development of the centerline velocity profile (such as the ones in Fig. 13) and the lack of any yielded/unyielded regions.

However, the same test calculation showed that the domain length, the pressure distribution, and the resulting operating variables were within 1% of the values found by LAT. This is also confirmed by previous work [10] regarding the calendering of pseudoplastic fluids, which has proved that this approach, compared to a full 2D analysis of the process, gives good results for the pressure distribution and hence for all the resulting integrated quantities.

7 Conclusions

The lubrication approximation has been used to derive numerical solutions for the dimensionless leave-off distance and common engineering quantities, for the case of calendering pseudoplastic and viscoplastic fluids from an infinite reservoir. The inclusion of a slip boundary condition has been the major thrust of the present work, as it is known that many calendered materials exhibit slip at the wall.

In the case of power-law fluids, it was found that shear thinning increases the final sheet thickness, while shear thickening reduces it. In the case of viscoplastic materials, the increase in the dimensionless yield stress also leads to increased values for the leave-off distance (and consequently for the sheet thickness). The presence of slip increases sheet thickness for both cases. Operating variables, such as maximum pressure, roll-separating force, and power input to both rolls, all increase—in dimensionless form—with increasing pseudoplasticity (*shear thinning*) and viscoplasticity, and decrease under slip conditions.

It has been argued that the use of LAT is adequate with regard to obtaining such quantities as the ones presented here, although the prediction for the yielded and unyielded regions is erroneous. The present work is a prelude for more elaborate results to be obtained from a fully 2D analysis of the process, as done earlier for Newtonian and power-law fluids (see, e.g., [10]). Such work is currently under way by the authors.

Acknowledgment

Financial assistance from NTUA is gratefully acknowledged.

References

[1] Gaskell, R. E., 1950, "The Calendering of Plastic Materials," ASME J. Appl. Mech., **17**, pp. 334–337.

[2] Middleman, S., 1977, *Fundamentals of Polymer Processing*, McGraw-Hill, New York.

[3] Bird, R. B., Dai, G. C., Yarusso, B. J., 1983, "The Rheology and Flow of Viscoplastic Materials," Rev. Chem. Eng., **1**, pp. 1–70.

[4] Sofou, S., and Mitsoulis, E., 2004, "Calendering of Pseudoplastic and Viscoplastic Fluids Using the Lubrication Approximation," J. Polym. Eng., **24**, pp. 505–522.

[5] Vlachopoulos, J., and Hrymak, A. N., 1980, "Calendering Poly(Vinyl Chloride): Theory and Experiments," Polym. Eng. Sci., **20**, pp. 725–731.

[6] Mitsoulis, E., 1986, "Finite Elements Analysis of Wire Coating," Polym. Eng. Sci., **26**, pp. 171–186.

[7] Gerald, C. F., and Wheatley, P. O., 1994, *Applied Numerical Analysis*, 5th Ed., Addison-Wesley, Reading, MA.

[8] Sofou, S., and Mitsoulis, E., 2004, "Calendering Pseudoplastic and Viscoplastic Sheets of Finite Thickness," J. Plast. Film Sheeting, **20**, pp. 185–222.

[9] Lipscomb, G. G., and Denn, M. M., 1984, "Flow of Bingham Fluids in Complex Geometries," J. Non-Newtonian Fluid Mech., **14**, pp. 337–346.

[10] Mitsoulis, E., Vlachopoulos, J., and Mirza, F. A., 1985, "Calendering Analysis Without the Lubrication Approximation," Polym. Eng. Sci., **25**, pp. 6–18.

Prediction of Falling Cylinder Through Air-Water-Sediment Columns

Peter C. Chu[1]
e-mail: chu@nps.navy.mil

Chenwu Fan

Naval Ocean Analysis and Prediction Laboratory,
Department of Oceanography,
Naval Postgraduate School,
Monterey, CA 93943

A falling rigid body through air, water, and sediment is investigated experimentally and theoretically. Two experiments were conducted to drop rigid cylinders with density ratio around 1.8 into shallow water (around 13 m deep) in the Monterey Bay (Exp-1) and into the Naval Postgraduate School's swimming pool (Exp-2). During the experiments, we carefully observe cylinder track and burial depth while simultaneously taking gravity cores (in Exp-1). After analyzing the gravity cores, we obtain the bottom sediment density and shear strength profiles. The theoretical work includes the development of a 3D rigid body impact burial prediction model (IMPACT35) that contains three components: triple coordinate transform and hydrodynamics of a falling rigid object in a single medium (air, water, or sediment) and in multiple media (air-water and water-sediment interfaces). The model predicts the rigid body's trajectory in the water column and burial depth and orientation in the sediment. The experimental data (burial depth, sediment density, and shear strength) show the capability of IMPACT35 in predicting the cylinder's trajectory and orientation in a water column and burial depth and orientation in sediment.
[DOI: 10.1115/1.2125975]

1 Introduction

Study on the movement of a rigid body in fluid has wide scientific significance and technical application. The scientific studies of the hydrodynamics of a rigid cylinder in fluid involve the nonlinear dynamics, flight theory, body-fluid interaction, and instability theory. The body forces include the gravity and the buoyancy force. The hydrodynamic forces include the drag and lift forces that depend on the fluid-to-body velocity and the impact force as the body penetrates the air-water or water-sediment interfaces. Usually, a nonlinear dynamical system is needed to predict a falling rigid body in fluid, e.g., [1].

Recently, the scientific problem about rigid body movement in the air-water-sediment columns drew attention to the naval research. This is due to the threat of mines in the naval operations. Within the past 15 years three U.S. ships, the *USS Samuel B. Roberts* (FFG-58), *Tripoli* (LPH-10), and *Princeton* (CG-59) have fallen victim to mines. Total ship damage was $125 million while the mines cost approximately $30,000 [2]. Mines have evolved over the years from the dumb "horned" contact mines that damaged the *Tripoli* and *Roberts* to ones that are relatively sophisticated—nonmagnetic materials, irregular shapes, anechoic coatings, multiple sensors, and ship count routines. Despite their increased sophistication, mines remain inexpensive and are relatively easy to manufacture, keep, and place. Water mines are characterized by three factors [3,4]: position in water (bottom, moored, rising, and floating), method of delivery (aircraft, surface, and subsurface), and method of actuation (acoustic and/or magnetic influence, pressure, contact, and controlled). Accurate mine burial predictions are inherently difficult to make because of uncertainties in both mine deployment conditions and the relevant environmental parameters [5]. The U.S. Navy developed operational models to predict the environmental parameters for mine burial prediction [6]. Recently, statistical methods such as the Monte Carlo method [7] and the expert system method [5] have been developed. These methods need a core-physical model for describing the movement of falling rigid body through air-water-sediment columns.

When the rigid body is cylindrical, this dynamical system can be simplified using three coordinate systems: earth-fixed coordinate (E-coordinate), cylinder's main-axis following coordinate (M-coordinate), and hydrodynamic force following coordinate (F-coordinate). The origin of both M- and F-coordinates is at the cylinder's center of mass (COM). The body forces and their moments are easily calculated using the E-coordinate system. The hydrodynamic forces and their moments are easily computed using the F-coordinate. The cylinder's moments of gyration are simply represented using the M-coordinate. Recently, Chu et al. [8] developed a recursive model to predict the cylinder's translation velocity and orientation in the water column (single phase) on the base of the triple coordinate transformation.

To extend the recursive model from single medium (water column) to multi-media (air, water, sediment), a falling cylinder through air-water and water-sediment interfaces (i.e., cylinder contacting with two media) should be particularly analyzed. The cylinder is decomposed into two parts with each one contacting one medium. For the air-water penetration, the cylinder is decomposed into air and water parts. For the water-sediment penetration, the cylinder is decomposed into water and sediment parts. The body forces (such as the buoyancy force) and surface forces (such as pressure and hydrodynamic force) are computed separately for the two parts. A fully three-dimensional model is developed for prediction of the translation velocity and orientation of falling rigid cylinder through air, water, and sediment. Theoretical model development and a cylinder drop experiment for the model evaluation are depicted in this paper.

The outline of this paper is as follows: Section 2 depicts the triple coordinate systems. Section 3 describes the dynamics for determining the cylinder's translation velocity and orientation. Section 4 presents the equivalent cylinder method for computing hydrodynamic forces and torques when the cylinder penetrates the air-water and water-sediment interfaces. Section 5 describes forces and torques in air and water. Section 6 describes the resis-

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received March 2, 2004; final manuscript received August 4, 2005. Review conducted by D. 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.

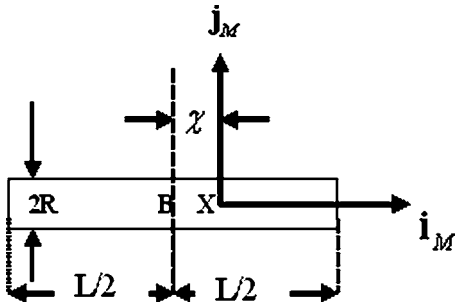


Fig. 1 M-coordinate with the COM as the origin X and (i_m, j_m) as the two axes. Here, χ is the distance between the COV (B) and COM (X), (L, R) are the cylinder's length and radius.

tance from sediments. Section 7 shows the model integration. Section 8 describes two cylinder drop experiments and observational data processing. Section 9 presents the model-data inter comparison. The conclusions are listed in Sec. 10.

2 Triple Coordinate Systems

Consider an axially symmetric cylinder with the centers of mass (COM) **X** [or called gravity center (GC) in literatures] and center of volume (COV) **B** on the main axis (Fig. 1). Let (L, R, χ) represent the cylinder's length, radius, and the distance between the two points $(\mathbf{X}, \mathbf{B})$. The positive χ values refer to nose-down case, i.e., the point **X** is lower than the point **B**. Three coordinate systems are used to model the falling cylinder through the air, water, and sediment phases: earth-fixed coordinate (E-coordinate), main-axis following coordinate (M-coordinate), and force following coordinate (F-coordinate) systems. All the systems are three-dimensional, orthogonal, and right-handed [8].

2.1 E-Coordinate. The E-coordinate is represented by $F_E(O, \mathbf{i}, \mathbf{j}, \mathbf{k})$ with the origin "O" and three axes: x, y axes (horizontal) with the unit vectors $(\mathbf{i}, \mathbf{j})$ and z axis (vertical) with the unit vector **k** (upward positive). The position of the cylinder is represented by the position of the COM,

$$\mathbf{X} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \tag{1}$$

which is translation of the cylinder. The translation velocity is given by

$$\frac{d\mathbf{X}}{dt} = \mathbf{V}, \quad \mathbf{V} = (u, v, w). \tag{2}$$

2.2 M-Coordinate. Let orientation of the cylinder's main axis (pointing downward) be given by $\mathbf{i}_M$. The angle between $\mathbf{i}_M$ and **k** is denoted by $\psi_2 + \pi/2$. Projection of the vector $\mathbf{i}_M$ onto the (x, y) plane creates angle (ψ_3) between the projection and the x axis (Fig. 2). The M-coordinate is represented by $F_M(\mathbf{X}, \mathbf{i}_M, \mathbf{j}_M, \mathbf{k}_M)$ with the origin "**X**", unit vectors $(\mathbf{i}_M, \mathbf{j}_M, \mathbf{k}_M)$, and coordinates (x_M, y_M, z_M). In the plane consisting of vectors $\mathbf{i}_M$ and **k** (passing through the point M), two new unit vectors $(\mathbf{j}_M, \mathbf{k}_M)$ are defined with $\mathbf{j}_M$ perpendicular to the $(\mathbf{i}_M, \mathbf{k})$ plane, and $\mathbf{k}_M$ perpendicular to $\mathbf{i}_M$ in the $(\mathbf{i}_M, \mathbf{k})$ plane. The unit vectors of the M-coordinate system are given by (Fig. 2)

$$\mathbf{j}_M = \mathbf{k}_M \times \mathbf{i}_M, \quad \mathbf{k}_M = \mathbf{i}_M \times \mathbf{j}_M. \tag{3}$$

The M-coordinate system is solely determined by orientation of the cylinder's main axis $\mathbf{i}_M$. Let the vector **P** be represented by ${}^E\mathbf{P}$ in the E-coordinate and by ${}^M\mathbf{P}$ in the M-coordinate, and let ${}^E_M\mathbf{R}$ be the rotation matrix from the M-coordinate to the E-coordinate,

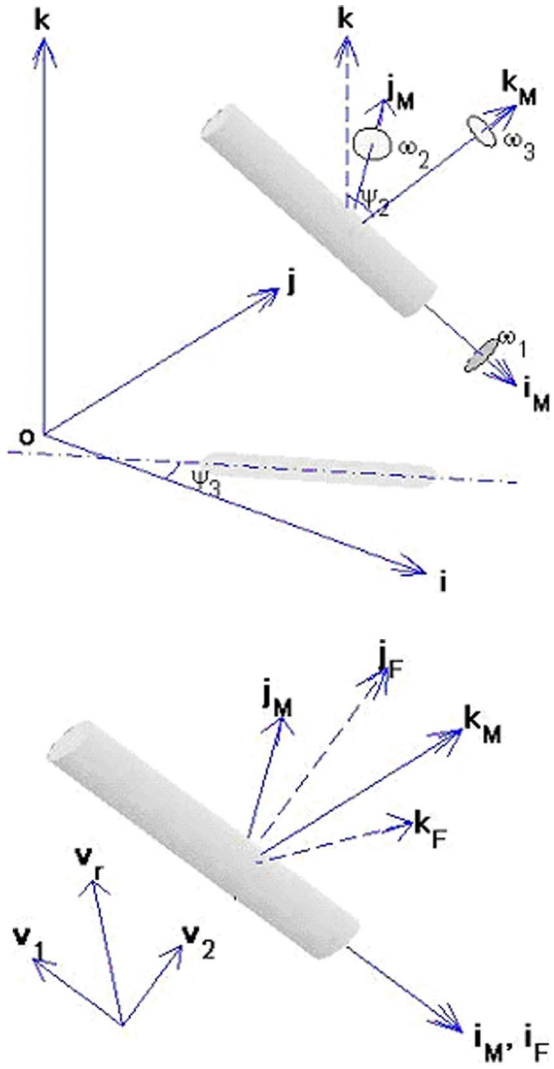


Fig. 2 Three coordinate systems

$$ {}^E_M\mathbf{R}(\psi_2, \psi_3) \equiv \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} \cos\psi_3 & -\sin\psi_3 & 0 \\ \sin\psi_3 & \cos\psi_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos\psi_2 & 0 & \sin\psi_2 \\ 0 & 1 & 0 \\ -\sin\psi_2 & 0 & \cos\psi_2 \end{bmatrix}, \tag{4}$$

which represents $(\mathbf{i}_M, \mathbf{j}_M, \mathbf{k}_M)$,

$$\mathbf{i}_M = \begin{bmatrix} r_{11} \\ r_{21} \\ r_{31} \end{bmatrix}, \quad \mathbf{j}_M = \begin{bmatrix} r_{12} \\ r_{22} \\ r_{32} \end{bmatrix}, \quad \mathbf{k}_M = \begin{bmatrix} r_{13} \\ r_{23} \\ r_{33} \end{bmatrix}. \tag{5}$$

Transformation of ${}^M\mathbf{P}$ into ${}^E\mathbf{P}$ contains rotation and translation,

$${}^E\mathbf{P} = {}^E_M\mathbf{R}(\psi_2, \psi_3)\,{}^M\mathbf{P} + \mathbf{X}. \tag{6}$$

Let the cylinder rotate around $(\mathbf{i}_M, \mathbf{j}_M, \mathbf{k}_M)$ with angles $(\varphi_1, \varphi_2, \varphi_3)$ (Fig. 2). The angular velocity of cylinder is calculated by

$$\omega_1 = \frac{d\varphi_1}{dt}, \quad \omega_2 = \frac{d\varphi_2}{dt}, \quad \omega_3 = \frac{d\varphi_3}{dt}, \tag{7}$$

and

$$\psi_1 = \varphi_1, \quad \frac{d\psi_2}{dt} = \frac{d\varphi_2}{dt} = \omega_2, \quad \frac{d\psi_3}{dt} \neq \frac{d\varphi_3}{dt}. \tag{8}$$

If $(\omega_1, \omega_2, \omega_3)$ are given, time integration of (7) with the time interval Δt leads to

$$\Delta\varphi_1 = \omega_1 \Delta t, \quad \Delta\varphi_2 = \omega_2 \Delta t, \quad \Delta\varphi_3 = \omega_3 \Delta t. \tag{9}$$

The increments $(\Delta\psi_2, \Delta\psi_3)$ are determined by the relationship between the two rotation matrices ${}^{E}_{M}R(\psi_2+\Delta\psi_2, \psi_3+\Delta\psi_3)$ and ${}^{E}_{M}R(\psi_2, \psi_3)$;

$$ {}^{E}_{M}R(\psi_2+\Delta\psi_2, \psi_3+\Delta\psi_3) = {}^{E}_{M}R(\psi_2, \psi_3) \begin{bmatrix} \cos(\Delta\varphi_3) & -\sin(\Delta\varphi_3) & 0 \\ \sin(\Delta\varphi_3) & \cos(\Delta\varphi_3) & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos(\Delta\varphi_2) & 0 & \sin(\Delta\varphi_2) \\ 0 & 1 & 0 \\ -\sin(\Delta\varphi_2) & 0 & \cos(\Delta\varphi_2) \end{bmatrix}. \tag{10}$$

2.3 F-Coordinate. The F-coordinate is represented by $F_F(\mathbf{X}, \mathbf{i}_F, \mathbf{j}_F, \mathbf{k}_F)$ with the origin $\mathbf{X}$, unit vectors $(\mathbf{i}_F, \mathbf{j}_F, \mathbf{k}_F)$, and coordinates (x_F, y_F, z_F). Let $\mathbf{V}_w$ be the fluid velocity. The fluid-to-cylinder velocity is represented by $\mathbf{V}_r = \mathbf{V}_w - \mathbf{V}$ that is decomposed into two parts,

$$\mathbf{V}_r = \mathbf{V}_1 + \mathbf{V}_2, \quad \mathbf{V}_1 = V_1 \mathbf{i}_F, \quad \mathbf{V}_2 = V_2 \mathbf{j}_F, \tag{11}$$

where

$$\mathbf{V}_1 = (\mathbf{V}_r \cdot \mathbf{i}_F)\mathbf{i}_F$$

is the component parallel to the cylinder's main axis (i.e., along $\mathbf{i}_M$), and

$$\mathbf{V}_2 = \mathbf{V}_r - (\mathbf{V}_r \cdot \mathbf{i}_F)\mathbf{i}_F$$

is the component perpendicular to the cylinder's main-axial direction. The unit vectors for the F-coordinate are defined by (column vectors)

$$\mathbf{i}_F = \mathbf{i}_M = \begin{bmatrix} r_{11} \\ r_{21} \\ r_{31} \end{bmatrix}, \quad \mathbf{j}_F = \mathbf{V}_2/|\mathbf{V}_2|, \quad \mathbf{k}_F = \mathbf{i}_F \times \mathbf{j}_F. \tag{12}$$

The F-coordinate system is solely determined by orientation of the cylinder's main-axis $(\mathbf{i}_M)$ and the water-to-cylinder velocity. Note that the M- and F-coordinate systems have one common unit vector $\mathbf{i}_M$ (orientation of the cylinder).

Let ${}^{E}_{F}\mathbf{R}$ be the rotation matrix from the F-coordinate to the E-coordinate,

$${}^{E}_{F}\mathbf{R}(\psi_2, \psi_3, \phi_{MF}) \equiv \begin{bmatrix} r_{11} & r'_{12} & r'_{13} \\ r_{21} & r'_{22} & r'_{23} \\ r_{31} & r'_{32} & r'_{33} \end{bmatrix}, \quad \phi_{MF} \equiv (\mathbf{j}_M, \mathbf{j}_F), \tag{13}$$

which leads to

$$\mathbf{i}_F = \begin{bmatrix} r_{11} \\ r_{21} \\ r_{31} \end{bmatrix}, \quad \mathbf{j}_F = \begin{bmatrix} r'_{12} \\ r'_{22} \\ r'_{32} \end{bmatrix}, \quad \mathbf{k}_F = \begin{bmatrix} r'_{13} \\ r'_{23} \\ r'_{33} \end{bmatrix}. \tag{14}$$

Here, ϕ_{MF} is the angle between the two unit vectors $(\mathbf{j}_M, \mathbf{j}_F)$. Let the vector $\mathbf{P}$ be represented by ${}^{F}\mathbf{P}$ in the F-coordinate. Transformation of ${}^{F}\mathbf{P}$ into ${}^{E}\mathbf{P}$ contains rotation and translation,

$${}^{E}\mathbf{P} = {}^{E}_{F}\mathbf{R}(\psi_2, \psi_3, \phi_{MF}){}^{F}\mathbf{P} + \mathbf{X}. \tag{15}$$

Use of the F-coordinate system simplifies the calculations for the lift and drag forces and torques acting on the cylinder. Since the M- and F-coordinates share a common axis $\mathbf{i}_M = \mathbf{i}_F$, the rotation matrix from the F- to M- coordinate systems is given by

$${}^{M}_{F}\mathbf{R} = {}^{M}_{E}\mathbf{R}\,{}^{E}_{F}\mathbf{R} = {}^{E}_{M}\mathbf{R}^{-1}(\psi_2, \psi_3)\,{}^{E}_{F}\mathbf{R}(\psi_2, \psi_3, \phi_{MF}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e_{22} & e_{23} \\ 0 & e_{32} & e_{33} \end{bmatrix} \tag{16}$$

it is two-dimensional with the rotation matrix given by

$${}^{M}_{F}\mathbf{E} = \begin{bmatrix} \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} e_{22} \\ e_{32} \end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix} e_{23} \\ e_{33} \end{bmatrix}. \tag{17}$$

Let the cylinder rotate around $(\mathbf{i}_F, \mathbf{j}_F, \mathbf{k}_F)$ with the angular velocity components represented by $(\omega_1^F, \omega_2^F, \omega_3^F)$ (Fig. 2). They are connected to the angular velocity components in the M-coordinate system by

$$\omega_1^F = \omega_1, \quad \begin{bmatrix} \omega_2^F \\ \omega_3^F \end{bmatrix} = {}^{F}_{M}\mathbf{E} \begin{bmatrix} \omega_2 \\ \omega_3 \end{bmatrix}. \tag{18}$$

3 Dynamics

3.1 Momentum Balance. The translation velocity of the cylinder $(\mathbf{V})$ is governed by the momentum equation in the E-coordinate system,

$$\frac{d}{dt}\begin{bmatrix} u \\ v \\ w \end{bmatrix} = -\begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} + \frac{\mathbf{F}_{nh} + \mathbf{F}_h}{\rho\Pi}, \tag{19a}$$

where g is the gravitational acceleration, Π is the cylinder volume, ρ is the rigid body density, $\rho\Pi = m$, is the cylinder mass, $\mathbf{F}_{nh}$ is the nonhydrodynamic force, and $\mathbf{F}_h$ is the hydrodynamic force (i.e., surface force including drag, lift, impact forces). Both $\mathbf{F}_{nh}$ and $\mathbf{F}_h$ are integrated for the cylinder. The drag and lift forces are calculated using the drag and lift laws with the given water-to-cylinder velocity $(\mathbf{V}_r)$. In the F-coordinate, $\mathbf{V}_r$ is decomposed into along-cylinder $(\mathbf{V}_1)$ and across-cylinder $(\mathbf{V}_2)$ components.

The nonhydrodynamic force $\mathbf{F}_{nh}$ is the buoyancy force $(\mathbf{F}_b)$ for the air and water phases,

$$\mathbf{F}_{nh} = \mathbf{F}_b = \mathbf{k}(\rho_a \Pi g, \rho_w \Pi g),$$

where (ρ_a, ρ_w) are the air and water densities. $\mathbf{F}_{nh}$ is the resultant of the buoyancy force $(\mathbf{F}_b)$, pore water pressure force $(\mathbf{F}_{pw})$, and shearing resistance force $(\mathbf{F}_s)$ for the sediment phase (see Sec. 6).

3.2 Moment of Momentum Equation. It is convenient to write the moment of momentum equation,

$$\mathbf{J} \cdot \frac{d\boldsymbol{\omega}}{dt} = -2\mathbf{J} \cdot (\boldsymbol{\Omega} \times \boldsymbol{\omega}) + \mathbf{M}_{nh} + \mathbf{M}_h, \tag{20}$$

in the M-coordinate system with the cylinder's angular velocity components $(\omega_1, \omega_2, \omega_3)$ defined by (19*a*) and (19*b*). Here, the first term on the right-hand side is an apparent torque (similar to the Corilois term in earth science) due to the use of the rotating coordinate system (i.e., the M-coordinate), and

$$\boldsymbol{\Omega} = \omega_2 \mathbf{j}_M + \omega_3 \mathbf{k}_M \tag{21}$$

is the angular velocity of the M-coordinate system. If $\omega_1 = 0$, then $\boldsymbol{\Omega} = \boldsymbol{\omega}$, which leads to

$$-2\mathbf{J} \cdot (\boldsymbol{\Omega} \times \boldsymbol{\omega}) = \begin{cases} 0, & \text{if } \omega_1 = 0 \text{ (i.e., } \boldsymbol{\Omega} = \boldsymbol{\omega}), \\ -2J_2\omega_1\omega_3\mathbf{j}_M + 2J_3\omega_1\omega_2\mathbf{k}_M, & \text{if } \omega_1 \neq 0. \end{cases} \tag{22}$$

In this study, the apparent torque is neglected. The gravity force, passing the COM, does not induce the moment. $\mathbf{M}_{nh}$ and $\mathbf{M}_h$ are the nonhydrodynamic and hydrodynamic force torques. In the M-coordinate system, the moment of gyration tensor for the axially symmetric cylinder is a diagonal matrix

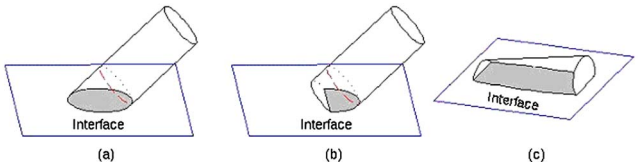


Fig. 3 Three patterns of cylinder penetration with the cross section being (*a*) a complete ellipse, (*b*) a cutoff ellipse with one side straight line, and (*c*) a cutoff ellipse with two side straight lines

$$\mathbf{J} = \begin{bmatrix} J_1 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_3 \end{bmatrix}, \tag{23}$$

where J_1, J_2, and J_3 are the moments of inertia. The buoyancy force induces the moment in the $\mathbf{j}_M$ direction if the COM does not coincide with the COV (i.e., $\chi \neq 0$),

$$\mathbf{M}_b = |\mathbf{F}_b|\chi \cos \psi_2 \mathbf{j}_M. \tag{24}$$

Computation of nonhydrodynamic and hydrodynamic forces $(\mathbf{F}_{nh}, \mathbf{F}_h)$ and torques $(\mathbf{M}_{nh}, \mathbf{M}_h)$ is more complicated for a cylinder penetrating through air-water and water-sediment interfaces than falling through a single medium such as water. At the instance when the cylinder penetrates into an interface, three situations may exist: the cross section is a complete ellipse [Fig. 3(*a*)], a cutoff ellipse with one side straight line [Fig. 3(*b*)], or a cutoff ellipse with two straight lines [Fig. 3(*c*)]. The interface separates the cylinder to two parts. Each part contains a noncylinder D and a subcylinder C (Fig. 4). Let (L_c, L_d), (Ω_c, Ω_d) and (Π_c, Π_d) be the lengths, surfaces, and volumes of $[C,D]$, and (h_1, h_2) the depths of the two sides of D (Fig. 5). The characteristics of the geometric parameters (L_c, h_1, h_2) are listed in Table 1. The COV for the portion $[C,D]$ is called the partial COV (PCOV).

4 Equivalent Cylinder Method

4.1 Equivalent Cylinder. During penetration, the part that contacts fluid (air or water) is treated as a cylinder [E] with the same mass and PCOV location and with the assumption that the buoyancy and hydrodynamic forces and torques for $[C,D]$ are the same for [E]. The cylinder [E], called the equivalent cylinder, is used to represent the part $[C,D]$. Thus, the theoretical procedure developed for calculating external forcing (buoyancy and hydrodynamic forces and torques) for a cylinder [8] can be easily used for [E].

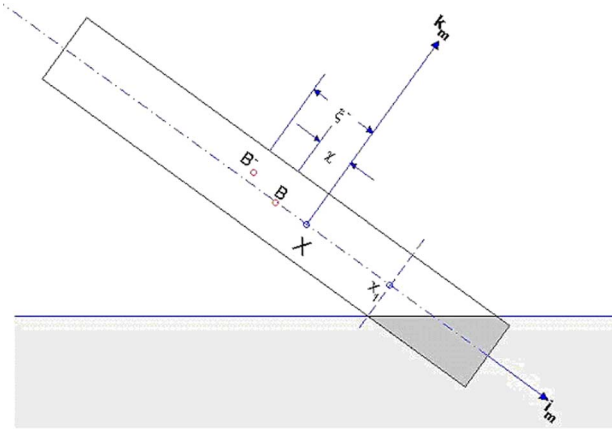


Fig. 4 Illustration of PCOV (B^-), x_1, and ξ^- for the tail part $[C^{(1)}, D^{(1)}]$ for the case in Fig. 3(*a*)

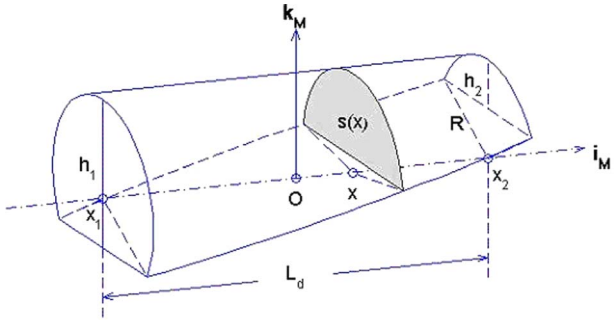


Fig. 5 Geometry of the part $D^{(1)}$

4.2 Volume of $[C,D]$. In the M-coordinate system, the area of the vertical cross section of D is given by

$$s(x) = R^2 \cos^{-1}\left(1 - \frac{h(x)}{R}\right) - [R - h(x)]\sqrt{R^2 - [R - h(x)]^2}, \tag{25a}$$

where $h(x)$ is the depth of the cross section,

$$h(x) = h_1 + \frac{\Delta h}{L_d}(x - x_1), \quad \Delta h = h_2 - h_1, \tag{25b}$$

where L_d is the length of D (see Fig. 5). Integration of $s(x)$ along x axis gives the volume of D,

$$\Pi_d = \int_{x_1}^{x_2} s(x)dx = \frac{R^3 L_d}{\Delta h}\beta(\kappa_1, \kappa_2) = \pi R^2 l_d, \tag{26}$$

where

$$\kappa_1 = 1 - \frac{h_1}{R}, \quad \kappa_2 = 1 - \frac{h_2}{R}, \tag{27a}$$

$$\beta(\kappa_1, \kappa_2) \equiv \kappa_1 \cos^{-1}(\kappa_1) - \sqrt{1 - \kappa_1^2} + \tfrac{1}{3}(1 - \kappa_1^2)^{3/2} - \kappa_2 \cos^{-1}(\kappa_2) + \sqrt{1 - \kappa_2^2} - \tfrac{1}{3}(1 - \kappa_2^2)^{3/2}, \tag{27b}$$

and

$$l_d = \frac{RL_d}{\pi \Delta h}\beta(\kappa_1, \kappa_2). \tag{27c}$$

Here, l_d is the equivalent length of D. The volume of C is calculated by

$$\Pi_c = \pi R^2 L_c. \tag{28}$$

The total volume of $[C,D]$ is

$$\Pi = \pi R^2 l,$$

and

$$l = L_c + l_d$$

is the length of the equivalent cylinder E.

Table 1 Geometric parameters during the cylinder penetration

	L_c	h_1	h_2
Upper and lower parts of Fig. 3(*a*)	>0	2R	0
Upper part of Fig. 3(*b*)	>0	2R	0∼2R
Lower part of Fig. 3(*b*)	0	0∼2R	0
Upper and lower parts of Fig. 3(*c*)	0	0∼2R	0∼2R

4.3 PCOV of $[C,D]$. Let (ξ^+,η^+) and (ξ^-,η^-) be the PCOV of the head $[C,D]$ (in the direction of $\mathbf{i}_M$) or tail (in the opposite direction of $\mathbf{i}_M$) $[C,D]$ (denoted by $B^\pm$, positive sign for the head part) in the M-coordinate system,

$$(\xi^\pm,\eta^\pm)=\frac{1}{\Pi}\Big[\iiint_{\Pi_c}(x,z)dv+\iiint_{\Pi_d}(x,z)dv\Big]$$
$$=\frac{1}{\Pi_c+\Pi_d}\left[\left(x_1\pm\frac{L_c}{2},0\right)\Pi_c+\int_{x_1}^{x_2}(x,z)s(x)dx\right]. \tag{29}$$

where x_1 is defined as the location of interface between C and D. Substitution of (25*a*), (25*b*), (26), and (28) into (29) leads to

$$\xi^\pm=x_1\mp\frac{R\Delta h}{\beta(\kappa_1,\kappa_2)L_d(1+\pi\Delta hL_c\beta^{-1}L_d^{-1})}\times\left[\left(\frac{L_d}{\Delta h}\right)^2\mu_x(\kappa_1,\kappa_2)\pm\frac{1}{2}\left(\frac{L_c}{R}\right)^2\right], \tag{30}$$

$$\eta^\pm=\pm\,\mathrm{sign}(\cos\Psi_2)\frac{R}{6\beta(\kappa_1,\kappa_2)(1+\pi\Delta hL_c\beta^{-1}L_d^{-1})}\mu_z(\kappa_1,\kappa_2), \tag{31}$$

where

$$\mu_x(\kappa_1,\kappa_2)\equiv\frac{1}{4}[(2\kappa_2^2-1)\cos^{-1}\kappa_2-(2\kappa_1^2-1)\cos^{-1}\kappa_1+\kappa_1\sqrt{1-\kappa_1^2}-\kappa_2\sqrt{1-\kappa_2^2}]+\frac{1}{4}[\kappa_2\sqrt{(1-\kappa_2^2)^3}-\kappa_1\sqrt{(1-\kappa_1^2)^3}]$$
$$-\frac{1}{8}(\kappa_2\sqrt{1-\kappa_2^2}-\kappa_1\sqrt{1-\kappa_1^2}+\sin^{-1}\kappa_2-\sin^{-1}\kappa_1)-\kappa_1(\kappa_2\cos^{-1}\kappa_2-\kappa_1\cos^{-1}\kappa_1+\sqrt{1-\kappa_1^2}-\sqrt{1-\kappa_2^2})-\frac{\kappa_1}{3}[\sqrt{(1-\kappa_2^2)^3}-\sqrt{(1-\kappa_1^2)^3}],$$

$$\mu_z(\kappa_1,\kappa_2)\equiv\kappa_1\sqrt{(1-\kappa_1^2)^3}-\kappa_2\sqrt{(1-\kappa_2^2)^3}+\frac{3}{2}(\kappa_1\sqrt{1-\kappa_1^2}-\kappa_2\sqrt{1-\kappa_2^2}+\sin^{-1}\kappa_1-\sin^{-1}\kappa_2). \tag{32}$$

Let (ξ,η) be (ξ^+,η^+) for the head part and (ξ^-,η^-) for the tail part. The position vector of PCOV in the M-coordinate system is represented by

$$\mathbf{r}_{\mathrm{PCOV}}=\xi\mathbf{i}_M+\eta\mathbf{k}_M. \tag{33}$$

5 Forces and Torques in Air and Water

Calculation of the buoyancy force and torque is straightforward. Calculation of the surface force and torque is not simple. Assume that the surface force and torque on the equivalent cylinder E are the same on the $[C,D]$. If $[C,D]$ moves in fluid (air or water), the recursive model recently developed [8] can be used to calculate for equivalent cylinder E. Thus, the water column is taken as the example to illustrate the calculation of the hydrodynamic force and torque. Computation of the surface force and torque due to sediment is described in Sec. 6.

5.1 Buoyancy Force and Torque. The buoyancy force $\mathbf{F}_b$ is the product of the air (or water) density and volume,

$$\mathbf{F}_b=\rho(\Pi_c+\Pi_d)\mathbf{k}=\rho\pi R^2(L_c+l_d)\mathbf{k}. \tag{34}$$

The torque due to the buoyancy force for the upper or lower part is given by

$$\mathbf{M}_b=\mathbf{r}_{\mathrm{PCOV}}\times\mathbf{F}_b. \tag{35}$$

Substitution of (33) and (34) into (35) leads to

$$\mathbf{M}_b=-\rho\pi R^2(L_c+l_d)(\xi\cos\psi_2+\eta\sin\psi_2)\mathbf{j}_M. \tag{36}$$

5.2 Drag and Lift Forces. The drag and lift forces exerted on the cylinder is represented by

$$\mathbf{F}_h=(F_{d1}\mathbf{i}_F+F_{d2}\mathbf{j}_F+F_{d3}\mathbf{k}_F)+\mathbf{F}_l, \tag{37}$$

where (F_{d1},F_{d2},F_{d3}) are the components of drag force along $\mathbf{i}_F$ (along-cylinder), $\mathbf{j}_F$ (across-cylinder) and $\mathbf{k}_F$ directions. $\mathbf{F}_l$ represented the lift force. Linearization of drag and lift laws is used in the computation.

Let (C_{d1},C_{d2}) be the drag coefficients in the along- and across-cylinder directions (Reynolds number dependent). The drag force coefficients are calculated on the base of steady flow; they is different from the fluid around an accelerated solid body. The added mass correction is represented by the ratios (f_1,f_2,f_3) in the three directions of the F-coordinate system.

The drag force along $\mathbf{i}_F$ is calculated by

$$F_{d1}=C_{td1}(t)V_1, \tag{38}$$

$$C_{td1}(t)\equiv C_{d1}\frac{\pi R^2}{2}\frac{\rho_w}{(1+f_1)}|V_1(t)|. \tag{39}$$

C_{d1} is almost independent on the axial Reynolds number (Re) when $\mathrm{Re}>10^4$, but dependent on the cylinder's aspect ratio [9],

$$C_{d1}=\begin{cases}1.0, & \text{if } \delta>8,\\ 0.75+\delta/32.1934+0.09612/\delta^2, & \text{if } 8\geqslant\delta>0.5,\\ 1.15, & \text{if } \delta\leqslant 0.5.\end{cases} \tag{40}$$

Substitution of (11) and (12) into (38) leads to

$$F_{d1}\mathbf{i}_F=-C_{td1}(t)\mathbf{I}_{11}\cdot\left(\begin{bmatrix}u\\ v\\ w\end{bmatrix}-\begin{bmatrix}u_w\\ v_w\\ w_w\end{bmatrix}\right),\quad \mathbf{I}_{11}=\mathbf{i}_F\mathbf{i}_F^{\mathrm{T}}, \tag{41}$$

where the superscript "T" denotes the transpose.

The drag force along $\mathbf{j}_F$ is calculated by

$$F_{d2}=R\int_{-L/2-\chi}^{L/2-\chi}C_{d2}(V_2')^2\frac{\rho_w}{(1+f_2)}d\eta=C_{td2}(t)V_2+f_{rd2}(t), \tag{42}$$

where

$$V_2'(\eta)=V_2-\omega_3^F\eta$$

is the water-to-cylinder velocity at the surface in the $\mathbf{j}_F$ direction and

$$C_{td2}(t)\equiv 2C_{d2}LR\frac{\rho_w}{(1+f_2)}\left(\frac{V_2}{2}+\chi\omega_3^F\right), \tag{43a}$$

$$f_{rd2}(t)\equiv 2C_{d2}LR\frac{\rho_w}{(1+f_2)}\left(\frac{1}{2}\chi^2+\frac{1}{24}L^2\right)(\omega_3^F)^2. \tag{43b}$$

An empirical formula is used for calculating C_{d2} [10]

$$C_{d2}=\begin{cases}1.9276+8/\mathrm{Re}, & \text{if } \mathrm{Re}\leqslant 12,\\ 1.261+16/\mathrm{Re}, & \text{if } 12<\mathrm{Re}\leqslant 180,\\ 0.855+89/\mathrm{Re}, & \text{if } 180<\mathrm{Re}\leqslant 2000,\\ 0.84+0.000\,03\,\mathrm{Re}, & \text{if } 2000<\mathrm{Re}\leqslant 12{,}000,\\ 1.2-4/\delta, & \text{if } 12{,}000<\mathrm{Re}\leqslant 150{,}000,\quad \delta\geqslant 10,\\ 0.835-0.35/\delta, & \text{if } 12{,}000<\mathrm{Re}\leqslant 150{,}000,\quad 2\leqslant\delta<10,\\ 0.7-0.08/\delta, & \text{if } 12{,}000<\mathrm{Re}\leqslant 150{,}000,\quad \delta<2,\\ 1.875-0.000\,004\,5\,\mathrm{Re}, & \text{if } 150{,}000<\mathrm{Re}\leqslant 350{,}000,\\ 1/(641{,}550/\mathrm{Re}+1.5), & \text{if } \mathrm{Re}>350{,}000.\end{cases} \tag{44}$$

Substitution of (11) and (12) into (42) leads to

$$F_{d2}\mathbf{j}_{\mathrm{F}}=-C_{td2}(t)\mathbf{I}_{22}\cdot\left(\begin{bmatrix}u\\v\\w\end{bmatrix}-\begin{bmatrix}u_w\\v_w\\w_w\end{bmatrix}\right)+f_{rd2}(t)\mathbf{j}_{\mathrm{F}},\quad \mathbf{I}_{22}=\mathbf{j}_{\mathrm{F}}\mathbf{j}_{\mathrm{F}}^{\mathrm{T}}. \tag{45}$$

The angular velocity (ω_2^F) causes nonuniform water-to-cylinder velocity in the $\mathbf{k}_{\mathrm{F}}$ direction,

$$V_3=\omega_2^F\eta. \tag{46}$$

The drag force along $\mathbf{k}_{\mathrm{F}}$ is calculated by

$$\mathbf{F}_{d3}=\left[C_{d2}R\frac{\rho_w}{(1+f_2)}\omega_2^F|\omega_2^F|\left(\int_0^{\frac{L}{2}-\chi}\eta^2d\eta-\int_{-\frac{L}{2}-\chi}^{0}\eta^2d\eta\right)\right]\mathbf{k}_{\mathrm{F}}=f_{rd3}(t)\mathbf{k}_F, \tag{47}$$

where

$$f_{rd3}(t)\equiv-\frac{1}{6}C_{d2}\frac{\rho_wR}{(1+f_2)}\chi(3L^2+4\chi^2)|\omega_2^F|\omega_2^F \tag{48}$$

is the rotational drag force in the $\mathbf{k}_{\mathrm{F}}$ direction.

The water-to-cylinder velocity determines the lift force [11]

$$\mathbf{F}_l=\left[\frac{C_{tl}(t)}{L}\int_{-L/2-\chi}^{L/2-\chi}V_2'(\eta)d\eta\right]\mathbf{k}_{\mathrm{F}},\quad C_{tl}(t)\equiv C_lLR\frac{\rho_w}{(1+f_2)}|V_2|, \tag{49}$$

where C_l is the lift coefficient. An empirical formula is used for calculating C_l [12],

$$C_l=\begin{cases}2\omega_1R/V_2, & \text{if } \omega_1R/V_2\leqslant 4,\\ 8+0.24(\omega_1R/V_2-4), & \text{if } \omega_1R/V_2>4.\end{cases} \tag{50}$$

Substitution of (11) and (12) into (49) leads to

$$\mathbf{F}_l=-C_{tl}(t)\mathbf{I}_{32}\cdot\left(\begin{bmatrix}u\\v\\w\end{bmatrix}-\begin{bmatrix}u_w\\v_w\\w_w\end{bmatrix}\right)+f_{rl}(t)\mathbf{k}_{\mathrm{F}},\quad \mathbf{I}_{32}=\mathbf{k}_{\mathrm{F}}\mathbf{j}_{\mathrm{F}}^{\mathrm{T}}, \tag{51}$$

where

$$f_{rl}(t)\equiv C_{tl}(t)\chi\omega_3^F$$

is the rotational lift force. Substitution of (41), (45), (47), and (51) into (37) and use of (14) lead to

$$\mathbf{F}_h=-[C_{td1}(t)\mathbf{I}_{11}+C_{td2}(t)\mathbf{I}_{22}+C_{tl}(t)\mathbf{I}_{32}]\cdot\left(\begin{bmatrix}u\\v\\w\end{bmatrix}-\begin{bmatrix}u_w\\v_w\\w_w\end{bmatrix}\right)+f_{rd2}(t)\times\begin{bmatrix}r_{12}'\\r_{22}'\\r_{32}'\end{bmatrix}+[f_{rd3}(t)+f_{rl}(t)]\begin{bmatrix}r_{13}'\\r_{23}'\\r_{33}'\end{bmatrix}. \tag{52}$$

Substitution of (52) into (19*a*) leads to the cylinder's momentum equation,

$$\frac{d}{dt}\begin{bmatrix}u\\v\\w\end{bmatrix}=-\mathbf{D}\cdot\begin{bmatrix}u\\v\\w\end{bmatrix}+\boldsymbol{\alpha}_1, \tag{19b}$$

where

$$\boldsymbol{\alpha}_1\equiv\mathbf{D}\cdot\begin{bmatrix}u_w\\v_w\\w_w\end{bmatrix}-\begin{bmatrix}0\\0\\(1-\rho_w/\rho)g\end{bmatrix}+b_1\begin{bmatrix}r_{12}'\\r_{22}'\\r_{32}'\end{bmatrix}+b_2\begin{bmatrix}r_{13}'\\r_{23}'\\r_{33}'\end{bmatrix},$$

$$\mathbf{D}\equiv\frac{C_{td1}(t)\mathbf{I}_{11}+C_{td2}(t)\mathbf{I}_{22}+C_{tl}(t)\mathbf{I}_{23}}{\rho\Pi},\quad b_1\equiv\frac{f_{rd2}(t)}{\rho\Pi},$$

$$b_2\equiv\frac{f_{rd3}(t)+f_{rl}(t)}{\rho\Pi}.$$

5.3 Drag and Lift Torques. For an axially symmetric cylinder, the moment of the hydrodynamic force in the $\mathbf{i}_{\mathrm{F}}$ direction is not caused by the drag and lift forces, but by the viscous fluid. The moment of the viscous force of steady flow between two rotating cylinders with the common axis is calculated by [1]

$$M=4\pi\mu\frac{r_1^2\cdot r_0^2}{r_1^2-r_0^2}(\omega_1-\omega_0),$$

where (r_1,r_0) and (ω_1,ω_0) are the radii and angle velocities of the inner and outer cylinders and μ is the viscosity. The moment of the viscous force on one rotating cylinder is th limit case of the two rotating cylinders as $r_0\to\infty$ and $\omega_0=0$. The moment of the viscous force around $\mathbf{i}_{\mathrm{F}}$ is calculated by

$$\mathbf{M}_{v1}=-C_{m1}\omega_1\mathbf{i}_{\mathrm{F}},\quad C_{m1}\equiv\pi\mu Ld^2. \tag{53}$$

Same as the hydrodynamic forces, the torques along the $\mathbf{j}_{\mathrm{F}}$ and $\mathbf{k}_{\mathrm{F}}$ axes, $(\mathbf{M}_{d1},\mathbf{M}_{d2},\mathbf{M}_l)$, are calculated. When the cylinder rotates around $\mathbf{j}_{\mathrm{F}}$ with the angular velocity ω_2^F, the drag force causes a torque on the cylinder in the $\mathbf{j}_{\mathrm{F}}$ direction,

$$\mathbf{M}_{d2}=\left[-\omega_2^F|\omega_2^F|\int_{-L/2-\chi}^{L/2-\chi}C_{d2}R\frac{\rho_w}{(1+f_r)}\eta^2|\eta|d\eta\right]\mathbf{j}_{\mathrm{F}}=-[C_{m2}(t)\omega_2^F]\mathbf{j}_{\mathrm{F}}, \tag{54}$$

$$C_{m2}(t) \equiv \frac{1}{2} C_{d2} R \frac{\rho_w}{(1+f_r)} \left(\frac{1}{16} L^4 + \frac{3}{2} L^2 \chi^2 + \chi^4 \right) |\omega_2^F|,$$

where f_r is the added mass factor for the moment of drag and lift forces. If the water-to-cylinder velocity or the cylinder mass distribution is nonuniform ($\chi \neq 0$), the drag force causes a torque on the cylinder in the $\mathbf{k}_F$ direction,

$$\mathbf{M}_{d3} = \left[\int_{-L/2-\chi}^{L/2-\chi} C_{d2} R \frac{\rho_w}{(1+f_r)} (V_2 - \omega_3^F \eta)^2 \eta d\eta \right] \mathbf{k}_F$$
$$= -[C_{m3}(t)\omega_3^F + M_3(t)]\mathbf{k}_F, \tag{55}$$

$$C_{m3}(t) \equiv C_{d2} R \frac{\rho_w}{(1+f_r)} \left(\frac{1}{6} V_2 L^3 + V_2 L \chi^2 + \frac{1}{4} L^3 \omega_3^F \chi + L \chi^3 \omega_3^F \right), \tag{56a}$$

$$M_3(t) \equiv C_{d2} R \frac{\rho_w}{(1+f_r)} V_2^2 L \chi. \tag{56b}$$

The lift force exerts a torque on the cylinder in the $\mathbf{j}_F$ direction,

$$\mathbf{M}_{l2} = \left[- \int_{-L/2-\chi}^{L/2-\chi} C_l R \frac{\rho_w}{f_{kr}} (V_2 - \omega_3^F \eta) \eta d\eta \right] \mathbf{j}_F$$
$$= [C_{ml}(t)\omega_3^F + M_l(t)]\mathbf{j}_F, \tag{57}$$

$$C_{ml}(t) \equiv C_l V_2 R \frac{\rho_w}{(1+f_r)} L \left(\frac{1}{12} L^2 + \chi^2 \right), \quad M_l(t) \equiv R \frac{\rho_w}{f_{kr}} L V_2^2 \chi. \tag{58}$$

After the angular velocity components (ω_2^F, ω_3^F) are transformed into (ω_2, ω_3) (from the F-coordinate to the M-coordinate) using (18), and the unit vectors $(\mathbf{j}_F, \mathbf{k}_F)$ are transformed into $(\mathbf{j}_M, \mathbf{k}_M)$ using the rotation matrix (17), the drag force torques in the $\mathbf{j}_F$ direction (54) and in the $\mathbf{k}_F$ direction (55) are represented by

$$\mathbf{M}_{d2} = -C_{m2}(t)\mathbf{H}_{22} \cdot \begin{bmatrix} \omega_2 \\ \omega_3 \end{bmatrix}, \quad \mathbf{H}_{22} = \mathbf{e}_2 \mathbf{e}_2^T, \tag{59}$$

$$\mathbf{M}_{d3} = -C_{m3}(t)\mathbf{H}_{33} \cdot \begin{bmatrix} \omega_2 \\ \omega_3 \end{bmatrix} - M_3(t)\mathbf{e}_3, \quad \mathbf{H}_{33} = \mathbf{e}_3 \mathbf{e}_3^T, \tag{60}$$

and the lift torque in the $\mathbf{j}_F$ (57) is represented by

$$\mathbf{M}_{l2} = C_{ml}(t)\mathbf{H}_{23} \cdot \begin{bmatrix} \omega_2 \\ \omega_3 \end{bmatrix} + M_l(t)\mathbf{e}_2, \quad \mathbf{H}_{23} = \mathbf{e}_2 \mathbf{e}_3^T. \tag{61}$$

Summation of (53) and (59)–(61) leads to

$$\mathbf{M}_h = \mathbf{M}_v + \mathbf{M}_{d2} + \mathbf{M}_{d3} + \mathbf{M}_{l2}$$
$$= -C_{m1}\omega_1 \mathbf{i}_F - [C_{m2}(t)\mathbf{H}_{22} + C_{m3}(t)\mathbf{H}_{33} - C_{ml}(t)\mathbf{H}_{23}] \cdot \begin{bmatrix} \omega_2 \\ \omega_1 \end{bmatrix}$$
$$+ M_l(t)\mathbf{e}_2 - M_3(t)\mathbf{e}_3. \tag{62}$$

6 Resistant Forces in Sediment

6.1 Water Cavity. As the cylinder impacts and penetrates into the sediment, it pushes the sediment and leaves space in the wake. This space is refilled by water right away and a water cavity is produced (Fig. 6). At the instant of the penetration, the total resistant force on the cylinder is represented by

$$\mathbf{F}^s = \int_{\sigma_{sed}} [\delta(\mathbf{f}_b^s + \mathbf{f}_{sh}) + \mathbf{f}_b^w + \mathbf{f}_h^w] d\sigma + \mathbf{F}_{pw}, \tag{63}$$

where $(\mathbf{f}_b^s, \mathbf{f}_{sh})$ and $(\mathbf{f}_b^w, \mathbf{f}_h^w)$ are the sediment buoyancy and shear resistance forces and water buoyancy and hydrodynamic forces

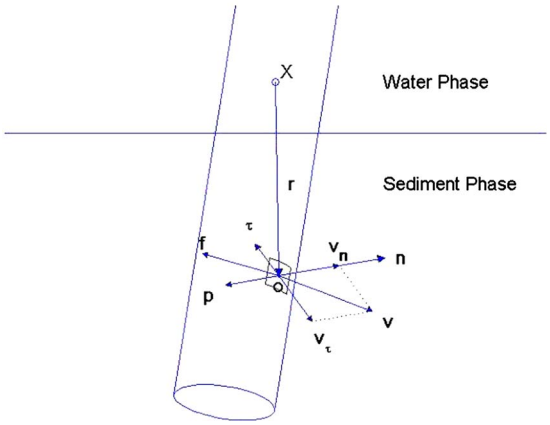


Fig. 6 The impact (resistant) force exerted on the part of the object's surface moving towards the sediment

(per unit area) at the point $\mathbf{r}$ over the cylinder's surface; σ_{sed} is the area of the cylinder's surface below the water-sediment interface; and $\mathbf{F}_{pw}$ is the pore water pressure force on the whole cylinder. In the sediment, the magnitude of the sediment nonhydrostatic force is much larger than the magnitude of the water hydrodynamic force,

$$|\mathbf{f}^s| \gg |\mathbf{f}_h^w|,$$

which means that $\mathbf{f}_h^w$ in (63) can be neglected. The water buoyancy force per unit area over the cylinder's surface is defined by

$$\mathbf{f}_b^w = -\rho_w g(z_{ws} - z)\mathbf{n}, \tag{64}$$

where z_{ws} is the depth of the water-sediment interface and $\mathbf{n}$ is the unit vector normal to the cylinder surface (outward positive).

Let $\mathbf{v}$ be the velocity at point $\mathbf{r}$ (represented in the M-coordinate) on the cylinder surface,

$$\mathbf{v} = \mathbf{V} + \boldsymbol{\omega} \times \mathbf{r}.$$

The step function δ is defined by

$$\delta = \begin{cases} 1, & \mathbf{v} \cdot \mathbf{n} \geqslant 0, \\ 0, & \mathbf{v} \cdot \mathbf{n} \leqslant 0, \end{cases} \tag{65}$$

which shows that the sediment buoyancy and shear resistance forces act when the cylinder moves towards it. Let $\mathbf{v}_n$ be the normal velocity. The tangential velocity is represented by

$$\mathbf{v}_\tau = \mathbf{v} - \mathbf{v}_n. \tag{66}$$

The tangential unit vector ($\boldsymbol{\tau}$) is defined by

$$\boldsymbol{\tau} = -\frac{\mathbf{v}_\tau}{|\mathbf{v}_\tau|}, \tag{67}$$

which is opposite to $\mathbf{v}_\tau$ (Fig. 7).

6.2 Sediment Resistant Forces. When the cylinder impacts and penetrates into the sediment, it will create a large transient pore pressure in the sediment that causes ruptures in the sediment which influences the lifting forces on the cylinder [13,14].

The sediment buoyancy force per unit area is defined by

$$\mathbf{f}_b^s = -\mathbf{n} \int_z^{z_{ws}} \rho_s(z') g dz', \tag{68}$$

where $\rho_s(z)$ is the sediment density.

The shear resistant force per unit area $\mathbf{f}_{sh}$ depends on the cylinder's penetration speed (V) and the sediment strength. Let $S(z)$ be the sediment shear strength. The shear strength is defined as the maximum stress that a material can withstand before failure in

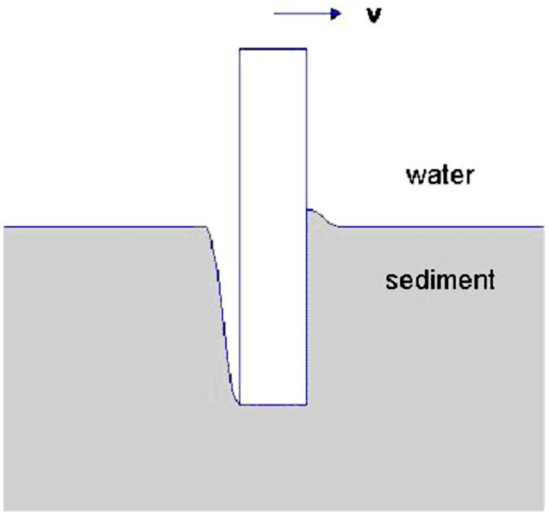


Fig. 7 Momentum and angular momentum balance for the cylinder's penetration through the water-sediment interface

shear. Calculation of shear strength depends upon the test method.

After entering the water-sediment interface, the cylinder reduces its speed (V), and the sediment shearing resistant force also decreases. When the cylinder ceases, the shearing resistant force should be the same as the sediment shear strength $S(z)$. Thus, the shearing resistant force is represented by

$$\mathbf{f}_{sh} = S(z)G(V)\boldsymbol{\tau}, \quad G(0) = 1, \tag{69}$$

where $G(V)$ is the impact function defined by

$$G(V) = A\left[1 - (1 - A^{-1})\exp\left(-\frac{V}{V_{\text{rest}}}\right)\right]. \tag{70}$$

Here, V_{rest} is an infinitesimally small value for V representing the cease of the cylinder in the sediment. The impact function has the following feature,

$$\lim_{V\to\infty} G(V) = A, \tag{71}$$

which shows that when the cylinder impacts on the sediment (usually with large penetration speed), the impact function takes the value of A. Thus, we may call A the impact factor. Note that A and V_{rest} are the two tuning parameters of the numerical model. In this study we use

$$A = 10, \quad V_{\text{rest}} = 0.04 \text{ m s}^{-1}. \tag{72}$$

The shear strength of the sediment is directly measured from the gravity cores using the fall cone apparatus (model G-200) (see Sec. 8.2).

The total force due to the pore water pressure on the cylinder is computed by [15]

$$\mathbf{F}_{pw} = \left[\frac{\pi}{8}\rho_s(z)\left(\frac{gw}{k_p} + \frac{1+e_v}{e_v}\frac{dw}{dt}\right)B^3\right]\mathbf{k}, \tag{73}$$

where k_p is the permeability coefficient (10^{-4} m s^{-1} [15]), $e_v(\sim 0.50)$ is the void ratio, and B is the length of the rupture line. Substitution of (64), (68), (69), and (73) into (63) leads to

$$\mathbf{F}^s = \int_{\sigma_{\text{sed}}} \boldsymbol{\tau}[\delta\mu G(V)S(z)]d\sigma - \int_{\sigma_{\text{sed}}} \mathbf{n}\left[\left(\delta\int_z^{z_{ws}} \rho_s(z')g dz'\right) + \rho_w g(z_{ws} - z)\right]d\sigma + \mathbf{k}\frac{\pi}{8}\rho_s(z)\left(\frac{gw}{k_p} + \frac{1+e_v}{e_v}\frac{dw}{dt}\right)B^3, \tag{74}$$

which is the external force acted on the cylinder in the sediment phase. The torque due to the sediment ($\mathbf{M}^s$) is calculated by

$$\mathbf{M}^s = \int_{\sigma_{\text{sed}}} (\mathbf{r}\times\boldsymbol{\tau})[\delta\mu G(V)S(z)]d\sigma + \int_{\sigma_{\text{sed}}} (\mathbf{r}\times\mathbf{n}) \times\left[\left(\delta\int_z^{z_{ws}} \rho_s(z')g dz'\right) + \rho_w g(z_{ws} - z)\right]d\sigma + (\mathbf{r}_{pw}\times\mathbf{k})\frac{\pi}{8}\rho_s(z)\left(\frac{gw}{k} + \frac{1+e_v}{e_v}\frac{dw}{dt}\right)B^3. \tag{75}$$

where $\mathbf{r}_{pw}$ is the position vector (in the M-coordinate) indicating the location of the cylinder's rupture line.

7 Model Integration

The momentum equation (19*a*) and (19*b*) and moment of momentum equation (20) are integrated numerically using the triple coordinate transformation. The momentum equation is integrated in the E-coordinate system. The hydrodynamic (drag and lift) force is transformed from the F-coordinate to the E-coordinate. The moment of momentum equation is integrated in the M-coordinate system. The hydrodynamic torque is transformed from the F-coordinate to the M-coordinate. After the cylinder penetrates into the sediment, the resistant force due to sediment $\mathbf{f}^s$ reduces the cylinder's speed and changes the turning angle.

7.1 Cylinder's Angular Velocity. Substitution of (24) and (62) into (20) leads to the equations for $(\omega_1, \omega_2, \omega_3)$,

$$\frac{d\omega_1}{dt} = -a_1\omega_1, \tag{76}$$

$$\frac{d}{dt}\begin{bmatrix}\omega_2\\ \omega_3\end{bmatrix} = -\mathbf{B}\cdot\begin{bmatrix}\omega_2\\ \omega_3\end{bmatrix} + \boldsymbol{\alpha}_2, \tag{77}$$

where

$$a_1 \equiv \frac{C_{m1}}{J_1} = \frac{8\pi\mu L}{\rho\Pi},$$

$$\mathbf{B} \equiv \begin{bmatrix}\frac{1}{J_2} & 0\\ 0 & \frac{1}{J_3}\end{bmatrix}\cdot[C_{m2}(t)\mathbf{H}_{22} + C_{m3}(t)\mathbf{H}_{33} - C_{ml}(t)\mathbf{H}_{23}],$$

$$\boldsymbol{\alpha}_2 \equiv \begin{bmatrix}\frac{1}{J_2} & 0\\ 0 & \frac{1}{J_3}\end{bmatrix}\cdot(M_l\mathbf{e}_2 - M_3\mathbf{e}_3) + \frac{\Pi\chi g\rho_w}{J_2}\cos\psi_2\begin{bmatrix}1\\ 0\end{bmatrix}. \tag{78}$$

Equation (76) has the analytical solution

$$\omega_1(t) = \omega_1(t_0)\exp[-a_1(t - t_0)], \tag{79}$$

which represents damping rotation of the cylinder around the main axis ($\mathbf{i}_{\text{M}}$). Substitution of (79) into (8) leads to

$$\frac{d\varphi_1}{dt} = \omega_1(t_0)\exp[-a_1(t - t_0)],$$

and its integration leads to

$$\varphi_1(t) = -\frac{\omega_1(t_0)}{a_1}\exp[-a_1(t - t_0)] + \varphi_1(t_0). \tag{80}$$

Equations (79) and (80) are the analytic formulas for predicting the angle and angular velocity around the cylinder's main axis (ω_1, φ_1).

7.2 Recursive Procedure. The basic equations (19*a*), (19*b*), (77), (79), and (80) describe the dynamics of the falling cylinder. It is noted that the coefficient matrices $\mathbf{B}$, $\mathbf{D}$ and the vectors $\boldsymbol{\alpha}_1$,

$\boldsymbol{\alpha}_2$ depend on drag/lift coefficients. Besides, $\mathbf{B}$, $\mathbf{D}$, $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ depend on the fluid-to-cylinder velocity and cylinder's angular velocity. Equations (19*a*), (19*b*), and (79) are nonlinear equations.

Let matrices $\mathbf{B}$ and $\mathbf{D}$ be separated into diagonal and nondiagonal parts,

$$\mathbf{D}=\mathbf{D}_1+\mathbf{D}_2,\quad \mathbf{D}_1\equiv\begin{bmatrix} d_1 & 0 & 0\\ 0 & d_2 & 0\\ 0 & 0 & d_3\end{bmatrix},\quad \mathbf{D}_2\equiv\begin{bmatrix} 0 & d_{12} & d_{13}\\ d_{21} & 0 & d_{23}\\ d_{31} & d_{32} & 0\end{bmatrix}, \tag{81}$$

$$\mathbf{B}=\mathbf{B}_1+\mathbf{B}_2,\quad \mathbf{B}_1\equiv\begin{bmatrix} b_1 & 0\\ 0 & b_2\end{bmatrix},\quad \mathbf{B}_2\equiv\begin{bmatrix} 0 & b_{12}\\ b_{21} & 0\end{bmatrix}. \tag{82}$$

Substitution of (81) into (19*a*) and (19*b*) leads to

$$\frac{d\mathbf{V}}{dt}=-\mathbf{D}_1\cdot\mathbf{V}+\boldsymbol{\beta},\quad \boldsymbol{\beta}\equiv\boldsymbol{\alpha}_1-\mathbf{D}_2\cdot\mathbf{V},\quad \mathbf{V}=\begin{bmatrix} u\\ v\\ w\end{bmatrix}\equiv\begin{bmatrix} v_1\\ v_2\\ v_3\end{bmatrix}, \tag{83a}$$

and substitution of (82) into (77) leads to

$$\frac{d}{dt}\begin{bmatrix}\omega_2\\ \omega_3\end{bmatrix}=-\mathbf{B}_1\cdot\begin{bmatrix}\omega_2\\ \omega_3\end{bmatrix}+\boldsymbol{\gamma},\quad \boldsymbol{\gamma}\equiv\boldsymbol{\alpha}_2-\mathbf{B}_2\cdot\begin{bmatrix}\omega_2\\ \omega_3\end{bmatrix}. \tag{83b}$$

If $\mathbf{B}_1$, $\mathbf{D}_1$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ are taken the values given at the present time step t_n, (83*a*) and (83*b*) can be treated as "linear" equations (local linearization) on $[t_n, t_n+\Delta t]$ and integrated analytically,

$$v_i(t_{n+1})=\begin{cases} v_i(t_n)+\beta_i(t_n)\Delta t, & \text{if } d_i(t_n)=0,\\ \left(v_i(t_n)-\dfrac{\beta_i(t_n)}{d_i(t_n)}\right)\exp[-d_i(t_n)\Delta t]+\dfrac{\beta_i(t_n)}{d_i(t_n)}, & \text{if } d_i(t_n)\neq 0,\quad i=1,2,3,\end{cases} \tag{84}$$

and

$$\omega_i(t_{n+1})=\begin{cases} \omega_i(t_n)+\gamma_i(t_n)\Delta t, & \text{if } d_i(t_n)=0,\\ \left(\omega_i(t_n)-\dfrac{\gamma_i(t_n)}{b_i(t_n)}\right)\exp[-b_i(t_n)\Delta t]+\dfrac{\gamma_i(t_n)}{b_i(t_n)}, & \text{if } d_i(t_n)\neq 0,\quad i=2,3.\end{cases} \tag{85}$$

Integration of (83*a*) twice from t_n leads to the translation position of the cylinder at t_{n+1},

$$x_i(t_{n+1})=\begin{cases} x_i(t_n)+v_i(t_n)\Delta t+\frac{1}{2}\beta_i(t_n)\Delta t^2, & \text{if } d_i(t_n)=0,\\ x_i(t_n)+\dfrac{\beta_i(t_n)}{d_i(t_n)}\Delta t-\dfrac{1}{d_i(t_n)}\left(v_i(t_n)-\dfrac{\beta_i(t_n)}{d_i(t_n)}\right)\{\exp[-d_i(t_n)\Delta t]-1\}, & \text{if } d_i(t_n)\neq 0,\end{cases} \tag{86}$$

where $x_1\equiv x$, $x_2\equiv y$, and $x_3\equiv z$. Integration of (83*b*) twice from t_n leads to the change of rotation angles (φ_2,φ_3) at t_{n+1},

$$\Delta\varphi_i(t_{n+1})=\begin{cases} \omega_i(t_n)\Delta t+\frac{1}{2}\gamma_i(t_n)\Delta t^2, & \text{if } b_i(t_n)=0,\\ \dfrac{\omega_i(t_n)}{b_i(t_n)}\Delta t-\dfrac{1}{b_i(t_n)}\left(\varphi_i(t_n)-\dfrac{\omega_i(t_n)}{b_i(t_n)}\right)\{\exp[-b_i(t_n)\Delta t]-1\}, & \text{if } b_i(t_n)\neq 0,\quad i=2,3.\end{cases} \tag{87}$$

Equations (84) and (85) are the recursive formulas for predicting the cylinder's translation velocity (u,v,w) and angular velocity $(\omega_1,\omega_2,\omega_3)$, and Eqs. (86) and (87) are the recursive formulas for predicting the cylinder's translation position (x,y,z) and rotation angle increments $(\Delta\varphi_2,\Delta\varphi_3)$ in the M-coordinate system. The cylinder's orientation is represented by angles (ψ_1,ψ_2,ψ_3) with $\psi_1=\varphi_1$, and a relationship between $(\Delta\psi_2,\Delta\psi_3)$ and $(\Delta\varphi_2,\Delta\varphi_3)$ given by (10).

Let $[x(t_0),y(t_0),z(t_0),u(t_0),v(t_0),w(t_0)]$ be the cylinder's initial translation and velocity and $[\psi_1(t_0),\psi_2(t_0),\psi_3(t_0),\omega_1(t_0),\omega_2(t_0),\omega_3(t_0)]$ be the cylinder's initial orientation and angular velocity. Following the procedures listed in Fig. 8, the values of these variables for next time step $(t=t_1)$ are calculated. Repeat of the procedures leads to predicting the cylinder's position and orientation as falling through the water column.

Theoretically, the model integration stops when the vertical coordinate of COM [i.e., $z(t)$] in the E-coordinate and the elevation angle $\psi_2(t)$ in the M-coordinate do not change with time (in the sediment column):

$$\frac{dz}{dt}=0,\quad \frac{d\psi_2}{dt}=0. \tag{88}$$

Practically, the following criteria are used to stop the integration,

$$\left|\frac{dz}{dt}\right|\leqslant\varepsilon_1,\quad \left|\frac{d\psi_2}{dt}\right|\leqslant\varepsilon_2, \tag{89}$$

where $(\varepsilon_1,\varepsilon_2)$ user-defined small values. In this study, we use

$$\varepsilon_1=10^{-6}\text{m},\quad \varepsilon_2=10^{-4}.$$

8 Cylinder Drop Experiments

Two cylinder drop experiments were conducted to collect data for the model evaluation. Exp-1 was designed to collect data on a cylinder's motion in the water column for various combinations of the cylinder's parameters. Exp-2 was designed to collect synchronized data on sediment parameters (shear strength and density) and the cylinder's burial depth and orientation.

8.1 Exp-1. Exp-1 was conducted at the Naval Postgraduate School swimming pool in June 2001 [16]. It consisted of dropping each of three model cylinders (Fig. 9) into the water where each drop was recorded underwater from two viewpoints. The physical parameters of the model cylinders are listed in Table 2. Figure 10 depicts the overall setup. The controlled parameters for each drop were L/R ratio, χ-value, initial velocity ($\mathbf{V}_{\text{in}}$), and drop angle. The E-coordinate system is chosen with the origin at the corner of the swimming pool with the two sides as x and y axes and the vertical z axis. The initial injection of cylinders was in the (y,z) plane

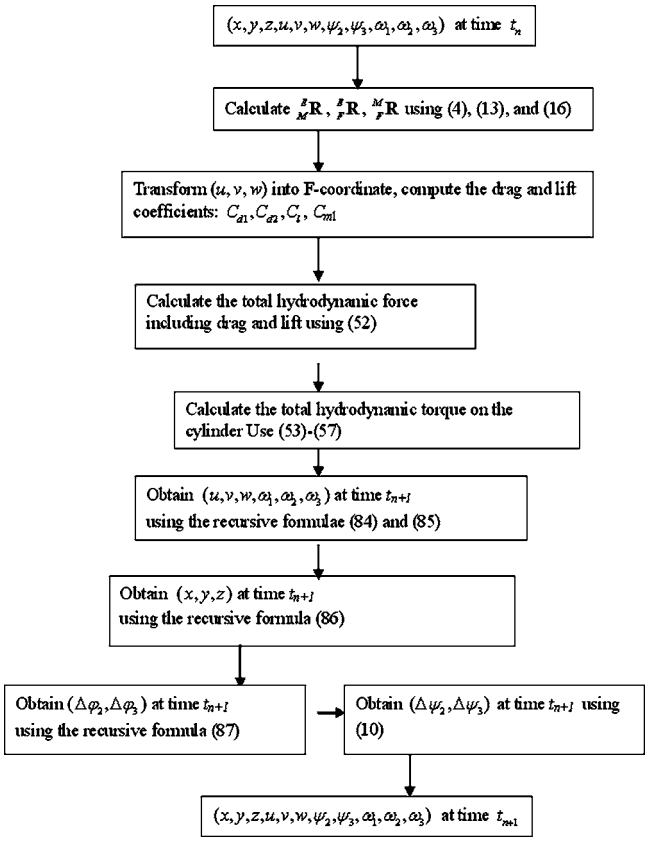


Fig. 8 Procedure of the recursive model

(Fig. 11).

Initial velocity ($\mathbf{V}_{in}$) was calculated by using the voltage return of an infrared photo detector located at the base of the cylinder injector. The infrared sensor produced a square wave pulse when no light was detected due to blockage caused by the cylinder's passage. The length of the square wave pulse was converted into time by using a universal counter. Dividing the cylinder's length by the universal counter's time yielded $\mathbf{V}_{in}$. The cylinders were dropped from several positions within the injector mechanism in order to produce a range of $\mathbf{V}_{in}$. The method used to determine $\mathbf{V}_{in}$ required that the infrared light sensor be located above the water's surface. This distance was held fixed throughout the experiment at 10 cm.

The drop angle (initial value of $\psi_2^{(in)}$) was controlled using the drop angle device. Five screw positions marked the 15 deg, 30 deg, 45 deg, 60 deg, and 75 deg. The drop angles were determined from the lay of the pool walkway, which was assumed to be parallel to the water's surface. A range of drop angles was chosen to represent the various entry angles that air- and surface-laid mines exhibit in naval operation. This range produced velocities whose horizontal and vertical components varied in magnitude. This allowed for comparison of cylinder trajectory sensitivity with the varying velocity components.

For each drop the cylinder was set to a χ value. For positive χ value, the cylinders were placed into the injector so that the COM was located below the geometric center. For negative χ value, the

Table 2 Physical parameters of the model cylinders

Cylinder	Mass (g)	L (cm)	Volume (cm^3)	ρ_m ($g\ m^{-3}$)	J_1 ($g\ m^2$)	χ (cm)	$J_2(J_3)$ $\times(g\ m^2)$
1	322.5	15.20	191.01	1.69	330.5	0.00	6087.9
						0.74	5783.0
						1.48	6233.8
2	254.2	12.10	152.05	1.67	271.3	0.06	3424.6
						0.53	3206.5
						1.00	3312.6
3	215.3	9.12	114.61	1.88	235.0	0.00	1695.2
						0.29	1577.5
						0.58	1556.8

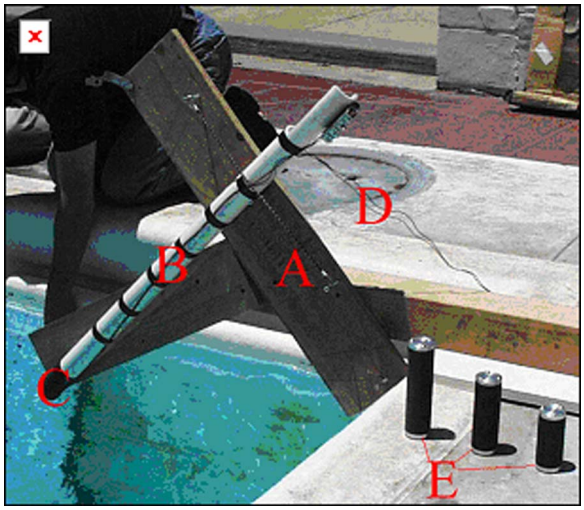


Fig. 10 Exp-1 equipments

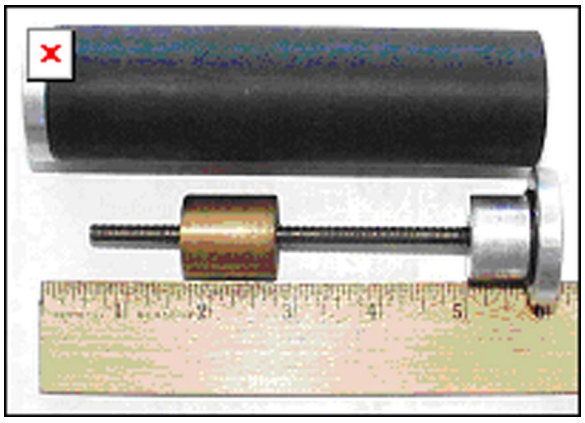


Fig. 9 Internal components of the model cylinder

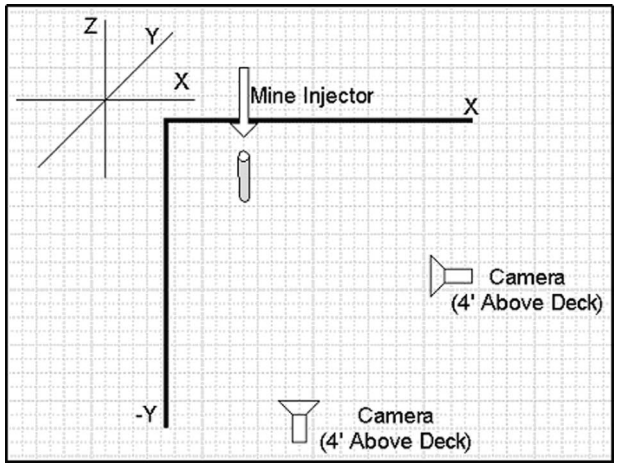


Fig. 11 Top view of Exp-1

Table 3 Number of drops conducted for different drop angles and χ values for $L/R=15/2$

$\psi_2^{(in)}$	15 deg	30 deg	45 deg	60 deg	75 deg
χ_2	13	15	15	15	12
χ_1	9	15	15	15	9
χ_0	12	14	15	18	6
χ_{-1}	0	6	6	6	0
χ_{-2}	2	6	6	0	0

COM was located above the geometric center to release. A series of drops was then conducted in order of decreasing mine length for each angle. Table 3 indicates number of drops conducted for different drop angles and χ value for $L/R=15/2$. The number of drops for other L/R ratios $(12/2, 9/2)$ is comparable to that for the L/R ratio of 15/2. All together there were 712 drops. Each video camera had a film time of approximately 1 h. At the end of the day, the tapes were replayed in order to determine clarity and optimum camera position.

Upon completion of the drop phase, the video from each camera was converted to digital format. The digital video for each view was then analyzed frame by frame (30 Hz) in order to determine the mine's position in x-z and y-z planes. The cylinder's top and bottom positions were input into a MATLAB generated grid, similar to the ones within the pool. The first point to impact the water was always plotted first. This facilitated tracking of the initial entry point throughout the water column. The cameras were not time synchronized; thus, the first recorded position corresponded to when the full length of the mine was in view.

8.2 Exp-2. Exp-2 was conducted on the *R/V John Martin* on May 23, 2000 [17]. The barrel with density ratio of 1.8 was released horizontally while touching the surface. The initial conditions are

$$\mathbf{V}_{\text{in}}=0, \quad \psi_2^{(in)}=90^\circ, \tag{90}$$

This would be to eliminate any chance of inertial effects caused by uneven introduction into the air-sea interface. This also set the initial velocity parameter in the code to zero. The barrel was to be released 17 times. The diver would snap the quick-release shackle on the barrel and then dive down to conduct measurements. The average depth of the water was 13 m. Since it was uncertain the path the barrel would follow, both the releasing diver and a second safety diver would stay on the surface until after the barrel had dropped. Once reaching the bottom, one diver would take penetration measurements using a meter stick marked at millimeter increments while the other would take a gravity core. After 17 drops, the divers began to run out of air and results were not varying greatly so the decision was made to end the experiment. Upon return to the Monterey Bay Aquarium Research Institute, the gravity cores were taken immediately to the USGS Laboratories in Menlo Park, CA where they were refrigerated until the analysis could be performed on May 31–June 1, 2000.

Analysis of the gravity cores was begun on May 31, 2000 at the USGS Laboratories in Menlo Park, CA. The gravity cores were sliced into 2 cm segments to a depth of 10 cms, and then sliced into 4 cm segments. A fall cone apparatus (Model G-200) was used to determine sediment density $\rho_s(z)$ and shear strength. In the test, it is assumed that the shear strength of sediment at constant penetration of a cone is directly proportional to the weight of the cone, and the relation between undrained shear strength s and the penetration h of a cone of weight Q is given by

$$S(z)=KQ/h^2, \tag{91}$$

where K is a constant which depends mainly on the angle of the cone, but is also influenced by the sensitivity of the clay/sediment. Four different cones are used with this instrument, each one having the measuring range listed in Table 4. The cones are suspended from a permanent magnet. By pressing a knob, the magnet is moved so that the magnetic field is broken momentarily and the cone is released. Measurements are taken of penetration depth and the evolution is repeated five times per sediment slice. These values are then averaged and correlated with a table which gives shear strength. Previous studies [18] showed that the sediment parameters are the most critical element in determining how deep an object was buried when it came to rest. During the experiment at the Monterey Bay, we obtained 17 gravity cores. Sediment bulk density and shear strength profiles (Fig. 12) generally show increase with depth until approximately 6–9 cm below the water-sediment interface.

Table 4 Measuring ranges of the gravity cores

Weight (g)	Apex-angle	Penetration (mm)	Undrained shear strength (kPa)
400	30 deg	4.0–15.0	25–1.8
100	30 deg	5.0–15.0	4–0.45
60	60 deg	5.0–15.0	0.6–0.067
10	60 deg	5.0–20.0	0.10–0.0063

9 Model-Data Comparison

The U.S. Navy has a 2D model in the x-z plane (IMPACT28) to predict the cylinder's trajectory and impact burial. Since the motion of the cylinder is 3D, the impact burial prediction using the 2D model has large errors [19–21]. In this study, a new 3D model (called IMPACT35) is developed on the base of momentum balance (19*a*) and (19*b*) and moment of momentum balance (20) using a triple coordinate transform [8] and cylinder decomposition. To evaluate the value added to the 3D model, comparison among the observed data (from Exp-1 and Exp-2) and predicted data using 2D (IMPACT28) and 3D (IMPACT35) models is conducted. Since position and orientation of the cylinder were tape recorded after it is submerged into the water, the free water surface effect was not detected from Exp-1 and Exp-2.

9.1 Comparison Using Exp-1 Data. Improvement from IMPACT28 to IMPACT35 in predicting cylinders' trajectory and orientation in the water column is verified using the Exp-1 data. Here, we list two examples.

Positive χ (Nose-Down): Cylinder 1 $(L=15.20\ \text{cm}, \rho=1.69\ \text{g cm}^{-3})$ with $\chi=0.74$ cm is injected to the water with the

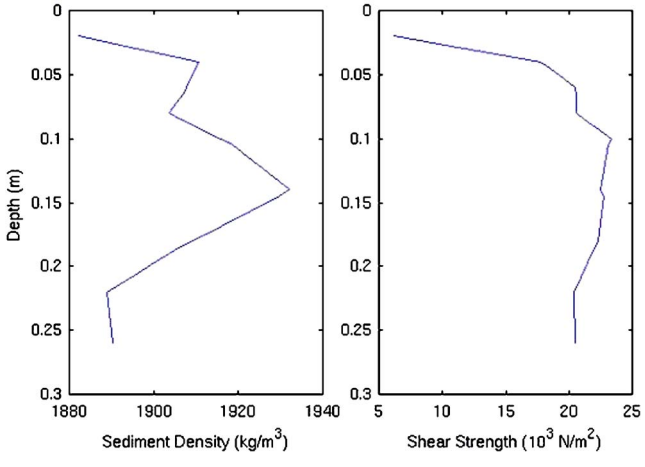


Fig. 12 Mean sediment density $\rho_s(z)$ and shear strength $S(z)$ profiles in the Monterey Bay collected during the cylinder drop experiment on May 31, 2000

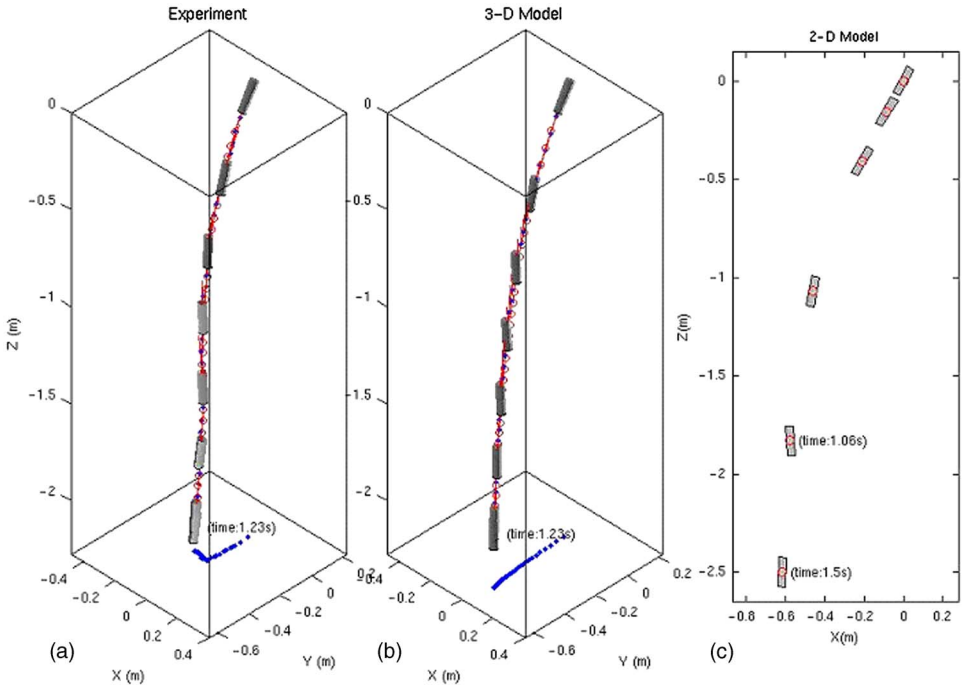


Fig. 13 Movement of cylinder 1 (L=15.20 cm, ρ=1.69 g cm^{-3}) with χ=0.74 m and drop angle 45 deg obtained from (*a*) experiment, (*b*) 3D IMPACT35 model, and (*c*) 2D Impact28 model

drop angle 45 deg. The physical parameters of this cylinder are given by

$$m = 322.5 \text{ g}, \quad J_1 = 330.5 \text{ g cm}^2, \quad J_2 = J_3 = 5783.0 \text{ g cm}^2. \tag{92a}$$

The initial conditions for the numerical models (IMPACT28 and IMPACT35) are taken the same as the experiment (see Sec. 8):

$$x_0 = 0, \quad y_0 = 0, \quad z_0 = 0, \quad u_0 = 0, \quad v_0 = -1.55 \text{ m s}^{-1},$$

$$w_0 = -2.52 \text{ m s}^{-1},$$

$$\psi_{10} = 0, \quad \psi_{20} = 60°, \quad \psi_{30} = -95°, \quad \omega_{10} = 0,$$

$$\omega_{20} = 0.49 \text{ s}^{-1}, \quad \omega_{30} = 0.29 \text{ s}^{-1}. \tag{92b}$$

Substitution of the model parameters (92*a*) and the initial conditions (92*b*) into IMPACT28 and IMPACT35 leads to the predictions of the cylinder's translation and orientation that are com-

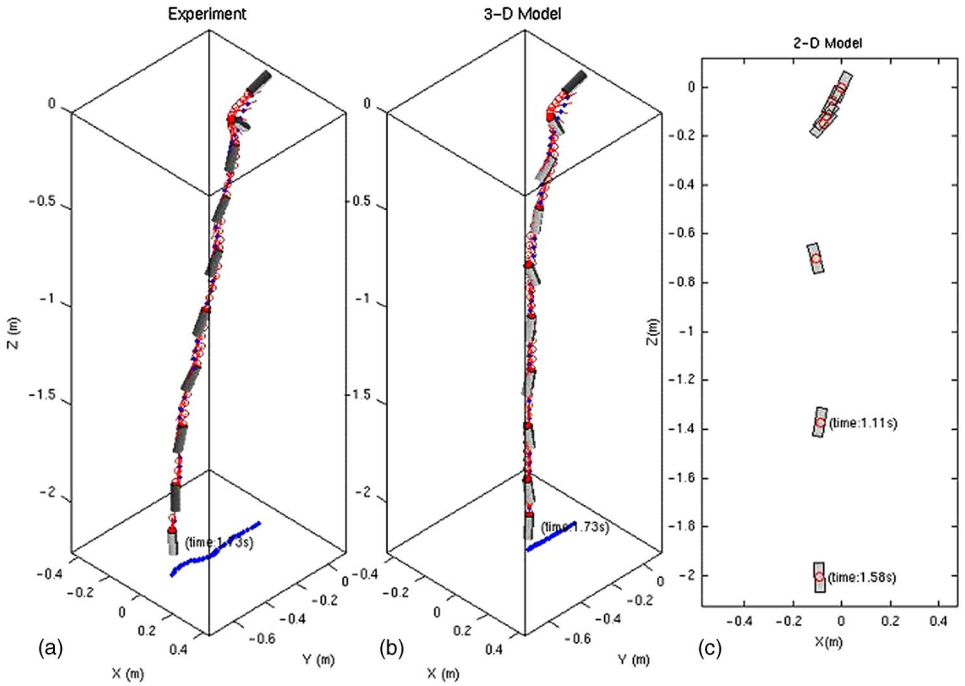


Fig. 14 Movement of cylinder 2 (L=12.10 cm, ρ=1.67 g cm^{-3}) with χ=−1.00 cm and drop angle 30 deg obtained from (*a*) experiment, (*b*) 3D IMPACT35 model, and (*c*) 2D Impact28 model

pared with the data collected during Exp-1 at time steps (Fig. 13). The new 3D model (IMPACT35) simulated trajectory agrees well with the observed trajectory. Both show the same slant-straight pattern and the same travel time (1.23 s) for the cylinder passing through the water column. However, the existing 2D model (IMPACT28) has less capability to predict the cylinder's movement in the water column. The travel time predicted by IMPACT28 is 1.5 s, much more than the observed value.

Negative χ (Nose-Up): Cylinder 2 ($L=12.10$ cm, $\rho=1.67$ g cm^{-3}) with $\chi=-1.00$ cm is injected to the water with the drop angle 30 deg. The physical parameters of this cylinder are given by

$$m=254.2 \text{ g}, \quad J_1=271.3 \text{ g cm}^2, \quad J_2=J_3=3312.6 \text{ g cm}^2. \tag{93a}$$

The initial conditions for the numerical models (IMPACT28 and IMPACT35) are taken the same as the experiment (see Sec. 8):

$$x_0=0, \quad y_0=0, \quad z_0=0, \quad u_0=0, \quad v_0=-0.75 \text{ m s}^{-1},$$
$$w_0=-0.67 \text{ m s}^{-1},$$
$$\psi_{10}=0, \quad \psi_{20}=24^\circ, \quad \psi_{30}=-96^\circ, \quad \omega_{10}=0,$$
$$\omega_{20}=-5.08 \text{ s}^{-1}, \quad \omega_{30}=0.15 \text{ s}^{-1}. \tag{93b}$$

The predicted cylinder's translation and orientation are compared with the data collected during Exp-1 at time steps (Fig. 14). The new 3D model (IMPACT35) simulated trajectory agrees well with the observed trajectory. Both show the same flip-spiral pattern and the same travel time (1.73 s) for the cylinder passing through the water column. The flip occurs at 0.11 s (0.13 s) after the cylinder enters the water in the experiment (IMPACT35). After the flip, the cylinder spirals down to the bottom. However, the existing 2D model (IMPACT28) does not predict the flip-spiral pattern.

9.2 Comparison Using Exp-2 Data. After running the two models (IMPACT35 and IMPACT28) for each gravity core regime $[\rho_s(z), S(z)]$ from the initial conditions (90), the burial depths were compared with measured burial depth data (Fig. 15). As evident, IMPACT35 improves the prediction capability. The existing 2D model (IMPACT25) overpredicts actual burial depth by an order of magnitude on average. However, the 3D model (IMPACT35) predicts the burial depth reasonably well without evident overprediction. Since the gravity cores were taken for approximately 2 to 3 m from the impact location, several cores were taken for each drop. This allowed an average to be calculated in order to yield more accurate data for each drop.

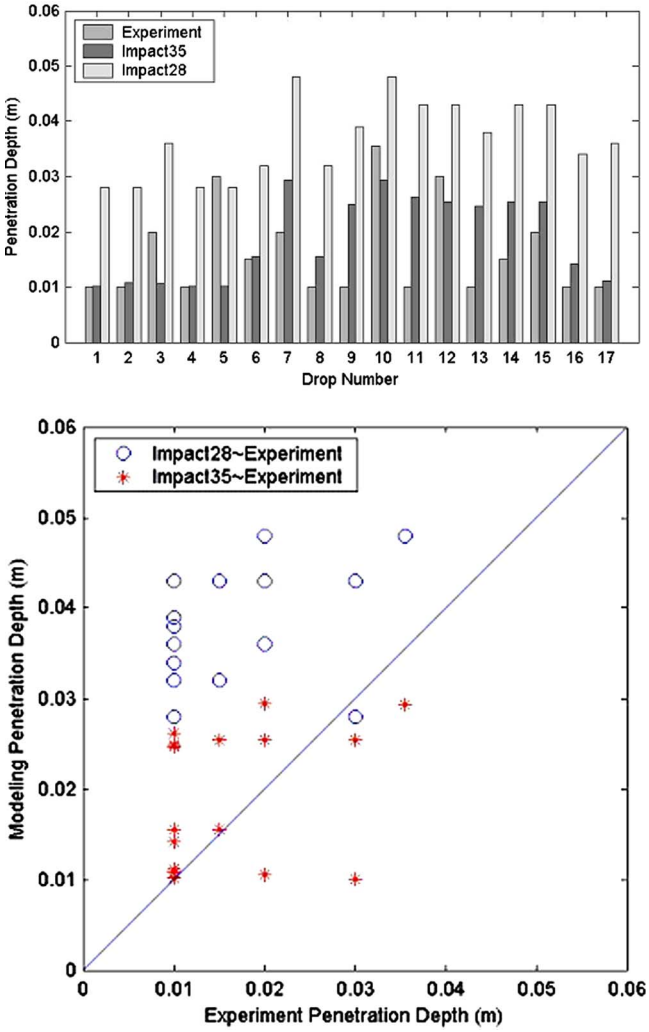


Fig. 15 Comparison among observed and predicted burial depths

10 Conclusions

1. A 3D model (IMPACT35) is developed to predict the translation and orientation of a falling rigid cylinder through air, water, and sediment. It contains three components: triple coordinate transform, cylinder decomposition, and hydrodynamics of a falling rigid object in a single medium (air, water, or sediment) and in multiple media (air-water and water-sediment interfaces).
2. Triple coordinate transform is useful for modeling the movement of a rigid body in air-water-sediment. The body forces (including buoyancy force) and torques are represented in the E-coordinate system, the hydrodynamic forces (such as the drag and lift forces) and torques are represented in the F-coordinate, and the cylinder's moments of gyration are represented in the M-coordinate. The momentum (moment of momentum) equation for predicting the cylinder's translation velocity (orientation) is represented in the E-coordinate (M-coordinate) system. Transformations among the three coordinate systems are used to convert the forcing terms into E-coordinate (M-coordinate) for the momentum (moment of momentum) equation.
3. During the penetration, the part that contacts the fluid (air or water) is treated as an equivalent cylinder with the same mass and PCOV location. The buoyancy and hydrodynamic forces and torques are computed in the equivalent cylinder. The procedure developed for calculating external forcing (buoyancy and hydrodynamic forces and torques) for a single cylinder is used for the equivalent cylinder.
4. Impact force and torque below the water-sediment interface are calculated on the basis of the fact that at the instance of penetration, the sediment exerts an impact force only on the portion of the cylinder's surface, which moves towards the sediment. The normal and tangential components of the impact force are calculated separately. The normal component is calculated using the sediment density and shear strength profiles. The tangential component is computed using the friction law between two solid bodies (i.e., proportional to the normal component). The torque is easily obtained after the impact force is determined.
5. The dynamic system for predicting trajectory and orientation of a rigid cylinder in air, water, and sediment are highly nonlinear. For example, the apparent torque in the moment of momentum equation (20) (represented in the F-coordinate) is nonlinear. The drag and lift forces are nonlinear terms which depend on the square of the fluid-to-body velocity. Two major assumptions are used to simplify the system. First, the apparent torque is neglected. Second, for

the given time step t_n, the nonlinear drag and lift forces and torques are linearized at any time instance with temporally varying coefficients (also dependent on the fluid-to-cylinder velocity) evaluated at the previous time step, t_{n-1}. With the given cylinder's parameters (translation, velocity, orientation, and angular velocity) at the time step t_n, $[x(t_n), y(t_n), z(t_n); u(t_n), v(t_n), w(t_n); \psi_1(t_n), \psi_2(t_n), \psi_3(t_n), \omega_1(t_n), \omega_2(t_n), \omega_3(t_n)]$, the model has analytical solutions at the time step t_{n+1}. The recursive procedure is established to predict the cylinder's translation, velocity, orientation, and angular velocity through air, water, and sediment from the initial conditions. The strength of such treatments guarantees the convergence of the model integration.

Since neglect of the apparent torque is feasible only for slow rotation around the cylinder's main axis (i.e., small self-spin angular velocity ω_1), and since local linearizations of drag and lift forces and torques are feasible for relatively small fluid-to-cylinder velocity, the model might not be valid if ω_1 or the fluid-to-cylinder velocity is large such as for fast water entry and fast self spinning. A fully numerical calculation (rather than the recursive procedure) should be developed for the prediction.

6. Two cylinder drop experiments were conducted to evaluate the 3D model. Model-data comparison shows that IMPACT35 improves the prediction capability drastically with an order of error reduction in the cylinder burial depth, more accurate cylinder track (depth and orientation) prediction, and more accurate travel time of the cylinder through air-water-sediment.

Acknowledgment

The Office of Naval Research Marine Geosciences Program (N0001403WR20178 and N0001404WR20067), Naval Oceanographic Office, and the Naval Postgraduate School supported this study.

Nomenclature

B = length of the sediment rupture line
(C_{d1}, C_{d2}) = drag coefficients along and across the cylinder
C_l = lift coefficient
C_{tl} = translational lift coefficient, kg s^{-1}
d = cylinder diameter, m
e_v = void ratio
(f_1, f_2, f_3) = added-mass ratios for drag and lift forces
f_r = added-mass ratio for moment of drag and lift forces
(f_{rd2}, f_{rd3}) = rotational drag force, N
$\mathbf{F}_b$ = buoyancy force, N
$\mathbf{F}_d$ = drag force, N
$(\mathbf{F}_{d1}, \mathbf{F}_{d2}, \mathbf{F}_{d3})$ = drag force in the F-coordinate, N
$\mathbf{F}_l$ = lift force, N
$(\mathbf{F}_{l1}, \mathbf{F}_{l2}, \mathbf{F}_{l3})$ = lift force in the F-coordinate, N
$\mathbf{F}_{pw}$ = pore water pressure force, N
$(\mathbf{i}_E, \mathbf{j}_E, \mathbf{k}_E)$ = unit vectors in the E-coordinate
$(\mathbf{i}_F, \mathbf{j}_F, \mathbf{k}_F)$ = unit vectors in the F-coordinate
$(\mathbf{i}_M, \mathbf{j}_M, \mathbf{k}_M)$ = unit vectors in the M-coordinate
(J_1, J_2, J_3) = moments of gyration, kg m^2
$(J_1^{(i)}, J_2^{(i)}, J_3^{(i)})$ = moments of gyration for cylindrical part i, kg m^2
k_p = permeability coefficient, m s^{-1}
L = length of the cylinder, m
(l_1, l_2, l_3) = lengths of the cylindrical Parts, m
$(m_1, \ldots, m_6)$ = masses of cylindrical parts, kg
$\mathbf{M}_b$ = torque due to the buoyancy force, kg m^2 s^{-2}
$\mathbf{M}_h$ = torque due to the hydrodynamic force, kg m^2 s^{-2}
$(\mathbf{M}_{d1}, \mathbf{M}_{d2}, \mathbf{M}_{d3})$ = torques due to the drag force in the M-coordinate, kg m^2 s^{-2}
$\mathbf{r}$ = position vector (in the M-coordinate) of point on the cylinder's surface
$\mathbf{r}_{pw}$ = position vector (in the M-coordinate) indicating the location of the cylinder's rupture line
R = radius of the cylinder
(R_1, R_2, R_3) = radii of cylindrical parts, m
Re = Reynolds number
$\mathbf{V}$ = translation velocity, m s^{-1}
$\mathbf{V}_r$ = water-to-cylinder velocity, m s^{-1}
$\mathbf{V}_1$ = component of $\mathbf{V}_r$ along the cylinder, m s^{-1}
$\mathbf{V}_2$ = component of $\mathbf{V}_r$ perpendicular to the cylinder, m s^{-1}
$\mathbf{V}_w$ = water velocity (m s^{-1})
$V^{(in)}$ = initial speed of dropping cylinder, m s^{-1}
ν = molecular viscosity of the water, $m^2 s^{-1}$
Π = volume of the cylinder, m^3
ρ = density of the cylinder kg m^{-3}
ρ_w = density of the water, kg m^{-3}
χ = distance between COM and COV, m
(ψ_1, ψ_2, ψ_3) = angles determining the cylinders' orientation
$\boldsymbol{\omega}$ = angular velocity, s^{-1}
$(\omega_1, \omega_2, \omega_3)$ = angular velocity components in the M-coordinate, s^{-1}
$(\omega_1^F, \omega_2^F, \omega_3^F)$ = angular velocity components in the F-coordinate, s^{-1}

References

[1] White, F. M., 1974, *Viscous Fluid Flow*, McGraw-Hill, New York.
[2] Boorda, J. M., 1999, "Mine Countermeasures - An Integral Part of Our Strategy and Our Forces," *Federation of American Scientists*, Washington DC (http://www.fas.org/man/dod-101/sys/ship/weaps/docs/cnopaper.htm).
[3] Inman, D. L., and Jenkins, S. A., 2002, "Scour and Burial of Bottom Mines, a Primer for Fleet Use. University of California, San Diego," *SIO Reference Series 02-8*, Scripps Institution of Oceanography, University of California, San Diego, CA.
[4] Inman, D. L., and Jenkins, S. A., 2005, "Scour and burial of objects in shallow water," *Encyclopedia of Coastal Science*, M. Schwartz, ed., Springer, New York.
[5] Rennie, S., Brandt, A., and Plant, N. 2004, "Utilization of an Expert System for Predicting Mine Burial: Quantifying Uncertainty," *Proceedings of Sixth International Symposium on Technology and the Mine Problem*, Naval Postgraduate School, Monterey, CA, May 10–13, DVD-Rom.
[6] Haeger, S., 2004, "Operational Ocean Modeling Support for Mine Warfare in Operation Iraqi Freedom," *Proceedings of Sixth International Symposium on Technology and the Mine Problem*, Naval Postgraduate School, Monterey, CA, May 10–13, DVD-Rom.
[7] Elmore, P. A., and Richardson, M. D., 2004, "Regional Mine Burial Prediction Using Monte Carlo and Deterministic Methods," *Proceedings of Sixth International Symposium on Technology and the Mine Problem*, Naval Postgraduate School, Monterey, CA, May 10-13, DVD-Rom.
[8] Chu, P. C., Fan, C. W., Evans, A. D., and Gilles, A. F., 2004, "Triple Coordinate Transforms for Prediction of Falling Cylinder Through the Water Column," ASME J. Appl. Mech., **71**, pp. 292–298.
[9] Crowe, C. T., Roberson, J. A., and Elger, D. F., 2001, *Engineering Fluid Mechanics*, 7th ed., John Wiley & Sons, New York.
[10] Rouse, H., 1938, *Fluid Mechanics for Hydraulic Engineers*, 1st ed., McGraw-Hill, New York.
[11] Von Mises, R., 1959, *Theory of Flight*, 1st ed., Dover, New York, pp. 564–585.
[12] Sumer, B. M., and Fredsøe, J., 1997, *Hydrodynamics Around Cylindrical Structures*, 1st ed., World Scientific, Singapore.
[13] Palmer, A., 1997, "Speed Effect in Cutting and Ploughing," Geotechnique, **49**(3), pp. 285–294.
[14] Simonsen, B. C., and Hansen, N. E. O., 1998, "Protection of Marine Structures by Artificial Islands," *Ship Collision Analysis*, H. Gluver and D. Olsen, eds.,

Balkema, Rotterdam, pp. 201–215.
[15] Hansen, N. E. O, Simonsen, B. C., and Sterndorff, M. J., 1994, "Soil Mechanics of Ship Beaching," *Proceedings on 24th International Conference on Coastal Engineering*, Kobe, Japan, pp. 3030–3044.
[16] Gilles, A. F., 2001, "Mine Drop Experiment," M.S. thesis, Naval Postgraduate School, Monterey, CA, p. 151.
[17] Smith, T. B., 2000. "Validation of the Mine Impact Burial Model Using Experimental Data," M.S. thesis, Naval Postgraduate School, Monterey, CA, p. 156.
[18] Chu, P. C., Gilles, A. F., Fan, C., and Fleischer, P., 2002, "Hydrodynamical Characteristics of a Falling Cylinder in the Water Column," *Advances in Fluid Mechanics*, M. Rahman, R. Verhoeven, and C. A. Brebbia, eds., WIT Press, Southampton, **4**, pp. 163–181.
[19] Taber, V. L., 1999, "Environmental Sensitivity Studies on Mine Impact Burial Prediction Model," M.S. thesis, Naval Postgraduate School, Monterey, CA, p. 50.
[20] Evans, A., 2002, "Hydrodynamics of Mine Impact Burial," M.S. thesis, Naval Postgraduate School, Monterey, CA, p. 470.
[21] Chu, P. C., Gilles, A. F., and Fan, C. W., 2005, "Experiment of Falling Cylinder Through the Water Column," Exp. Therm. Fluid Sci., **29**, pp. 555–568.

Coupling of Extension and Twist in Single-Walled Carbon Nanotubes

Karthick Chandraseker
Graduate Student

Subrata Mukherjee[1]
Professor
Fellow ASME
e-mail: sm85@cornell.edu

Department of Theoretical
and Applied Mechanics,
Kimball Hall,
Cornell University,
Ithaca, NY 14853

This paper presents a study of the deformation behavior of single-walled carbon nanotubes (SWNTs) subjected to extension and twist. The interatomic force description is provided by the Tersoff-Brenner potential for carbon. The rolling of a flat graphene sheet into a SWNT is first simulated by minimizing the energy per atom, the end result being the configuration of an undeformed SWNT. The Cauchy-Born rule is then used to connect the atomistic and continuum descriptions of the deformation of SWNTs, and leads to a multilength scale mechanics framework for simulating deformation of SWNTs under applied loads. Coupled extension and twist of SWNTs is considered next. As an alternative to the Cauchy-Born rule for coupled extension-twist problems, a direct map is formulated. Analytic expressions are derived for the deformed bond lengths using the Cauchy-Born rule and the direct map for this class of deformations. Numerical results are presented for kinematic coupling, for imposed extension and imposed twist problems, using the Cauchy-Born rule as well as the direct map, for representative chiral, armchair and zig-zag SWNTs. Results from both these approaches are carefully compared.
[DOI: 10.1115/1.2125987]

1 Introduction

Carbon nanotubes (CNTs) are extremely thin, hollow cylinders made of carbon atoms. They are very strong, very stiff materials and have good thermal conductivity. Depending on their structural arrangement, CNTs exhibit either conducting or semiconducting electrical behavior. These intriguing properties of CNTs have aroused a strong interest in their possible use in nanoelectro mechanical devices such as nanowires, or as active components in electronic devices such as field-effect transistors.

Some of the remarkable electrical and mechanical properties of CNTs stem from the close relation between carbon nanotubes and graphite. An ideal nanotube can be considered as a hexagonal network of carbon atoms that has been rolled up to make a seamless hollow cylinder [Fig. 1(*a*)]. Single-walled carbon nanotubes (SWNTs) are the fundamental structural units that form the building blocks of multiwalled nanotubes, nanotube ropes and rings. Examples of three kinds of SWNTs: chiral, armchair, and zig-zag, are shown in Figs. 1(*b*)–1(*d*). It is seen that the chiral SWNT has an asymmetry (its atoms lie on helices) which causes it to twist when extended and vice versa. The other two have nominally symmetric structures and do not twist when extended. Kinematic coupling analysis (for extension-twist problems) is carried out in this work in order to understand these phenomena. To the best of the authors' knowledge, this is the first time that such results have been presented for carbon nanotubes.

Deformation analysis of CNTs has been the subject of various studies undertaken to enhance the understanding of their mechanical behavior. Experiments at the nanoscale are still under development, and thus have resulted in a significant spread in the reported values for various mechanical properties (e.g., extensional stiffness) measured experimentally by employing various kinds of microscopy and other techniques (see, e.g., Refs. [2–11]). Natural frequencies of vibrating carbon nanotubes have been measured recently [12]. Atomistic simulations (e.g., [13–28]) have been effective in deformation analyses, but their computational expense limits the size of CNTs that can be studied. Examples of other recent papers that address mechanical properties of nanotubes are [29,30]. In spite of the robustness and economy of continuum mechanics, standard constitutive models cannot be directly applied to CNTs due to their small size, dominant presence of van der Waals forces, and ambiguities associated with the values of elastic moduli and CNT wall thickness. Hence, there is a great need to develop enriched continuum models that include atomistic effects, yet use a continuum framework, thus making reasonably sized calculations tractable.

The present work deals with a nonlinear atomistic-continuum constitutive framework that closely follows the constitutive model developed by Zhang et al. [31] and Jiang et al. [32] (see, also, Liu et al. [33], Tadmor et al. [36]). This approach makes use of the Tersoff-Brenner [34,35] multibody interatomic potential for carbon, and can be outlined as follows: First, the equilibrium positions of the atoms on a nanotube prior to deformation are obtained by establishing a correspondence between a flat sheet of graphene and a rolled-up tube. The positions of the atoms upon imposing a deformation on the CNT can then be found by adopting a modified Cauchy-Born rule [36]. The elasticity tensor (in the context of a hyperelastic continuum analysis) is obtained using an appropriate definition of the strain energy density, taken to be a function of the Lagrangian strain. Next, a particular class of deformation problems—imposed extension with allowable twist; and imposed twist with allowable extension, is considered in this work. It is also possible to use a direct map (instead of the Cauchy-Born rule) for this class of extension-twist problems. Analytic expressions are derived for the deformed bond lengths using the Cauchy-Born rule and the direct map for these deformations. Numerical results for kinematic coupling (e.g., variation of nanotube radius and angle of twist as functions of axial strain), using the Cauchy-Born rule as well as the direct map, for each of two reported parameter sets for the Tersoff-Brenner potential, are obtained and studied. Future plans call for determination of elastic moduli, and stress-strain plots for these extension-twist problems—these results will be compared with those reported in the literature

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received April 10, 2005; final manuscript received August 22, 2005. Review conducted by R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

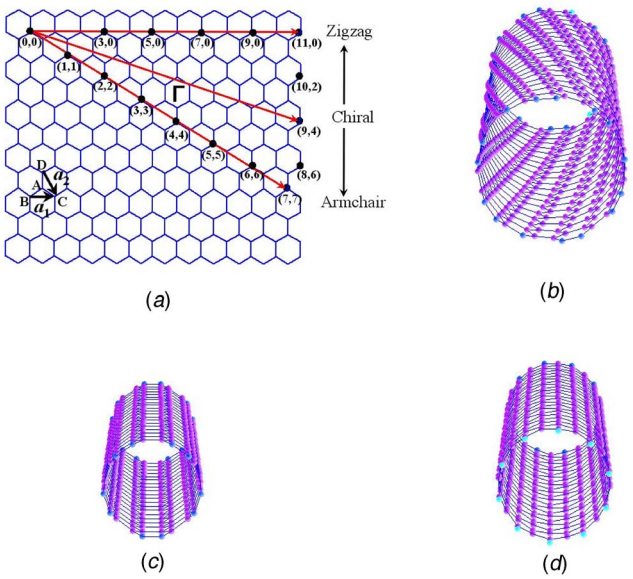


Fig. 1 (*a*) Rolling of a graphene sheet into a SWNT. Single walled carbon nanotubes drawn on the same scale: (*b*) Chiral (9,6); (*c*) Armchair (5,5) (d) Zig-zag (10,0). The figures (*b–d*) have been drawn by using the software in [1]

[31,32]. This work will set the stage for the use of this constitutive model to study the deformation behavior of CNTs under more complex loading conditions.

As mentioned before, the present work follows the excellent research presented in [31–33]. Certain features of the present work, that are different from that reported earlier, are pointed out below.

- The present work clearly follows a two-step process. The rolling of a flat sheet of graphene into a single-walled CNT (SWNT) is first simulated by following the movement of a representative carbon atom and its nearest neighbors, that start out on a flat graphene sheet and move to their new positions on the CNT. Their new positions are found by minimizing the energy per atom. These atoms are then allowed to move (shift) at the onset of deformation, and do so, thereby relaxing to a lower energy state. Next, suitable deformation (extension or twist) is applied to the CNT. Thus, the CNT configuration (both undeformed and deformed), in this work, is cylindrical, while [32] (p. 435) carries out mapping of the cylindrical CNT onto a planar sheet—both initially and during deformation. An energy minimization approach to obtain the cylindrical coordinates of the atoms has been adopted in previous work (e.g., [32]), as well as in the present work.
- A direct map, which is an alternative to the use of the Cauchy-Born rule for this problem, is presented here in addition to the use of the rule. Use of the direct map overcomes certain deficiencies in direct application of the Cauchy-Born rule to a CNT.
- Another important aspect of this work is an a posteriori check on the assumed, analytically convenient, simplified version of the Tersoff-Brenner potential that has been employed here as well as in previous work reported in Refs. [31–33].

The present paper is organized as follows. The Tersoff-Brenner [34,35] interatomic potential for carbon is briefly reviewed first, followed by a discussion of the simulation of rolling of a flat graphene sheet into a SWNT. Next, an enriched continuum mechanics model for the deformation of a SWNT is presented. The noncentrosymmetric structure of graphite is accounted for in this model by introducing the shift vector [36,31]. Simulation of the mechanical behavior of a SWNT under applied extension and applied twist follows. This is carried out for two cases: the Cauchy-Born rule as well as the direct map; and for both sets of parameters of the Tersoff-Brenner potential. Numerical results, a discussion of the results, and concluding remarks, complete the paper.

Table 1 Two sets of Tersoff-Brenner parameters for carbon

Parameter	Value (set 1)	Value (set 2)
$D^{(e)}$	6.325 eV	6.00 eV
S	1.29	1.22
β	15 nm^{-1}	21 nm^{-1}
$R^{(e)}$	0.1315 nm	0.1390 nm
$R^{(1)}$	0.17 nm	0.17 nm
$R^{(2)}$	0.20 nm	0.20 nm
δ	0.80469	0.5
a_0	0.011304	0.00020813
c_0^2	19^2	330^2
d_0^2	2.5^2	3.5^2

2 Constitutive Model for Nonlinear Deformation of Single-Walled Carbon Nanotubes

2.1 Interatomic Potential for Carbon. Expressions for the Tersoff-Brenner potential [35] for carbon are given below. It is noted that a "simplified" version of the Tersoff-Brenner potential is used here—the same model as has been employed in Refs. [31–33]. (Please see Sec. 3.3 for more details.)

Following Brenner [35], an expression for the bonding energy between atoms i and j for carbon is

$$V(r(i,j)) = V_R(r(i,j)) - \bar{B}(i,j)V_A(r(i,j)) \tag{1}$$

where $r(i,j)$ is the bond length (i.e., the distance between carbon atoms i and j at the two ends of the bond), V_R and V_A are the repulsive and attractive pair terms given as:

$$V_R(r) = \frac{D^{(e)}}{S-1}\exp[-\sqrt{2S}\beta(r-R^{(e)})]f_c(r) \tag{2}$$

$$V_A(r) = \frac{D^{(e)}S}{S-1}\exp[-\sqrt{2/S}\beta(r-R^{(e)})]f_c(r) \tag{3}$$

where the parameters $D^{(e)}, S, \beta$, and $R^{(e)}$ (a parameter that is fit to the lattice constant of diamond and graphite) are given in Table 1. The function f_c is a smooth cut-off function that limits the range of the potential. This is given as:

$$f_c(r) = \begin{cases} 1 & r < R^{(1)} \\ (1/2)\left\{1+\cos\left[\dfrac{\pi(r-R^{(1)})}{R^{(2)}-R^{(1)}}\right]\right\} & R^{(1)} \leqslant r \leqslant R^{(2)} \\ 0 & r > R^{(2)} \end{cases} \tag{4}$$

where $R^{(1)}$ and $R^{(2)}$ are cut-off distances that are available in Table 1.

The parameter $\bar{B}(i,j)=(1/2)(B(i,j)+B(j,i))$ in (1) represents a multibody coupling between the bond from atom i to atom j, and the local environment of atom i. This has the form:

$$B(i,j) = \left[1 + \sum_{k\neq i,j} G[\theta(i,j,k)]f_c[r(i,k)]\right]^{-\delta} \tag{5}$$

where δ is a parameter, k denotes other carbon atoms besides i and j, $\theta(i,j,k)$ is the angle between the bonds $i-j$ and $i-k$, and

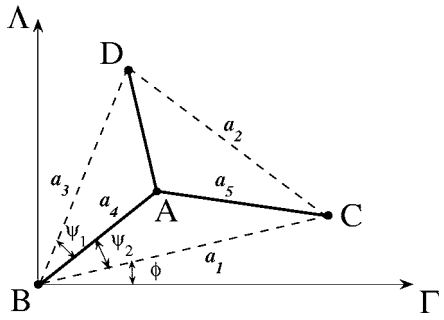


Fig. 2 A representative atom *A* and its nearest neighbors *B*, *C*, *D* and bonds *AB*, *AC*, *AD*. $a_1 = \overrightarrow{BC}$ and $a_2 = \overrightarrow{DC}$

$$G(\theta) = a_0\left[1 + \frac{c_0^2}{d_0^2} - \frac{c_0^2}{d_0^2 + (1+\cos\theta)^2}\right] \tag{6}$$

with, as before, the parameters a_0, c_0, d_0, δ given in Table 1.

Brenner [35] gives two sets of values of the parameters $D^{(e)}, S, \beta$ and $R^{(e)}$ in (2) and (3), δ in (5) and a_0, c_0, d_0 in (6) for carbon. The first set is a good fit (with experiments) for the bond lengths, while the second fits the stretching force constants well. The values of these parameters, as well as those of $R^{(1)}, R^{(2)}$ in (4), are given in Table 1.

2.2 Mapping of a Flat Graphene Sheet Into a SWNT. This section closely follows the discussion in [32]. A representative carbon atom A, and its nearest neighbors B, C, D on a flat graphene sheet, are shown in Fig. 2. The vectors $\mathbf{a}_k, k=1, 2, 3, 4, 5$, shown in the figure, have lengths a_k. The length of the vector $\overrightarrow{AD}$, as well as the angles ψ_1 and ψ_2 in Fig. 2, can be easily obtained as functions of $a_k, k=1, 2, 3, 4, 5$, from the geometry of this figure. In this work, the values of $a_k, k=1, 2, 3, 4, 5$, are obtained from an optimization process that minimizes the potential energy per atom in a SWNT. This matter is discussed in Sec. 2.3.

The graphene sheet is rolled into a cylinder whose circumferential vector is the chiral vector $\mathbf{\Gamma}$. Its length is the circumference of the SWNT, so that the diameter of the SWNT is $d = \Gamma/\pi$, where $\Gamma = |\mathbf{\Gamma}|$. The translational vector $\mathbf{\Lambda}$ is also denoted in Fig. 2. Its magnitude equals the axial period of the CNT. By usual convention,

$$\mathbf{\Gamma} = n\boldsymbol{a}_1 + m\boldsymbol{a}_2 \tag{7}$$

where n and m are integers $(n \geqslant |m| \geqslant 0)$ and (n, m) is the chirality of the CNT; $(n, 0)$ is called zig-zag, (n, n) is called armchair and the general case $n > |m| > 0$ is called a chiral CNT (Fig. 1). Referring to Fig. 2, one can easily show that:

$$\Gamma = \sqrt{n^2a_1^2 + m^2a_2^2 + nm(a_1^2 + a_2^2 - a_3^2)}, \quad \phi = \cos^{-1}\left(\frac{\mathbf{\Gamma}\cdot\boldsymbol{a}_1}{\Gamma a_1}\right) \tag{8}$$

Now consider the mapped positions, in polar coordinates, of the points A, C, D in Fig. 2, in the rolled SWNT. With B as the reference point, one gets

$$R(A) = R(B) = R(C) = R(D) = d/2$$

$$Z(A) = a_4\sin(\psi_2 + \phi), \quad Z(B) = 0, \quad Z(C) = a_1\sin(\phi),$$

$$Z(D) = a_3\sin(\psi_1 + \psi_2 + \phi)$$

$$\Theta(A) = \frac{2a_4\cos(\psi_2 + \phi)}{d}, \quad \Theta(B) = 0, \tag{9}$$

$$\Theta(C) = \frac{2a_1\cos(\phi)}{d}, \quad \Theta(D) = \frac{2a_3\cos(\psi_1 + \psi_2 + \phi)}{d}$$

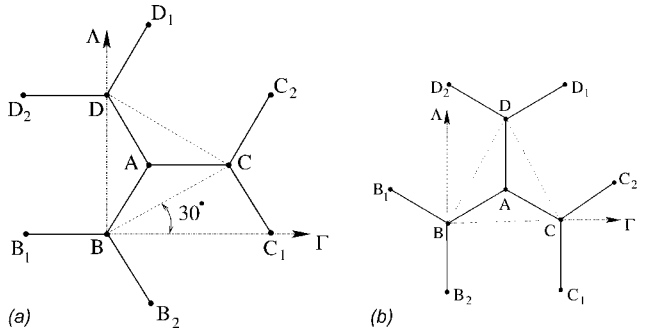


Fig. 3 Atomic structure for (*a*) armchair $r^{(0)}(A,B) = r^{(0)}(A,D), r^{(0)}(C,B) = r^{(0)}(C,D)$ (*b*) zig-zag $r^{(0)}(A,B) = r^{(0)}(A,C), r^{(0)}(D,B) = r^{(0)}(D,C)$ The figures show nearest neighbors *B*, *C*, *D* of atom *A*, and the nearest neighbors of *B*, *C*, *D*

Finally, let $r^{(0)}(X,Y)$ denote the Euclidean distance (which is also the bond length when X and Y are nearest neighbors) between two atoms X and $Y(X, Y = A, B, C, D)$ *when* X, Y *are located on the surface of the SWNT*. One has

$$r^{(0)}(X,Y) = \sqrt{\frac{d^2}{2}\left[1 - \cos(\Theta(Y) - \Theta(X))\right] + (Z(Y) - Z(X))^2} \tag{10}$$

The angle θ in (6) is obtained from $r^{(0)}(X,Y)$ by using the cosine rule.

Given m and n, and the parameters in the Tersoff-Brenner potential; the bond lengths $r^{(0)}(X,Y)$, and the energy per atom (V) associated with atom A (half of the energy from each bond is counted as the energy for atom A):

$$V = (1/2)[V(r(A,B)) + V(r(A,C)) + V(r(A,D))] \tag{11}$$

are now known as functions of $a_k, k=1, 2, 3, 4, 5$. The next step is to minimize the energy for atom A. Therefore, one has

$$\frac{\partial V}{\partial a_k} = 0, \quad k = 1,2,3,4,5 \tag{12}$$

Implementation of Eq. (5), for the atoms B, C, D in Fig. 2 requires special care. In this case, nearest neighbors X_1, X_2 of X, where atom X represents B, C, D (see Fig. 3), must be included as well. The equations needed for including these interactions are listed in Tables 2–5. These equations are motivated by the need to enforce appropriate periodicity conditions. Those in Tables 3 and 4 are correct on a nanotube. Those in Table 2 are correct on a flat graphene sheet and are approximations on a nanotube. Those in Table 5 are assumptions. Equation (5) is implemented in this man-

Table 2 Relationships between angles for all SWNTs (see Fig. 3)

$\theta(B,C,B_1) = \theta(B,A,B_1) + \theta(B,A,C)$
$\theta(B,C,B_2) = \theta(B,A,B_2) - \theta(B,A,C)$
$\theta(B,D,B_1) = \theta(B,A,B_1) - \theta(B,A,D)$
$\theta(B,D,B_2) = \theta(B,A,B_2) + \theta(B,A,D)$
$\theta(C,B,C_1) = \theta(C,A,C_1) - \theta(C,A,B)$
$\theta(C,B,C_2) = \theta(C,A,C_2) + \theta(C,A,B)$
$\theta(C,D,C_1) = \theta(C,A,C_1) + \theta(C,A,D)$
$\theta(C,D,C_2) = \theta(C,A,C_2) - \theta(C,A,D)$
$\theta(D,B,D_1) = \theta(D,A,D_1) + \theta(D,A,B)$
$\theta(D,B,D_2) = \theta(D,A,D_2) - \theta(D,A,B)$
$\theta(D,C,D_1) = \theta(D,A,D_1) - \theta(D,A,C)$
$\theta(D,C,D_2) = \theta(D,A,D_2) + \theta(D,A,C)$

Table 3 Relationships between bond lengths and angles for zig-zag SWNTs [see Fig. 3(*b*)]

$r(B,B_1)=r(C,C_2)=r(D,D_1)=r(D,D_2)=(1/2)(r(A,B)+r(A,C))$
$r(B,B_2)=r(C,C_1)=r(A,D)$
$\theta(B,A,B_1)=\theta(C,A,C_2)=\theta(A,B,C)$
$\theta(B,A,B_2)=\theta(C,A,C_1)=\theta(D,A,D_1)=\theta(D,A,D_2)=(1/2)(\theta(A,B,D)+\theta(A,C,D))$

Table 7 Energy per atom for the starting guesses in Table 6

Type of SWNT	V (eV) for set 1	V (eV) for set 2
(9, 6) Chiral	−7.1174	−7.1849
(5, 5) Armchair	−7.0531	−7.1111
(10, 0) Zig-zag	−7.0809	−7.1431

ner for simulation of rolling of a graphene sheet into a SWNT, as well as for simulation of deformation of SWNTs under imposed extension or twist.

2.3 Numerical Results for Bond Lengths in an Undeformed SWNT. Equation (12) is solved for the parameters $a_k, k=1, 2, 3, 4, 5$, with MATLAB using the subroutine *fminunc*. This is a Quasi-Newton algorithm. In this function, the BFGS method [37–40] is used for choosing the search direction (by updating the Hessian matrix) and a cubic polynomial interpolation (*cubicpoly*) is used, together with gradient and function evaluations, to obtain estimates of step length in a chosen direction. This subroutine *fminunc* is used for all the minimization calculations in the rest of this paper.

The problem is solved for examples of a chiral (9,6), armchair (5,5), and zig-zag (10,0) CNT. A sample starting guess, based on an initial equilateral triangle *BCD* in Fig. 2, is given in Table 6. Please note that $R^{(e)}$, a parameter that is fit to the lattice constant of diamond and graphite, has values of $R^{(e)}=0.1315$ nm for parameter set #1 and $R^{(e)}=0.1390$ nm for parameter set #2 (see Table 1). The starting guesses for a_4 and a_5 in Table 6 are chosen to be equal to the corresponding values of $R^{(e)}$ for the two parameter sets. The corresponding values of the energy per atom, for the three kinds of nanotubes mentioned above, appear in Table 7.

Various starting guesses were tried and the best converged values for $a_k, k=1, 2, 3, 4, 5$, for the two sets of parameters, are given in Tables 8 and 9, respectively. The values of the a_k are very important because these are used to calculate the polar coordinates of the atoms A,B,C,D in the SWNT [from Eqs. (7)–(9)]; and these polar coordinates are inputs for simulations of deformations of the various SWNTs.

It is noted that the values of a_4 and a_5 in Tables 8 and 9 lie between the lengths of single and double carbon-carbon bonds as reported in Table 2 of [35]. This is consistent with the fact that graphene has a resonant bond structure with each of its bonds being "between single and double."

Various other quantities are now calculated from a_k. The corresponding values of the bond lengths $r_{XY}^{(0)}$, in the SWNT, are given in Tables 10 and 11. The values of the energy/atom V, in each case, are also given in these tables. It is noted that the bond lengths are close to the values reported in [32,33,35]. Also, the energy values are close to the binding energy for graphite (−7.3768 eV/atom for parameter set 1 and −7.3756 eV/atom for parameter set 2) as reported in [35]. Finally, the values of the SWNT radius R, and the angle ϕ in Fig. 2, for the two sets of parameters, appear in Tables 12 and 13, respectively. It is noted that the corresponding values, for the two sets of parameters, are quite close to each other.

Table 4 Relationships between bond lengths and angles for armchair SWNTs [see Fig. 3(*a*)]

$r(B,B_2)=r(C,C_1)=r(C,C_2)=r(D,D_1)=(1/2)(r(A,B)+r(A,D))$
$r(B,B_1)=r(D,D_2)=r(A,C)$
$\theta(B,A,B_2)=\theta(D,A,D_1)=\theta(A,B,D)$
$\theta(B,A,B_1)=\theta(C,A,C_1)=\theta(C,A,C_2)=\theta(D,A,D_2)=(1/2)(\theta(A,B,C)+\theta(A,C,D))$

Table 5 Relationships between bond lengths and angles for chiral SWNTs (see Fig. 3)

$r(X,X_1)=r(X,X_2)=(1/3)(r(A,B)+r(A,C)+r(A,D))$
$\theta(X,A,X_1)=\theta(X,A,X_2)=(1/3)(\theta(A,B,C)+\theta(A,B,D)+\theta(A,C,D))$
for X=B,C,D

2.4 Nonlinear Elastic Deformation of SWNTs. The starting point here is the undeformed SWNT with the atom positions described above. The deformation gradient $\mathbf{F}=\partial\mathbf{x}/\partial\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{x}$ denote the positions of a material point in the undeformed and deformed configurations, respectively. The Cauchy-Born rule has been applied to Bravais lattices in [36] and has been used to connect the atomistic and continuum descriptions of a SWNT in [31–33]. It states that the vector defined by a pair of atoms deforms according to the local deformation gradient. Thus, one has

$$\mathbf{r}(i,j)=\mathbf{F}\cdot\mathbf{r}^{(0)}(i,j) \tag{13}$$

where $\mathbf{r}(i,j)=r(i,j)\mathbf{n}(i,j)$ with $r(i,j)$ denoting the bond length between atoms i and j, and $\mathbf{n}(i,j)$ a unit vector along the bond (directed from atom i to atom j), respectively. A superscript (0) denotes the undeformed, and its absence, the deformed configuration.

One can easily show that

$$r^2(i,j)=\mathbf{r}^{(0)}(i,j)\cdot(\mathbf{I}+2\mathbf{E})\cdot\mathbf{r}^{(0)}(i,j) \tag{14}$$

where the Lagrange strain $\mathbf{E}=(1/2)(\mathbf{F}^T\cdot\mathbf{F}-\mathbf{I})$ with $\mathbf{I}$ the second order identity tensor, and $\mathbf{F}^T$ is the transpose of $\mathbf{F}$.

A centrosymmetric lattice is one that has pairs of bonds in opposite directions around each atom. The Cauchy-Born rule ensures equilibrium for such a structure for arbitrary imposed homogeneous deformation. A SWNT, however, is not centrosymmetric, but consists of two different sublattices, each of which is centrosymmetric. A shift vector ζ becomes necessary (when using the Cauchy-Born rule) for a SWNT to relate an atom pair when each of the atoms in the pair lies on different sublattices (see, e.g., [31–33,36]).

Atom A in Fig. 2 lies on one sublattice and B,C,D lie on another. Also, define $\boldsymbol{\zeta}=\mathbf{F}\cdot\boldsymbol{\eta}$. Now, one has the modified Cauchy-Born rule as follows:

$$\mathbf{r}(X,Y)=\mathbf{F}\cdot\mathbf{r}^{(0)}(X,Y)+\boldsymbol{\zeta}=\mathbf{F}\cdot\left(\mathbf{r}^{(0)}(X,Y)+\boldsymbol{\eta}\right) \tag{15}$$

Table 6 Sample starting values for $a_k, k=1, 2, 3, 4, 5$, for two parameter sets, based on an initial equilateral triangle *BCD* (see Fig. 2) on the graphene sheet

Parameter set	a_1 (nm)	a_2 (nm)	a_3 (nm)	a_4 (nm)	a_5 (nm)
1	0.2278	0.2278	0.2278	0.1315	0.1315
2	0.2408	0.2408	0.2408	0.1390	0.1390

Table 8 Optimal values of a_k, k=1, 2, 3, 4, 5, for parameter set 1

Type of SWNT	a_1 (nm)	a_2 (nm)	a_3 (nm)	a_4 (nm)	a_5 (nm)
(9, 6) Chiral	0.2464	0.2462	0.2469	0.1422	0.1427
(5, 5) Armchair	0.2476	0.2476	0.2468	0.1423	0.1437
(10, 0) Zig-zag	0.2484	0.2462	0.2462	0.1428	0.1428

$$r^2(X,Y) = \left(\mathbf{r}^{(0)}(X,Y) + \boldsymbol{\eta}\right) \cdot (\mathbf{I} + 2\mathbf{E}) \cdot \left(\mathbf{r}^{(0)}(X,Y) + \boldsymbol{\eta}\right) \quad (16)$$

when either of atoms X or Y is A; and $\boldsymbol{\eta}=0$ otherwise. It is noted from above that a deformed bond length r, in general, is a function of $\mathbf{E}$ and $\boldsymbol{\eta}$. Strictly speaking, the standard Cauchy-Born rule is valid only for homogeneous deformations at the atomic scale, i.e., when $\mathbf{F}$ is not a function of position. Such is not the case in the problems considered later in this paper in which $\mathbf{F}$ is a function of the cylindrical coordinates. In the present work, Eq. (13) is implemented by evaluating $\mathbf{F}$ at atom i, although the reported results with the Cauchy-Born rule have been verified to remain unchanged even if $\mathbf{F}$ is evaluated at atom j. Further, $\mathbf{F}$ maps the tangent spaces of the undeformed and deformed surfaces at each corresponding point (see [41]). As pointed out in [41], (13) implies an approximation of a tangent vector by a chord vector, and is one of the shortcomings of applying the standard Cauchy-Born rule to a nanotube. Use of a modification like the exponential map [41] is currently underway [42].

2.5 Strain Energy Density in Terms of Interatomic Potentials. The strain energy density W (in a continuum sense) must be related to the interatomic potentials for a SWNT. This relationship is

$$W(\mathbf{E}, \boldsymbol{\eta}(\mathbf{E})) = \frac{V(r(A,B)) + V(r(A,C)) + V(r(A,D))}{2\Omega} \quad (17)$$

where Ω is the area of the triangle BCD in the flat graphene sheet in Fig. 2. (This area is assumed to be preserved in the rolled SWNT.)

It is important to point out that, due to the uncertainty (in the literature) in the value of the thickness t of a SWNT (because this thickness is of atomic dimension), its explicit use is avoided in this work. The strain energy density W in (17), therefore, is defined per unit area (rather than per unit volume). A consequence of this choice is that the stress and elastic modulus tensors $\mathbf{T}$ and $\mathbf{C}$, defined below in (18) and (19), are actually stress × thickness and modulus × thickness, respectively.

Equation (17) can be understood by referring to Fig. 4. In view of the two atom basis for graphene, its entire structure can be generated by replicating the parallelogram in Fig. 4. Such a cell includes one full and four half bonds, so that, if the bonding energy of each atom pair equals V, one gets $W=3V/2\Omega$; 2Ω being the area of the parallelogram [43]. It is important to mention here that the dependence of W on $\mathbf{F}$ must satisfy material frame indifference. $W(\mathbf{C})$ (where $\mathbf{C}=\mathbf{F}^T\cdot\mathbf{F}$) is both necessary and sufficient for satisfaction of this requirement.

2.6 Stress and Elastic Moduli. The PKII stress tensor $\mathbf{T}$ is obtained by differentiating the strain energy density W with respect to the Lagrange strain $\mathbf{E}$ as follows:

$$\mathbf{T} = \frac{dW}{d\mathbf{E}} = \frac{\partial W}{\partial \mathbf{E}} + \frac{\partial W}{\partial \boldsymbol{\eta}}\frac{d\boldsymbol{\eta}}{d\mathbf{E}} = \frac{\partial W}{\partial \mathbf{E}} \quad (18)$$

since the shift vector is chosen such that $\partial W/\partial \boldsymbol{\eta}=0$.

For this problem, the elasticity tensor $\mathbf{C}$ at the onset of deformation has the form [32]:

$$\mathbf{C} = \left.\frac{d\mathbf{T}}{d\mathbf{E}}\right|_{\mathbf{E}=\mathbf{0}} = \left[\frac{\partial^2 W}{\partial \mathbf{E}\,\partial \mathbf{E}} - \frac{\partial^2 W}{\partial \mathbf{E}\,\partial\,\boldsymbol{\eta}} \cdot \left[\frac{\partial^2 W}{\partial \boldsymbol{\eta}\,\partial\,\boldsymbol{\eta}}\right]^{-1} \cdot \frac{\partial^2 W}{\partial \boldsymbol{\eta}\,\partial \mathbf{E}}\right]_{\mathbf{E}=\mathbf{0}} \quad (19)$$

It is noted in the above definition that $\mathbf{C}=d\mathbf{T}/d\mathbf{E}|_{\mathbf{E}=0}$ is only true when the reference configuration is stress free.

3 A SWNT Under Tension and Torsion

The deformation of a SWNT under tension and torsion is discussed in this section.

Table 9 Optimal values of a_k, k=1, 2, 3, 4, 5, for parameter set 2

Type of SWNT	a_1 (nm)	a_2 (nm)	a_3 (nm)	a_4 (nm)	a_5 (nm)
(9, 6) Chiral	0.2518	0.2516	0.2522	0.1453	0.1458
(5, 5) Armchair	0.2529	0.2529	0.2520	0.1454	0.1468
(10, 0) Zig-zag	0.2538	0.2515	0.2515	0.1459	0.1459

Table 10 Undeformed SWNT bond lengths and energy/atom values, obtained from the optimal values of a_k, for parameter set 1

Type of SWNT	$r_{AB}^{(0)}$ (nm)	$r_{AC}^{(0)}$ (nm)	$r_{AD}^{(0)}$ (nm)	$r_{BC}^{(0)}$ (nm)	$r_{CD}^{(0)}$ (nm)	$r_{DB}^{(0)}$(nm)	V (eV)
(9, 6) Chiral	0.1421	0.1423	0.1421	0.2447	0.2452	0.2469	–7.3425
(5, 5) Armchair	0.1422	0.1427	0.1422	0.2445	0.2445	0.2468	–7.2982
(10, 0) Zig-zag	0.1424	0.1424	0.1421	0.2443	0.2460	0.2460	–7.3174

Table 11 Undeformed SWNT bond lengths and energy/atom values, obtained from the optimal values of a_k, for parameter set 2

Type of SWNT	$r_{AB}^{(0)}$ (nm)	$r_{AC}^{(0)}$ (nm)	$r_{AD}^{(0)}$ (nm)	$r_{BC}^{(0)}$ (nm)	$r_{CD}^{(0)}$ (nm)	$r_{DB}^{(0)}$(nm)	V (eV)
(9, 6) Chiral	0.1452	0.1453	0.1452	0.2500	0.2506	0.2522	–7.3359
(5, 5) Armchair	0.1453	0.1457	0.1453	0.2498	0.2498	0.2520	–7.2844
(10, 0) Zig-zag	0.1454	0.1454	0.1452	0.2496	0.2513	0.2513	–7.3067

Table 12 Values of radius R and angle ϕ, for various undeformed SWNTs, for parameter set 1

Type of SWNT	R (nm)	ϕ (°)
(9, 6) Chiral	0.5122	23.4647
(5, 5) Armchair	0.3416	29.8927
(10, 0) Zig-zag	0.3953	0.0

Table 13 Values of radius R and angle ϕ, for various undeformed SWNTs, for parameter set 2

Type of SWNT	R (nm)	ϕ (°)
(9, 6) Chiral	0.5235	23.4563
(5, 5) Armchair	0.3490	29.8869
(10, 0) Zig-zag	0.4039	0.0

3.1 Basic Equations. The deformation gradient evaluated at atom i, $\mathbf{F}(i)$, for this problem has the form [44][2]

$$\mathbf{F}(i) = \frac{r}{R}(\mathbf{e}_{ri}\mathbf{e}_{Ri} + \mathbf{e}_{\theta i}\mathbf{e}_{\Theta i}) + kr\mathbf{e}_{\theta i}\mathbf{e}_{Zi} + (1+\epsilon)\mathbf{e}_{zi}\mathbf{e}_{Zi} \tag{20}$$

where the undeformed and deformed cylindrical coordinates of atom i are $(R, \Theta(i), Z(i))$ and $(r, \theta(i), z(i))$, respectively, with corresponding base vectors $(\mathbf{e}_{Ri}, \mathbf{e}_{\Theta i}, \mathbf{e}_{Zi})$ and $(\mathbf{e}_{ri}, \mathbf{e}_{\theta i}, \mathbf{e}_{zi})$. Therefore, R and r are the undeformed and deformed (here mean) radii of the nanotube. Also, with k the twist per unit undeformed length and ϵ the axial strain, one has (please note that k and ϵ are defined with respect to the *unrelaxed* (i.e., $\boldsymbol{\eta}=0$) undeformed tube):

$$\theta(i) = \Theta(i) + kZ(i), z(i) = (1+\epsilon)Z(i)$$

$$\mathbf{e}_{Ri} = \mathbf{e}_R\cos(\Theta(i)) + \mathbf{e}_\Theta\sin(\Theta(i)),$$

$$\mathbf{e}_{\Theta i} = -\mathbf{e}_R\sin(\Theta(i)) + \mathbf{e}_\Theta\cos(\Theta(i)), \quad \mathbf{e}_{Zi} = \mathbf{e}_Z$$

$$\mathbf{e}_{ri} = \mathbf{e}_{Ri}\cos(kZ(i)) + \mathbf{e}_{\Theta i}\sin(kZ(i)), \tag{21}$$

$$\mathbf{e}_{\theta i} = -\mathbf{e}_{Ri}\sin(kZ(i)) + \mathbf{e}_{\Theta i}\cos(kZ(i)), \quad \mathbf{e}_{zi} = \mathbf{e}_{Zi}$$

where $(\mathbf{e}_R, \mathbf{e}_\Theta, \mathbf{e}_Z)$ are the base vectors corresponding to the reference atom (here atom B, with $\Theta(B)=0$).

Since only the mean radius is of interest in this work, one can use a linear relationship $r=\gamma R$ (where γ is spatially independent), so that one has $dr/dR=r/R$. This assumption is made above to get (20), and also in [44] even though the tube of interest there does not necessarily have a very thin wall.

Use of the assumption $dr/dR=r/R$ can also be motivated as follows. Consider a thin-walled circular cylinder, made of a linearly elastic isotropic material, under pure axial tension. Now, one has (in usual notation):

$$\frac{du}{dR} = \epsilon_{rr} = -(\nu/E)\sigma_{zz} = \epsilon_{\theta\theta} = \frac{u}{R} \tag{22}$$

Since $r=R+u, dr/dR=r/R$ follows from (22).

The corresponding expression for the Lagrangian strain tensor evaluated at atom i is

$$\mathbf{E}(i) = \frac{1}{2}\left(\frac{r^2}{R^2} - 1\right)(\mathbf{e}_{Ri}\mathbf{e}_{Ri} + \mathbf{e}_{\Theta i}\mathbf{e}_{\Theta i}) + \frac{kr^2}{2R}(\mathbf{e}_{\Theta i}\mathbf{e}_{Zi} + \mathbf{e}_{Zi}\mathbf{e}_{\Theta i}) + \frac{1}{2}(k^2r^2 + 2\epsilon + \epsilon^2)\mathbf{e}_{Zi}\mathbf{e}_{Zi} \tag{23}$$

It is useful, at this stage, to consider W as a function of $(r, k, \epsilon, \boldsymbol{\eta})$. It is felt that this point of view is useful for a clear presentation of the rest of this section. The stationary potential energy principle is implemented by minimizing the total internal energy (body forces and nonbonded interactions are neglected in the present work) over the referential domain. Accordingly, the strain energy density must be minimized with respect to the variables r and $\boldsymbol{\eta}$. Therefore, one must have

$$\frac{\partial W}{\partial r} = 0, \frac{\partial W}{\partial \eta_R} = 0, \frac{\partial W}{\partial \eta_\Theta} = 0, \frac{\partial W}{\partial \eta_Z} = 0 \tag{24}$$

The first equation in (24), together with the boundary condition of a traction-free lateral surface (since $T_{RR}=0$ on the inner and outer surfaces of a SWNT, and the nanotube is very thin, it is assumed that $T_{RR}=0$ throughout the nanotube), yields one of the equilibrium equations (in local form) at the location of each atom i:

$$T_{\Theta_i\Theta_i} + 2kRT_{\Theta_iZ_i} + k^2R^2T_{Z_iZ_i} = 0 \tag{25}$$

which is Eq. (41b) in [32].

3.1.1 Imposed Extension With Allowable Twist. In this case, in addition to (24), one has $\partial W/\partial k=0$. This leads to the additional equation:

$$T_{\Theta_iZ_i} + kRT_{Z_iZ_i} = 0 \tag{26}$$

which satisfies one of the local equilibrium equations upon assuming the stresses to be axisymmetric i.e., no Θ dependence (the remaining local equilibrium equation is satisfied trivially in this case).

Therefore, in general (i.e., for $k\neq 0$), the nonzero stress components for this problem are $T_{Z_iZ_i}, T_{\Theta_iZ_i}=T_{Z_i\Theta_i}, T_{\Theta_i\Theta_i}$.[3]

3.1.2 Imposed Twist With Allowable Extension. In this case, in addition to (24), one has $\partial W/\partial\epsilon=0$. This leads to the additional equation:

$$(1+\epsilon)T_{Z_iZ_i} = 0 \tag{27}$$

which satisfies one of the local equilibrium equations upon assuming the stresses to be axisymmetric i.e., no Θ dependence (the remaining local equilibrium equation can easily be shown to be satisfied in this case).

Therefore, in this case, $T_{Z_iZ_i}=0$, and, from (25), one has $T_{\Theta_i\Theta_i} = -2kRT_{\Theta_iZ_i}$. This time, in general, one has the nonzero stress components $T_{\Theta_iZ_i}=T_{Z_i\Theta_i}, T_{\Theta_i\Theta_i}$

3.2 A Direct Map for the Extension-Twist Problem. Direct application of the standard Cauchy-Born rule to nanotubes has the following problems. It has to be modified [36] to account for the noncentrosymmetric structure of a SWNT—this has been done in this work. It is strictly valid for a homogeneous deformation gradient, while $\mathbf{F}$ for the torsion problem in this work is not homogeneous. Finally, as pointed out in [41], it must be modified (as mentioned earlier, this has not been carried out in the present work) for problems in which the bonds are chords (rather than tangents) on curved surfaces (such as on a SWNT). It is, therefore, interesting to compare the results from using the standard Cauchy-Born rule to an approach when this rule is not used. Use of a direct map for the extension-twist problem provides this opportu-

[2]It is noted that the term $(r/R)\mathbf{e}_{ri}\mathbf{e}_{Ri}$ is missing in Eq. (38) of [32] and in Eqs. (25), (29) of [33]. This is because 2 D deformation is considered in these papers while a cylindrical reference configuration is used for deformation analysis of a SWNT in the present work.

[3]Equation (32) ($T_{Z_i\Theta_i}=T_{\Theta_iZ_i}=0$) in [32] is expected to be true for armchair and zig-zag nanotubes but not for a chiral SWNT which tends to twist when extended. It is noted that for this problem, using $\mathbf{T}=J\mathbf{F}^{-1}\cdot\boldsymbol{\sigma}\cdot\mathbf{F}^{-T}$ [where $\boldsymbol{\sigma}$ is the Cauchy stress and $J=\det(\mathbf{F})$], one can show that $T_{\Theta_iZ_i}=T_{Z_i\Theta_i}=(r/R)\sigma_{\theta_iz_i}-[kr^2/(R(1+\epsilon))]\sigma_{z_iz_i}$. Now, if twist is not allowed, one has $k=0, \sigma_{\theta_iz_i}\neq 0$; while if twist is allowed, $\sigma_{\theta_iz_i}=0, k\neq 0$. In either case, in general, $T_{\Theta_iZ_i}\neq 0$. It is noted that numerical results are presented in [32] for only armchair and zig-zag nanotubes, and not for chiral ones.

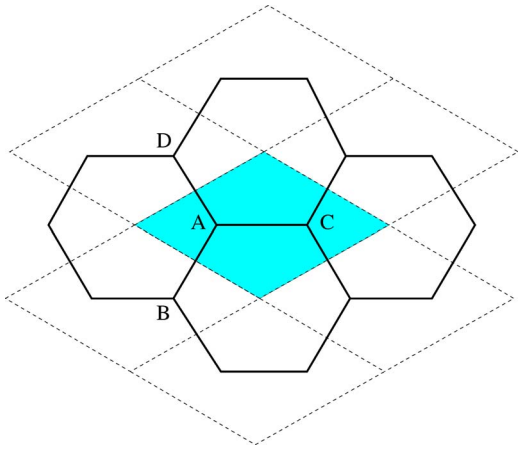


Fig. 4 The shaded parallelogram shown above is the unit cell in a graphene sheet

nity. The direct map is formulated as follows.

Referring to Fig. 5, atom B has cylindrical coordinates $(R,0,0)$ (with respect to the undeformed basis $\{\mathbf{e}_R,\mathbf{e}_\Theta,\mathbf{e}_Z\}$). The position vector of any atom i, in the undeformed configuration, is

$$\mathbf{r}^{(0)}(i) = R\cos(\Theta(i))\mathbf{e}_R + R\sin(\Theta(i))\mathbf{e}_\Theta + Z(i)\mathbf{e}_Z \tag{28}$$

A bond vector between atoms i and j, in the undeformed configuration, is

$$\mathbf{r}^{(0)}(i,j) = \mathbf{r}^{(0)}(j) - \mathbf{r}^{(0)}(i) \tag{29}$$

3.2.1 Cauchy-Born Rule Without Shift Vector. According to the Cauchy-Born rule:

$$r^2(i,j) = \mathbf{r}^{(0)}(i,j)\cdot\mathbf{C}\cdot\mathbf{r}^{(0)}(i,j) \tag{30}$$

where $\mathbf{C}=\mathbf{F}^T\cdot\mathbf{F}$, with the deformation gradient $\mathbf{F}$ given in Eq. (20).

The result is:

$$r^2(i,j) = 2r^2[1-\cos(\Theta(j)-\Theta(i))] + 2kr^2(Z(j)-Z(i))[\sin(\Theta(j) - \Theta(i))] + (Z(j)-Z(i))^2[k^2r^2+(1+\epsilon)^2] \tag{31}$$

3.2.2 Direct Map Without and With Shift Vector. The relevant equations for the direct map without the shift vector are (see table 14):

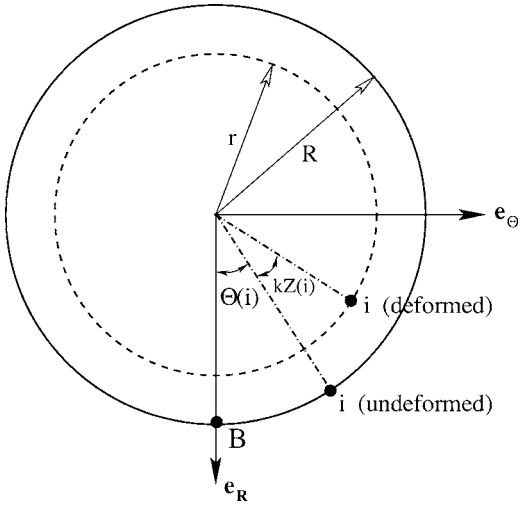


Fig. 5 Undeformed and deformed cross sections of a SWNT

Table 14 The direct map

$R\rightarrow r$
$\Theta\rightarrow\theta=\Theta+kZ$
$Z\rightarrow z=(1+\epsilon)Z$

$$\mathbf{r}(i) = \mathcal{F}(\mathbf{r}^{(0)}(i)) = r[\cos(\Theta(i)+kZ(i))]\mathbf{e}_R + r[\sin(\Theta(i)+kZ(i))]\mathbf{e}_\Theta + (1+\epsilon)Z(i)\mathbf{e}_Z \tag{32}$$

$$\mathbf{r}(i,j) = \mathbf{r}(j) - \mathbf{r}(i) \tag{33}$$

$$r^2(i,j) = |\mathbf{r}(i,j)|^2 = 2r^2[1-\cos(\Theta(j)-\Theta(i))\cos\{k(Z(j)-Z(i))\}] + 2r^2\sin(\Theta(j)-\Theta(i))\sin\{k(Z(j)-Z(i))\} + (1+\epsilon)^2(Z(j) - Z(i))^2 \tag{34}$$

It is noted that the (exact) map expression (34) reduces to the (approximate) Cauchy-Born expression (31) for small values of k. The reason for this is explained in the next paragraph.

The shift vector is now introduced in order to allow for relaxation at the onset of deformation and also for comparison with results from the Cauchy-Born rule with shift. Equation (15) with the shift vector is now replaced by

$$\mathbf{r}(i,j) = \mathcal{F}(\mathbf{r}^{(0)}(i,j) + \boldsymbol{\eta}) \tag{35}$$

and the components of the vector $\boldsymbol{\eta}$ are allowed to be free variables in the energy minimization process as before. Expressions (31) and (34) remain the same even with shift, with Θ replaced by $(\Theta+\eta_\Theta)$, and Z replaced by $(Z+\eta_Z)$.

3.2.3 A Comparison of the Cauchy-Born Rule and the Direct Map. The general case is illustrated in Fig. 6. From $p=\mathcal{F}(P)$, it is seen that:

$$\mathbf{r}(i,j) = \mathcal{F}(\mathbf{X}+\mathbf{r}^{(0)}(i,j)) - \mathcal{F}(\mathbf{X}) \tag{36}$$

By expanding the right-hand side of (36) in a Taylor series about $\mathbf{X}$, and using $\mathbf{F}=\nabla\mathcal{F}$, one has

$$\mathbf{r}(i,j) = \mathbf{F}(\mathbf{X})\cdot\mathbf{r}^{(0)}(i,j) + \frac{1}{2!}\nabla\mathbf{F}(\mathbf{X}):[\mathbf{r}^{(0)}(i,j)\otimes\mathbf{r}^{(0)}(i,j)] + h.o.t. \tag{37}$$

where : denotes the action of a third rank tensor on a second rank tensor that results in a vector, and $\otimes$ denotes the standard tensor product (see, for e.g. [45]). Equation (37) shows that now the deformed bond length vector depends not only on the deformation gradient $\mathbf{F}$, but on its gradients as well. Please note that this still leads to a strain energy density function (for implementation of the stationary potential energy principle) that is frame indifferent under spatially independent, orthogonal rigid body rotations. This

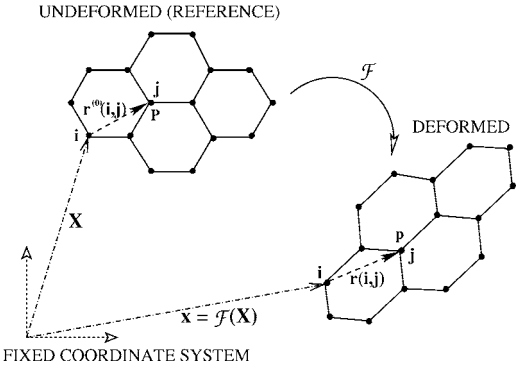


Fig. 6 The direct map $\mathcal{F}(\mathcal{F}(P)=p)$

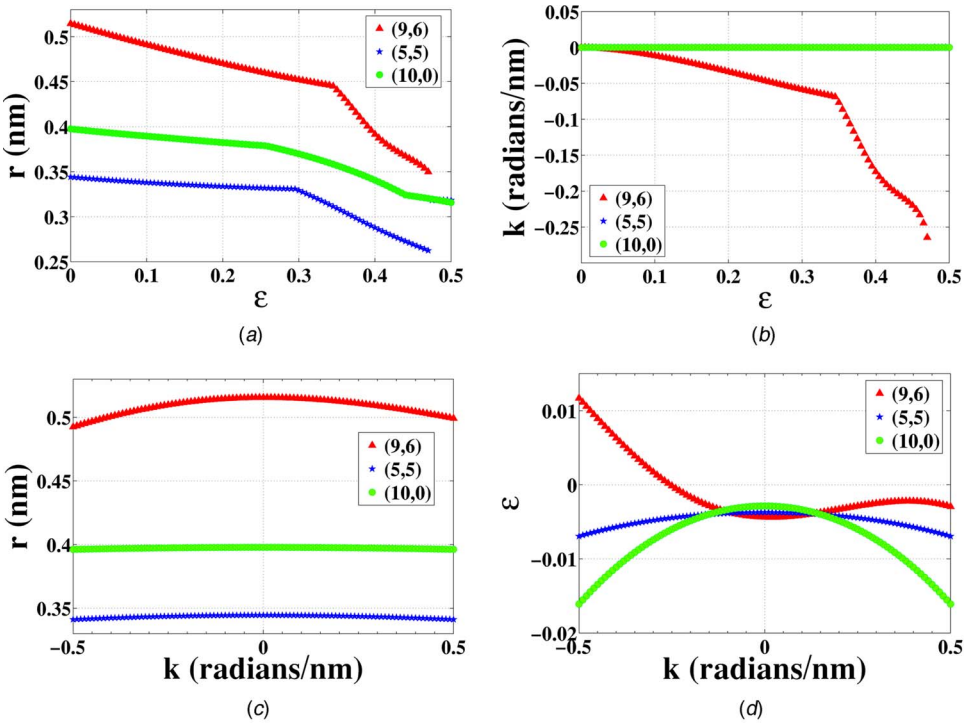


Fig. 7 Kinematic coupling plots for parameter set 1 (*k* for (5,5) and (10,0) coincide in (*b*))

can be seen from the fact that the expressions for the interatomic potential depend only on the magnitudes of the bond lengths (which can easily be shown to be frame indifferent) and the *relative* orientations between bonds (and not on the absolute orientation of any given bond). These comments on frame-indifference hold true even when the shift vector is included (35) because the internal relaxations are performed on the *undeformed* bond vectors. The expressions for the two methods (Cauchy-Born and direct map) coincide if the deformation gradient is homogeneous, so that all its gradients vanish. This is the case for the special situation of pure tension ($k=0$). It is also noted that the expressions in Eqs. (31) and (34) coincide for small k (i.e., when terms of order k^2 and higher are neglected).

3.3 Numerical Results for Imposed Extension and Twist Problems. Numerical results are presented in this section for two SWNT deformation problems: (a) imposed extension with allowable twist and (b) imposed twist with allowable extension. Kinematic coupling plots are presented for each of the Tersoff-Brenner parameter sets in Sec. 3.3.1. Also, there are two sets of results—those obtained by using the Cauchy-Born rule and those from a direct deformation map (see Sec. 3.3.2).

It is emphasized again that a "simplified" version of the Tersoff-Brenner potential [35] is employed in the present work. (This same simplified version has been used previously in Refs. [31–33]). In particular, $F(i,j)$ (called F_{ij} in [35]) is assumed to be zero in the expression for $\bar{B}(i,j)$ (called $\bar{B}_{ij}$ in [35]) in Eq. (10) of [35]. The primary reason for doing this is to be able to obtain analytical derivatives of the potential in the expressions for the PKII stress $\mathbf{T}$ and the elasticity tensor $\mathbf{C}$ in Eqs. (18) and (19) (there is no intrinsic difficulty in using the full Tersoff-Brenner expressions). Calculation of stresses and elastic moduli has not been carried out in the present paper but is planned for the near future. Consequences of using this "simplified" Tersoff-Brenner model are discussed in Sec. 3.3.1.

Another interesting and important observation is made here. As mentioned before in Sec. 2.2, *the carbon atoms are constrained to remain on the cylinder surface* during the simulation of rolling of a flat graphene sheet into a SWNT. At the onset of mechanical deformation, however, they are allowed to move (relax) (i.e., the components of the shift vector $\boldsymbol{\eta}$ are allowed to be nonzero) to a lower energy state. This is important since, in view of Eq. (19), a stress-free configuration must be chosen at the onset of deformation. They do, in fact, move (i.e., one gets nonzero values of the components of the shift vector at the onset of mechanical deformation), and it is possible that some of them migrate away from the cylinder surface. In this case, R and r are viewed as *mean* undeformed and deformed radii of the nanotube, respectively. This issue will be investigated further in the future by visualizing the positions of the atoms at the onset of and during deformation.

3.3.1 Numerical Results From the Cauchy-Born Rule. Numerical results for kinematic coupling are presented in Fig. 7 (for parameter set 1) and in Fig. 8 (for parameter set 2). Such figures have not been presented before in the literature. In each figure, (a) and (b) (for imposed extension with allowable twist) show the imposed strain in the abscissa and the resulting radius and twist per unit length, respectively, in the ordinates. Similarly, (c) and (d) (for imposed twist with allowable extension) show the imposed twist per unit length in the abscissa and the resulting radius and strain, respectively, in the ordinates. Three representative SWNTs, chiral, armchair and zig-zag, are considered in each case. It is noted that the initial atomic structure of the chiral SWNT is asymmetric while those of the other two are nominally symmetric. Also, it is important to note that the results in Figs. 7 and 8 are obtained by *minimizing W (the energy density) without imposing any symmetry constraints*.

It is first observed that the trends are similar for the two parameter sets. Also, certain kinks (in Figs. 7(*a*), 7(*b*), 8(*a*), and 8(*b*)) and jump discontinuities (in Figs. 8(*a*) and 8(*b*)) are observed at relatively large values of strain. These kinks and jumps are related to the use of the "simplified" Tersoff-Brenner model. This issue is discussed later in this section.

Comments on the results presented in Figs. 7 and 8 appear below.

3.3.1.1 Onset of deformation, zero stress, nonzero strain. It is

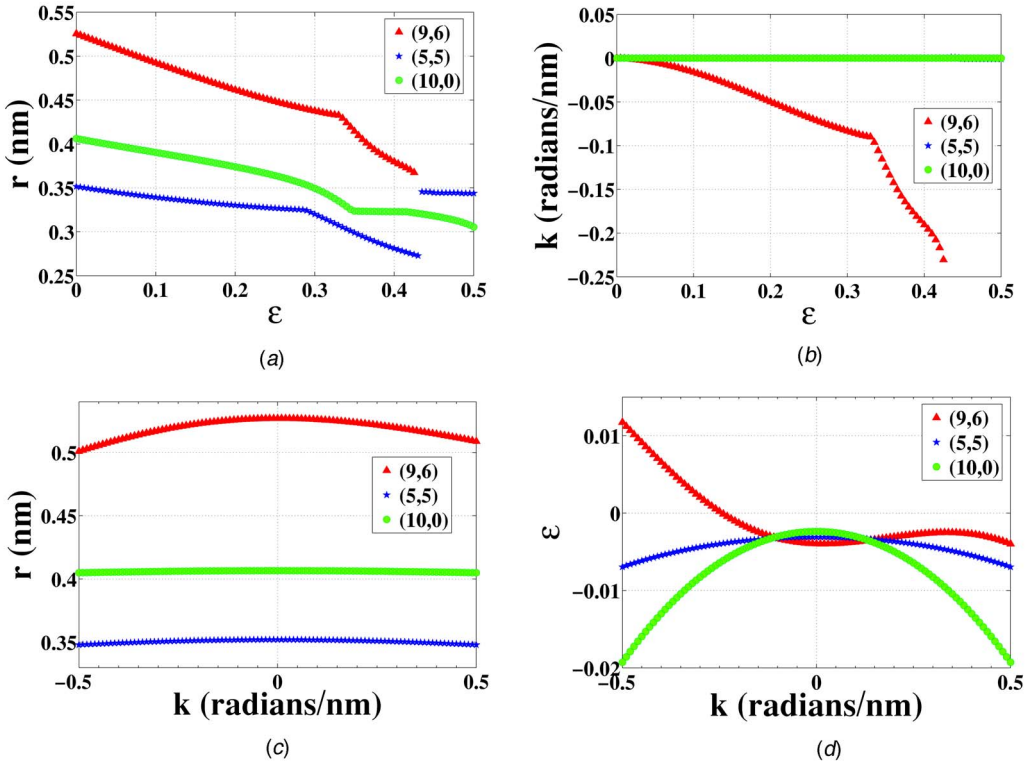


Fig. 8 Kinematic coupling plots for parameter set 2 (*k* for (5,5) and (10,0) coincide in (*b*))

noted that in Figs. 7(*d*), 8(*d*), and 9(*b*)–9(*d*), the SWNTs exhibit nonzero ϵ at $k=0$. The reason for this is as follows. As stated before, ϵ and k are measured with respect to the undeformed, unrelaxed ($\eta=0$) configuration. Relaxation at the onset of deformation causes the atoms to move to the lowest energy state ($\eta \neq 0$) at which the stress is zero but the strain ϵ might no longer be zero. A similar behavior has been reported in [36] (p. 239) for silicon. In fact, if the atoms are constrained to lie on the cylindrical surface (by setting $\eta_R=0$), one obtains $\epsilon=0$ at the onset of deformation ($k=0$). However, this does not ensure a lowest-energy undeformed configuration and hence fully three-dimensional relaxation is allowed in the present work, thereby guaranteeing a stress-free undeformed configuration.

3.3.1.2 Poisson effect. Figure 7(*a*) exhibits the well-known Poisson effect. The behavior in Fig. 7(*c*) can be called a "generalized" Poisson effect. It is noted that, as expected, the radius changes caused by twist [Fig. 7(*c*)] are, in general, much smaller than those caused by extension [Fig. 7(*a*)]. Also, as expected, the curves for the armchair and zig-zag SWNTs in Fig. 7(*c*) are symmetric with respect to k. Similar comments apply to Figs. 8(*a*), 8(*c*), 9(*a*), and 9(*b*). Finally, the chiral nanotube, with an initially asymmetric structure, shows the largest coupling.

3.3.1.3 Twist under imposed ϵ in Figs. 7(b) and 8(b). As expected, the chiral nanotube twists under imposed extension, while the armchair and zig-zag nanotubes do not (i.e., the induced twist for the armchair and zig-zag nanotubes is zero within the plotting scale throughout the range of imposed extension). This is related to the fact that the chiral nanotube has an asymmetric bond structure while the armchair and zig-zag nanotubes are nominally symmetric about the cylinder axis (Fig. 1).

3.3.1.4 Extension under imposed k in Figs. 7(d) and 8(d). It is observed that the chiral nanotube exhibits an asymmetric coupling between extension and twist, for imposed twist. For example, $\epsilon > 0$ for $k<0$ and can be <0 for $k>0$. The armchair and zig-zag nanotubes, however, exhibit a symmetric extension $\epsilon(k)$ about $k=0$ owing to their nominally axisymmetric atomic structure (Fig. 1). Further, the zig-zag (10,0) nanotube exhibits the largest $|\epsilon|$ for large values of $|k|$.

3.3.1.5 A posteriori evaluation of the use of the "simplified" Tersoff-Brenner potentials. It has been mentioned before at the start of Sec. 3.3 that a "simplified" version of the Tersoff-Brenner potential for carbon has been employed in this work. In particular, $F(i,j)$ is assumed to be zero in the expression for $\bar{B}(i,j)$ in Eq. (10) of [35]. An investigation of the relationship of kinks and jumps in some of the figures in Figs. 7 and 8, with the term $F(i,j)$, is carried out below.

Equation (10) of [35] (in slightly modified notation) reads

$$\bar{B}(i,j) = (1/2)[B(i,j) + B(j,i)] + F(N^{(t)}(i), N^{(t)}(j), N^{(\mathrm{conj})}(i,j)) \tag{38}$$

In the above, $N^{(t)}(i)$ is the total number of carbon atoms bonded to carbon atom i and $N^{(\mathrm{conj})}(i,j)$ depends on whether a bond between carbon atoms i and j is part of a conjugated system. They are given by the expressions:

$$N^{(t)}(i) = \sum_j f_c(r(i,j)) \tag{39}$$

$$N^{(\mathrm{conj})}(i,j) = 1 + \sum_{k(\neq i,j)} f_c(r(i,k)) F_c(x(i,k)) + \sum_{\ell(\neq i,j)} f_c(r(j,\ell)) F_c(x(j,\ell)) \tag{40}$$

where:

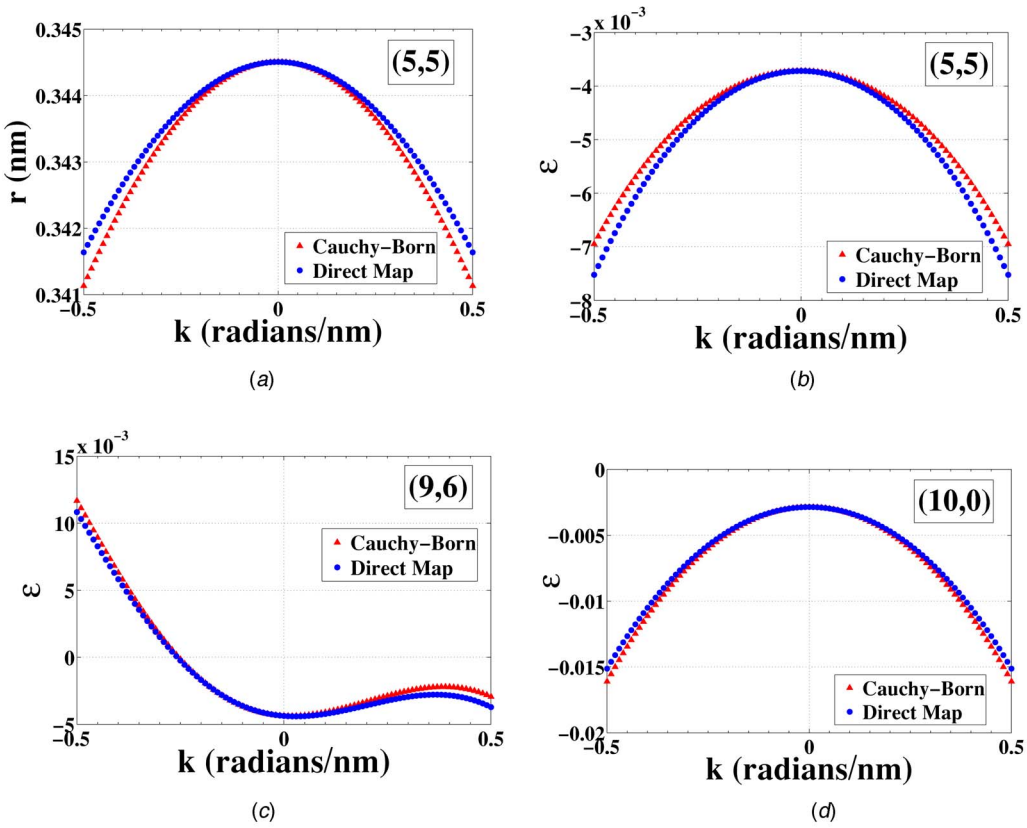


Fig. 9 Comparison of results from the Cauchy-Born rule and the direct map for parameter set 1. Very similar results are obtained for parameter set 2.

$$F_c(x(i,k)) = \begin{cases} 1 & x(i,k) \leqslant 2 \\ (1/2)[1+\cos\{\pi(x(i,k)-2)\}] & 2 < x(i,k) < 3 \\ 0 & x(i,k) \geqslant 3 \end{cases} \tag{41}$$

$$x(i,k) = N^{(t)}(k) - f_c(r(i,k)) \tag{42}$$

and $f_c(r)$ is given in Eq. (4).

In order to satisfy periodicity of the structure of a SWNT, it is assumed in (42) that (see Fig. 3):

$$N^{(t)}_{X_1} = N^{(t)}_{X_2} = N^{(t)}_X, \quad X = B,C,D \tag{43}$$

Table I on p. 9462 of [35] states that $F(2,3,1)=F(2,3,2)=-0.0465, F(1,2,2)=-0.0355$ for parameter set #1; Table III on page 9464 of the same paper states that $F(2,3,1)=F(2,3,2)=-0.0363, F(1,2,2)=-0.0243$ for parameter set #2. Also, $F(i,j,k)=F(j,i,k), F(i,j,k>2)=F(i,j,2)$ (where i,j,k are integers); and other values of F with integer arguments are zero. Values of F with noninteger arguments, however, can be nonzero because 3D splines are used in [35] to define smooth extensions of F near integer arguments.

Using the above equations, the values of $N^{(t)}(X)$ and $N^{(\text{conj})}(X,Y)$ (where X=A, B, C or D) have been calculated for various stages of imposed extension, for each of the three nanotubes considered before (see Figs. 7(*a*), 7(*b*), 8(*a*), and 8(*b*)). It is observed that most of these curves exhibit discontinuities in either the slope, or the value of the function, at certain values of strain. (In fact, certain jumps in values are so large as to go off-scale, and are not shown in these figures.) It has been established that each discontinuity is associated with at least one transition from integer to real or real to integer, of the arguments of the function F in Eq. (38). It can be concluded that the results in these figures portray those from the full Tersoff-Brenner model up to $\epsilon=\epsilon^\star$ in each case. These values are shown in Table 15. It is observed that the minimum transition strain is 26% for the Tersoff-Brenner parameter set #1 and 29% for the Tersoff-Brenner parameter set #2.

The results for imposed twist in Figs. 7(*c*), 7(*d*), 8(*c*), and 8(*d*) display no kinks or jumps. As expected, it has been verified that the function F is zero for these cases, so that the simplified Tersoff-Brenner expressions apply over the entire range $-0.5 \leqslant k \leqslant 0.5$ rad/nm.

3.3.2 Numerical Results From the Direct Map. All the results from the Cauchy-Born rule, shown in Figs. 7 and 8, have been recomputed using the direct map described in Sec. 3.2. The differences between the two sets of results are summarized below.

Results from parameter set #1 are considered first. The results shown in Fig. 7(*a*), 7(*b*), 8(*a*), and 8(*b*) (for imposed extension) remain unchanged while small changes are observed in those shown in Fig. 7(*c*), 7(*d*), 8(*c*), and 8(*d*) (for imposed twist) at large angles of twist. This is consistent with the analytical expressions for deformed bond lengths obtained using the Cauchy-Born rule and the direct map, which have been shown to be identical at

Table 15 Values of smallest transition strains indicating earliest occurrence of transition of arguments of *F* from integer to real values. Results for parameter set #1 appear in Figs. 7(*a*) and 7(*b*) while those from set #2 appear in Figs. 8(*a*) and 8(*b*).

SWNT	Set #1 $\epsilon^\star$	Set #2 $\epsilon^\star$
(9,6)	0.35	0.325
(5,5)	0.30	0.29
(10,0)	0.26	0.35

small angles of twist. In the imposed extension case, the induced angles of twist are not large enough to cause a noticeable difference between the Cauchy-Born and the direct map results for any of the three types of nanotubes (chiral, armchair, and zig-zag). Some of the results from using the Cauchy-Born rule and the direct map for imposed twist cases are compared in Fig. 9. Small differences between the two are observed at large values of k. The (5,5) nanotube, which has the smallest diameter among the three CNTs considered, exhibits the largest difference between the plots obtained from the Cauchy-Born rule and the direct map. The effect of CNT diameter on this difference was studied by obtaining similar comparison plots for a (3,3), (15,15), and a (25,25) nanotube (plots not shown). It has been observed that the difference decreases with increasing CNT diameter—the results from the two models (Cauchy-Born rule and direct map) nearly coinciding for the (25,25) CNT. This observation agrees with the fact that the approximation of a tangent vector by a chord vector (as elucidated in Sec. 2.4) improves with increasing CNT diameter.

Similar behavior is observed for parameter set #2 in all the cases. For both parameter set #1 and set #2, the jumps observed in Figs. 7(a) and 7(b) (as for, Figs. 8(a) and 8(b)) still occur at the same strain locations with the direct map, but the strengths of these jumps generally tend to be different from those observed with the Cauchy-Born rule.

4 Concluding Remarks

The following remarks are in order:

- Results from the two Tersoff-Brenner parameter sets are qualitatively similar but can have some quantitative differences in certain cases.
- An a posteriori check reveals that the analytically convenient, simplified version of the Tersoff-Brenner potential, employed in this work, remains valid for reasonably large strains for the SWNTs considered in this work (at least 26% for parameter set #1 and 29% for parameter set #2) and twist per unit length (± 0.5 rad/nm). Actual strains in SWNTs are expected to be much smaller than these values in most practical applications.
- The results from applying the standard Cauchy-Born rule directly to nanotubes (with no modifications) are found to be in excellent agreement with those from the direct map at small angles of twist and slightly different at larger angles of twist (see Fig. 9). However, the Cauchy-Born rule (to be made more accurate in future work with modifications as suggested in [41]) is very useful for (a) determination of stresses and moduli, and (b) more complicated problems in which a direct map is not available. The Cauchy-Born rule gives $W(\mathbf{C})$ from which the stresses can be obtained in a standard manner, while the direct map leads to a more complicated higher gradient theory with $W(\mathbf{F}, \nabla\mathbf{F}, \ldots)$ which requires a more involved procedure to perform stress analyses.
- Similar calculations with other SWNTs [for example (10,10) and (12,0)] yielded results that are qualitatively similar to those given in this paper.
- Each SWNT structure has been allowed to relax (i.e., the shift vector is allowed to be nonzero) at the onset of deformation. Nonzero shifts (i.e., $|\boldsymbol{\eta}| \neq 0$) are obtained at zero deformation in all cases.

Future work will focus on the use of this model with possible modifications to study more complicated deformations of a SWNT.

Acknowledgment

This research has been supported by Grant No. EEC-0303674 of the National Science Foundation to Cornell University. Several individuals at Cornell have generously contributed to this research through many discussions, and their contributions are gratefully acknowledged: Hande Üstünel and David Roundy from Physics, Yu Mukherjee, and Professors Paul Dawson (Mechanical and Aerospace Engineering) and Timothy Healey (Theoretical and Applied Mechanics).

References

[1] Maruyama, S., Department of Mechanical Engineering, University of Tokyo, http://www.photon.t.u-tokyo.ac.jp/~ maruyama/wrapping3/wrapping.html
[2] Treacy, M. M. J., Ebbesen, T. W., and Gibson. J. M., 1996, "Exceptionally High Young's Modulus Observed for Individual Carbon Nanotubes," Nature (London), **381**, pp. 678–680.
[3] Krishnan, A., Dujardin, E., Ebbesen, T. W., Yianilos, P. N., and Treacy, M. M. J., 1998, "Young's Modulus of Single-Walled Nanotubes," Phys. Rev. B, **58**, pp. 14013–14019.
[4] Wong, E. W., Sheehan, P. E., and Lieber, C. M., 1997, "Nanobeam Mechanics: Elasticity, Strength and Toughness of Nanorods and Nanotubes," Science, **277**, pp. 1971–1975.
[5] Salvetat, J.-P., Briggs, G. A. D., Bonard, J.-M., Bacsa, R. R., and Kulik, A. J., 1999, "Elastic and Shear Moduli of Single-Walled Carbon Nanotube Ropes," Phys. Rev. Lett., **82**, pp. 944–947.
[6] Tombler, T. W., Zhou, C., Alexseyev, L., Kong, J., Dai, H., Liu, L., Jayanthi, C. S., Tang, M., and Wu, S.-Y., 2000, "Reversible Electromechanical Characteristics of Carbon Nanotubes Under Local-Probe Manipulation," Nature (London), **405**, pp. 769–772.
[7] Yu, M.-F., Files, B. S., Arepalli, S., and Ruoff, R. S., 2000, "Tensile Loading of Ropes of Single Wall Carbon Nanotubes and Their Mechanical Properties." Phys. Rev. Lett., **84**, pp. 5552–5555.
[8] Yu, M.-F., Dyer, M. J., Skidmore, G. D., Rohrs, H. W., Lu, X., Ausman, K. D., von Ehr, J. R., and Ruoff, R. S., 1999, "Three-Dimensional Manipulation of Carbon Nanotubes Under a Scanning Electron Microscope," Nanotechnology, **10**, pp. 244–252.
[9] Muster, J., Burghard, M., Roth, S., Duesberg, G. S., Hernández, E., and Rubio, A, 1998, "Scanning Force Microscopy Characterization of Individual Carbon Nanotubes on Electrode Arrays," J. Vac. Sci. Technol. B, **16**, pp. 2796–2801.
[10] Lourie, O., and Wagner, H. D., 1998, "Evaluation of Young's Modulus of Carbon Nanotubes by Micro-Raman Spectroscopy," J. Mater. Res., **13**, pp. 2418–2422.
[11] Pan. Z. W., Xie, S. S., Lu, L., Chang, B. H., Sun, L. F., Zhou, W. Y., Wang, G., and Zhang, D. L., 1999, "Tensile Tests of Ropes of Very Long Aligned Multiwall Carbon Nanotubes," Appl. Phys. Lett., **74**, pp. 3152–3154.
[12] Sazonova, V., Yaish, Y., Üstünel, H., Roundy, D., Arias, T. A., and McEuen, P. L., 2004, "A Tunable Carbon Nanotube Electromechanical Oscillator," Nature (London), **431**, pp. 284–287.
[13] Robertson, D. H., Brenner, D. W., and Mintmire, J. W., 1992, "Energetics of Nanoscale Graphitic Tubules." Phys. Rev. B, **45**, pp. 12592–12595.
[14] Overney, G., Zhong, W., and Tománek, D., 1993, "Structural Rigidity and Low-Frequency Vibrational Modes of Long Carbon Tubules," Z. Phys. D: At., Mol. Clusters, **27**, pp. 93–96.
[15] Molina, J. M., Savinsky, S. S., and Khokhriakov, N. V., 1996, "A Tight-Binding Model for Calculations of Structures and Properties of Graphitic Nanotubes," J. Chem. Phys., **104**, pp. 4652–4656.
[16] Yakobson, B. I., Brabec, C. J., and Bernholc, J., 1996, "Nanomechanics of Carbon Tubes: Instabilities Beyond Linear Response," Phys. Rev. Lett., **76**, pp. 2511–2514.
[17] Cornwell, C. F., and Willi, L. T., 1997, "Elastic Properties of Single-Walled Carbon Nanotubes in Compression," Solid State Commun., **101**, pp. 555–558.
[18] Lu, J. P., 1997, "Elastic Properties of Carbon Nanotubes and Nanopores," Phys. Rev. Lett., **79**, pp. 1297–1300.
[19] Halicioglu, T., 1998, "Stress Calculations for Carbon Nanotubes," Thin Solid Films, **312**, pp. 11–14.
[20] Hernández, E., Goze, C., Bernier, P., and Rubio, A., 1998, "Elastic Properties of C and $B_xC_yN_z$ Composite Nanotubes," Phys. Rev. Lett., **80**, pp. 4502–4505.
[21] Hernández, E., Goze, C., Bernier, P., and Rubio, A., 1999, "Elastic Properties of Single-Wall Nanotubes," Appl. Phys. A: Mater. Sci. Process., **68**, pp. 287–292.
[22] Goze, C., Vaccarini, L., Henrard, L., Bernier, P., Hernández, E., and Rubio, A., 1999, "Elastic and Mechanical Properties of Carbon Nanotubes," Synth. Met., **103**, pp. 2500–2501.
[23] Sánchez-Portal D., Artacho, E., Soler, J. M., Rubio, A., and Ordejón, P., 1999, "*Ab Initio* Structural, Elastic and Vibrational Properties of Carbon Nanotubes," Phys. Rev. B, **59**, pp. 12678–12688.
[24] Lier, G. V., Alsenoy, C. V., Doran, V. V., and Geerlings, P., 2000, "Ab Initio Study of the Elastic Properties of Single-Walled Carbon Nanotubes and Graphene," Chem. Phys. Lett., **326**, pp. 181–185.
[25] Popov, V. N., van Doren, V. E., and Balkanski, M., 2000, "Elastic Properties of Single-Walled Carbon Nanotubes," Phys. Rev. B, **61**, pp. 3078–3084.
[26] Prylutskyy, Y. I., Durov, S. S., Ogloblya, O. V., Buzaneva, E. V., and Scharff P., 2000, "Molecular Dynamics Simulations of Mechanical, Vibrational and Electronic Properties of Carbon Nanotubes," Comput. Mater. Sci., **17**, pp. 352–355.
[27] Vaccarini, L., Goze, C., Henrard, L., Hernández, E., Bernier, P., and Rubio, A., 2000, "Mechanical and Electronic Properties of Carbon and Boron-Nitride Nanotubes," Carbon, **38**, pp. 1681–1690.
[28] Zhou, G., Duan, W., and Gu, B., 2001, "First Principles Study on Morphology

and Mechanical Properties of Single-Walled Carbon Nanotubes," Chem. Phys. Lett., **333**, pp. 344–349.
[29] Friesecke, G., and James, R. D., 2000, "A Scheme for the Passage from Atomic to Continuum Theory for Thin Films, Nanotubes and Nanorods," J. Mech. Phys. Solids, **48**, pp. 1519–1540.
[30] Li, C., and Chou, T.-W., 2003, "A Structural Mechanics Approach for the Analysis of Carbon Nanotubes," Int. J. Solids Struct., **40**, pp. 2487–2499.
[31] Zhang, P., Huang, Y., Geubelle, P. H., Klein, P. A., and Hwang, K. C., 2002, "The Elastic Modulus of Single-Wall Carbon Nanotubes: A Continuum Analysis Incorporating Interatomic Potentials," Int. J. Solids Struct., **39**, pp. 3893–3906.
[32] Jiang, H., Zhang, P., Liu, B., Huang, Y., Geubelle, P. H., Gao, H., and Hwang, K. C., 2003, "The Effect of Nanotube Radius on the Constitutive Model for Carbon Nanotubes," Comput. Mater. Sci., **28**, pp. 429–442.
[33] Liu, B., Jiang, H., Johnson, H. T., and Huang, Y., 2004, "The Influence of Mechanical Deformation on the Electrical Properties of Single Wall Carbon Nanotubes," J. Mech. Phys. Solids, **52**, pp. 1–26.
[34] Tersoff, J., 1988, "New Empirical Approach for the Structure and Energy of Covalent Systems," Phys. Rev. B, **37**, pp. 6991–7000.
[35] Brenner, D. W., 1990, "Empirical Potential for Hydrocarbons for use in Simulating the Chemical Vapor Deposition of Diamond Films," Phys. Rev. B, **42**, pp. 9458–9471.
[36] Tadmor, E. B., Smith, G. S., Bernstein, N., and Kaxiras, E., 1999, "Mixed Finite Element and Atomistic Formulation for Complex Crystals," Phys. Rev. B, **59**, pp. 235–245.
[37] Broyden, C. G., 1970, "The Convergence of a Class of Double-Rank Minimization Algorithms," J. Inst. Math. Appl. **6**, pp. 76–90.
[38] Fletcher, R., 1970, "A New Approach to Variable Metric Algorithms," Comput. J., **13**, pp. 317–322.
[39] Goldfarb, D., 1970, "A Family of Variable Metric Updates Derived by Variational Means," Math. Comput., **24**, pp. 23–26.
[40] Shanno, D. F. 1970, "Conditioning of Quasi-Newton Methods for Function Minimization," Math. Comput., **24**, pp. 647–656.
[41] Arroyo, M., and Belytschko, T., 2002, "An Atomistic-Based Finite Deformation Membrane for Single Layer Crystalline Films," J. Mech. Phys. Solids, **50**, pp. 1941–1977.
[42] Chandraseker, K., Mukherjee, S., Mukherjee, Y. X., 2005, "Modifications to the Cauchy-Born Rule: Applications in the Deformation of Single-walled Carbon Nanotubes" (submitted).
[43] Dresselhaus, M. S., Dresselhaus, G., and Jorio, A., 2004, "Unusual Properties and Structure of Carbon Nanotubes," Annu. Rev. Mater. Res., **34**, pp. 247–278.
[44] Fung, Y. C., and Tong, P., 2001, *Classical and Computational Solid Mechanics*, World Scientific, Singapore, pp. 539–543.
[45] Steigmann, D. J., Ogden, R. W., 1999, "Elastic Surface-Substrate Interactions," Proc. R. Soc. London, Ser. A, **455**, pp. 437–474.

William W. Gerberich
e-mail: wgerb@umn.edu

W. M. Mook

M. D. Chambers

M. J. Cordill

C. R. Perrey

C. B. Carter

Chemical Engineering and Materials Science,
University of Minnesota,
421 Washington Avenue, SE,
Minneapolis, MN 55455

R. E. Miller
Mechanical and Aerospace Engineering,
Carleton University,
Ottawa, ON K1S 5B6, Canada

W. A. Curtin
Division of Engineering,
Brown University,
Providence, RI 02912

R. Mukherjee

S. L. Girshick

Mechanical Engineering,
University of Minnesota,
Minneapolis, MN 55455

An Energy Balance Criterion for Nanoindentation-Induced Single and Multiple Dislocation Events

Small volume deformation can produce two types of plastic instability events. The first involves dislocation nucleation as a dislocation by dislocation event and occurs in nanoparticles or bulk single crystals deformed by atomic force microscopy or small nanoindenter forces. For the second instability event, this involves larger scale nanocontacts into single crystals or their films wherein multiple dislocations cooperate to form a large displacement excursion or load drop. With dislocation work, surface work, and stored elastic energy, one can account for the energy expended in both single and multiple dislocation events. This leads to an energy balance criterion which can model both the displacement excursion and load drop in either constant load or fixed displacement experiments. Nanoindentation of Fe-3% Si (100) crystals with various oxide film thicknesses supports the proposed approach. [DOI: 10.1115/1.2125988]

Introduction

Different types of yield instabilities under nanoscale contacts have been observed for over 30 years [1–3]. One type involves displacement excursions of multiple dislocations occurring multiple times. This has been called "staircase yielding" as shown in Fig. 1(*a*) for single crystal gold [4]. We denote this as a multiple dislocation avalanche event. On the other hand, atomistic simulations of perfect surfaces, such as that shown in Fig. 1(*b*) for aluminum [5], can detect the onset of the first dislocation nucleated. Henceforth, we will denote this type as a single dislocation event. Here, "staircase yielding," which occurs subsequently as multiple load drops due to displacement control, can also be identified with single dislocation events. These events can occur during indentation of perfect surfaces. Surfaces with thin oxide films can also undergo multiple dislocation avalanche events, usually at much higher loads [6]. What we will show in the present study is a simple model for these single dislocation and multiple dislocation avalanche events.

In many metals and semiconductors with oxide films there are in fact two yield points previously identified by nanoindentation which will eventually have to be considered by nanostructural designers. The first yield point of interest, often alluded to experimentally but most often not detected, is the nucleation of the first dislocation. This can occur at extremely early stages of contact at asperities or ledges or at defects in oxide films [3,7–13]. Most often this should be represented by a displacement excursion on the order of a Burgers vector. Originally, experiments such as those run by Gane and Bowden [1], Gane [2], and Gane and Cox [3] associated with either large displacement excursions or resistivity drops were probably not detecting the first dislocation. In later experiments under ultra high vacuum (UHV) the first dislocations were probably detected in Ni [7] since the displacement, showing a plastic response, was only on the order of a nanometer. Much later Page et al. [14] proved by transmission electron microscopy (TEM) that dislocations in Al_2O_3 had been emitted during a displacement excursion. Even later in a series of experiments on Fe-3% Si [12,13], GaAs [15], and tungsten [13] it was indirectly shown that dislocations were injected well below the first obvious yield discontinuity. For example, one could load and unload an Fe-3% Si single crystal to 2.0 mN with a diamond tip

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received August 27, 2004; final manuscript received August 1, 2005. Review conducted by 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 of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

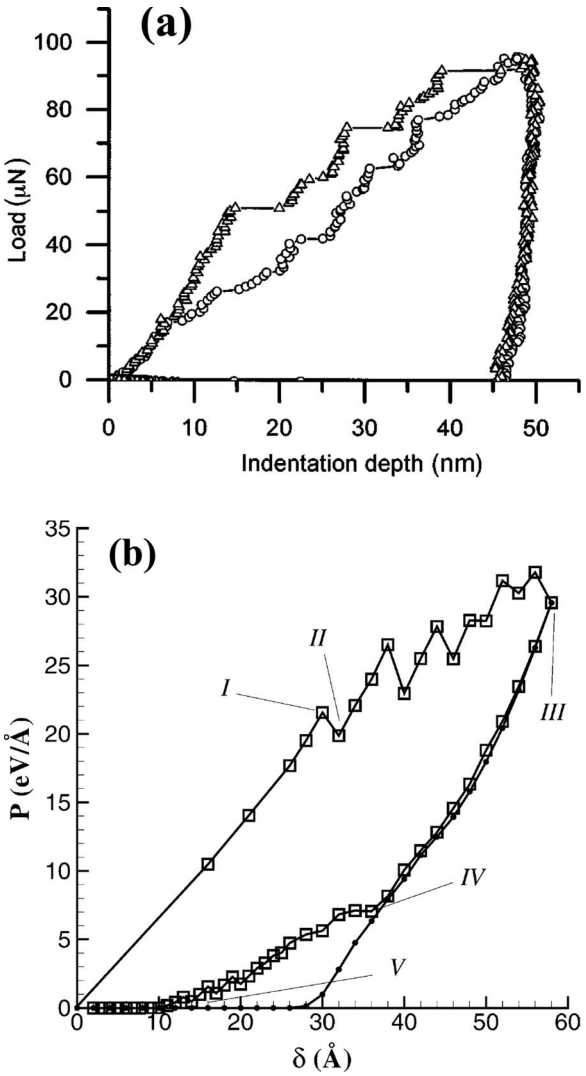


Fig. 1 Examples of "staircase yielding" in both (*a*) experimental [4] and (*b*) simulation [5] results

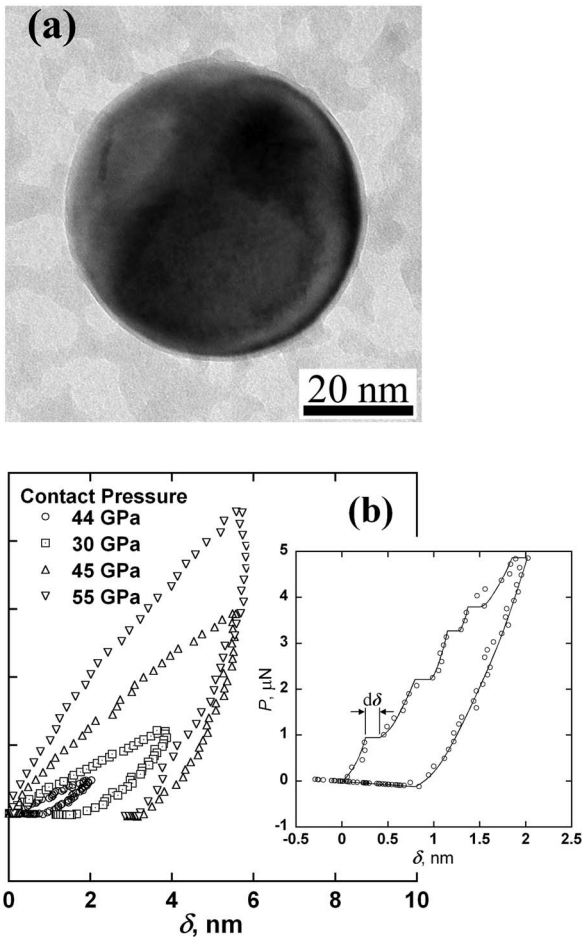


Fig. 2 (*a*) Representative transmission electron micrograph of a silicon nanoparticle. (*b*) Load-displacement results of the repeated compression of a 39 nm diameter silicon nanosphere. The inset shows the initial compression of the nanosphere with distinct displacement excursions.

and the behavior appeared perfectly elastic. However, if the materials were cycled to one-third that load, after about six cycles a yield excursion representative of a large avalanche of dislocations occurred [12]. It was proposed that a sufficient number of dislocations were emitted during each cycle to provide a local backstress to break through a thin oxide film at the Fe-3% Si surface. In the same set of experiments a hold at a constant load half of that required for a monotonically-produced yield excursion produced load relaxation. Permanent displacement was evident upon unloading again suggesting dislocation nucleation prior to the first discontinuous event observed on monotonic loading. At almost the same time Lilleodden et al. [15] observed reverse plasticity in GaAs after unloading from indentations prior to any yield excursion. Following this, Kramer et al. [13] in a series of elegant experiments on W (001) surfaces demonstrated that nanoindentations, produced at loads prior to a well-defined yield excursion, developed a depression. Even more significant, these plastic impressions disappeared by reverse plasticity during repeat scanning. These studies gave evidence that single dislocation events could be followed by multiple dislocation avalanches in materials with thin surface oxide films. On the other hand, in Au with an absence of an oxide film, Corcoran et al. [4] and Houston et al. [16,17] found yield excursions that seemed to imply that the 1.8 GPa shear stresses observed were consistent with dislocation nucleation. This was also consistent with molecular dynamics modeling of homogeneous loop nucleation beneath the contact [18] in Au. Since then a large number of atomistic simulations on Au [18], Ag [19], Al [5,20], and Cu [21] have all indicated either displacement excursions or load drops associated with the first dislocations nucleated (see Fig. 1(*b*)). These ideal situations are seldom if ever achieved experimentally unless one uses ultra high vacuum (with an adhesion problem) or passivates the surface with a self-assembling monolayer [17]. Recently, the evaluation of silicon nanoparticles shown in Fig. 2 strongly suggested that single dislocation events could be detected [22]. In expanded form, this is shown in the Fig. 2(*b*) insert where it is clear that the sum of the displacement excursions is equivalent to the residual displacement on unloading. In a later paper [23] at higher loads, this equivalency breaks down due to reverse plasticity. For the first yield event, then, this most often occurs at very small nN or μN forces and results in subnanometer displacement excursions. This is indicated by both atomistic simulations and nanometer scale contacts of nanospheres and nanocubes of silicon.

The second yield event of interest occurs on oxidized surfaces. Because of thin oxide films initially preventing egress of dislocations, the load at film fracture can be quite large resulting in a dislocation avalanche. We have proposed this occurs by dislocations created at lower loads. At increasing load, their backstress eventually creates a local stress sufficient to break the oxide film resulting in a dislocation avalanche [12]. In these systems, single dislocation nucleation events can still occur which initially do not

lead to dislocation avalanches. This would represent initial yielding. As to the second avalanche event, we will demonstrate that an energy balance criterion can model the behavior associated with an oxide breakthrough mechanism in Fe-3% Si single crystals having different oxide film thicknesses. It is emphasized that both of these events will most often occur at the same contact provided the load is increased to sufficient levels for oxide breakthrough.

To recapitulate then, we propose there are two yield point events in many small volume films and nanostructures covered with thin oxide films. The first is a single dislocation nucleation event which can be important to small force contacts which might compromise the operational characteristics of microelectronic, magnetic, optoelectronic, and microelectromechanical system (MEMS) devices. The second is a dislocation avalanche event which can be important to larger contact forces initially sustainable by oxide films. This represents the yield point most frequently observed except perhaps in gold which is not oxidized, but may have a weakly bonded carbonaceous film at the surface. As the second event can be cycle and time dependent, it is of importance to friction and wear processes in MEMS, shape memory products and biomedical leads which can be exposed to multiple cycles and long lifetimes. The following theoretical background, proposed model and experimental observations address these single dislocation and avalanche events, predominately in Si, Au, Al, and Fe-3% Si.

Single Dislocation Events

Recently, we have been evaluating the plasticity characteristics of nanospheres of silicon and titanium similar to that shown in Fig. 2(*a*). As explained elsewhere [22], these rapidly cooled particles from the melt are generally spherical and free of line and planar defects. Occasionally crystallographic facets at the surface and twin boundaries in the interior are found by TEM [24]. For four reasons these make nearly ideal test specimens for single dislocation events. First, for small diameter nanospheres the contact area at initial displacements is very small allowing for contact stresses large enough to nucleate dislocations but insufficient to break the oxide film away from the contact. As the nanosphere is generally line or planar defect free the initial event is either a single loop nucleated from a step or contact edge at the free surface or one homogeneously nucleated in the interior beneath the contact. Second, no dislocation avalanche forms at low contact forces where the initial plasticity occurs. Third, even at larger plastic strains, avalanches of dislocations have not been detected. This is proposed to be due to the smaller dislocation content and the lack of long-range pileups producing large backstresses sufficient to break oxide films. Fourth, the volumes of the nanospheres are sufficiently small so that 1:1 experiment/simulation studies are feasible. Previous tests of these relatively ideal single crystal samples demonstrated that silicon had higher than normal strength, produced yield discontinuities at relatively low loads and under larger displacements work-hardened substantially [22,23]. As they also exhibited reverse plasticity [23], it was concluded that dislocation plasticity was the main, if not the only source of plasticity. One can see that the Burgers vector sized steps in the load-displacement curve in the inset of Fig. 2(*b*) sum to the total residual displacement. Additionally, some small volume TEM *in situ* indentation experiments into wedge shaped silicon samples demonstrated dislocation loops and no phase transformation [25]. Finally, TEM images of a plastically deformed nanostructure of a similar size to the nanosphere provided evidence of dislocation activity but no phase transformation or amorphous phase [26].

The intriguing aspect of the multiple steps seen in the inset of Fig. 2(*b*) described here and elsewhere [22,23] is that these each are on the order of 0.2 nm, slightly less than a Burgers vector in silicon. We have taken this as evidence of dislocation emission. As stored elastic energy was released when each of the yield excursions in the Fig. 2 inset occurred, it was first necessary to speculate on where the energy was absorbed. An obvious choice was

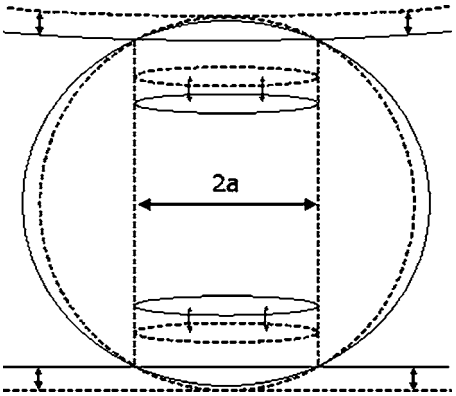


Fig. 3 Idealized compression of a nanosphere

plastic work associated with moving a dislocation. To solve this problem one needs to know where the dislocations nucleate, of what character they are and where they end up inside the nanosphere. For a single loop inside a sphere a solution by Willis [27] exists but it was desirable to have this for all of the steps shown. The situation was idealized to sequential prismatic loops of the same size as the contact radius, a. This is schematically shown in Fig. 3. Emitting the first dislocation at the edge of the contact requires a force needed to exceed an image force combined with a resisting Peierls barrier. When it exceeds this, the dislocation glides to a new equilibrium position. There are three caveats here regarding the validity of the image force as used. One is that the loop would have to be slightly outside the contact to experience a free surface. The second caveat is that a vacuum is used to represent the image force condition meaning that the oxide film is broken. As this might act as the nucleation site, this is also possible and even if not, the modulus of SiO_2 is substantially less than Si (70 GPa $<$ 160 GPa). The third caveat is that we consider prismatic loops while shear loops might be just as likely or more so. There is currently a work in progress to more clearly identify the dislocation character.

Given those caveats, a Hertizan stress distribution was utilized for a given spherical contact as provided by Johnson [28]. This was further coupled to a dislocation model for crack tip shielding of a plastically polarizable material by Zhou and Thomson [29]. By substituting the spherical contact stress field for the crack-tip stress field one finds a shear stress of

$$\sigma_{z_i}=\frac{p_o}{2}\left[\frac{1+\nu}{1-\frac{z_i}{a}\tan^{-1}\left(\frac{a}{z_i}\right)+\frac{1}{2}\left(1+\frac{z_i^2}{a^2}\right)}-\frac{a^2}{a^2+z_i^2}\right]-\frac{\mu b_i}{4\pi(1-\nu)z_i}+\sum_{\substack{j\\ j\neq i}}\frac{\mu b_j}{2\pi(1-\nu)(z_i-z_j)}. \tag{1}$$

Here, p_o is the maximum pressure, z_i is the distance below the contact, a is the contact radius, ν is the Poisson's ratio, μ is the shear modulus, and b_i and b_j are the Burgers vectors of the ith and jth dislocations in this case assumed to be the same. Additionally, at $\sigma_{zi}=\sigma_f$, this friction stress may stop further dislocation motion. Then, by increasing the driving force and exceeding σ_f, the dislocation may move freely to a deeper position below the surface where again $\sigma_{zi}=\sigma_f$ and arrest occurs. A simulation using Eq. (1) is described more fully in the discussion of Fig. 3 later. The maximum pressure for a spherical contact is $3P/2\pi a^2$, where P is the external load. The first term in Eq. (1) is the external driving force, the second term is the image force, and the third is the interaction force from any previously injected dislocations.

To check if Eq. (1) represents a reasonable description of the dislocations at equilibrium arrest positions we utilized experimen-

tal observations from the study of Rodriguez de La Fuente et al. [30]. Their experiments involved nucleation of dislocations by a tungsten scanning tunneling microscope (STM) tip driven into an Au (001) surface. Hillocks representing surface displacements associated with dislocations just beneath the surface were detected at 40.4, 58, and 82 nm from the center of the indentation. As these dislocations were traveling parallel to the free surface as pushed by the tungsten tip, two caveats or modifications of the use of Eq. (1) were required. First, we assume semicircular isostress contours which, while not the elastic stress distribution, has been shown to be fairly representative of the von Mises stress for shallow indentations into tungsten. For this continuum simulation, a three-dimensional, elastic-plastic finite difference formulation used the principle of virtual power [31]. The modification is that the image force for the driving of dislocations away from the tungsten tip involves a bimaterial interface. This can be seen by tracing back the dislocations to their origin and arriving at the tungsten tip. In terms of Dundurs parameters, α and β, the image force is given by

$$F_i = \frac{\mu_1(b_1^2 + b_2^2)(\alpha + \beta^2)}{4\pi(1-\nu)(1-\beta^2)x_i} \tag{1a}$$

where b_1^{Au} is $a_o/\sqrt{6}$ for the partial dislocation in Au and b_2^{W} is assumed to have the same form constituting the image dislocation. Also, μ_1 and ν_1 for Au are 30 GPa and 0.4 while μ_2 and ν_2 for W are 140 GPa and 0.28. The Dundurs parameters are

$$\alpha = \frac{\Gamma(1-\nu_1) - (1-\nu_2)}{\Gamma(1-\nu_1) + (1-\nu_2)}; \quad \beta = \frac{1}{2}\frac{\Gamma(1-2\nu_1) - (1-2\nu_2)}{\Gamma(1-\nu_1) + (1-\nu_2)} \tag{1b}$$

with $\Gamma = \mu_2/\mu_1 = 4.67$. Calculation of the friction stress at the three positions noted by the Hillocks in the STM micrograph of Fig. 1, Ref. 30, resulted in a friction stress of 90 ± 3 MPa. This required a use of $R_{\mathrm{tip}} = 16$ nm which is realistic since STM tips are often between 10–20 nm in radius. The satisfying aspect of this is that the friction stress is very close to the flow stress of 80 MPa previously used for indentation into Au (001) surfaces [32]. With this confirmation, we return to the silion experiments and apply this to the first four displacement excursions in the inset of Fig. 2(*b*). Using Eq. (1), the modeling is accomplished with $\mu = 66$ GPa, $\nu = 0.218$, and $b = 0.236$ nm for this 39 nm diameter silicon sphere. The resulting stress on the assumed prismatic loop at distances z below the surface are shown in Fig. 4 for the load associated with each excursion. For the first one at $P = 0.92$ μN, the dislocation sees an initial negative stress as the loop starts to form and gradually goes positive as it proceeds further beneath the surface. When the positive stress becomes sufficiently large to exceed the friction stress, here assumed to be a Peierls stress of 4 GPa, the dislocation can freely glide to a distance about 8 nm below the surface where it arrests also seen in Fig. 4. A friction stress of 4 GPa, about one-third of the 12 GPa hardness for silicon, is due to its high Peierls barrier of 2.4 eV. For the second dislocation emitted at the bottom, we allow the positions of both the top and bottom to readjust until the forces are zero consistent with a Peierls stress of 4 GPa. A similar iterative procedure was used for the third and fourth dislocations alternatively allowed to nucleate from the top and bottom sequentially. The final positions of all four dislocations are indicated in Fig. 4 with z representing the distance from the top of the sphere for dislocations 1 and 3 and from the bottom of the sphere for dislocations 2 and 4. While the stress would continue to fall at the far right below the dotted line we terminate the curves there to show the dislocation arrest positions. The real work associated with moving these dislocations is the force on the dislocation per unit length, τb, times the loop length, $2\pi a$, times the distance, z. For the first dislocation the position z is 1.02 nm. This is the distance the dislocation moves before it can glide freely to its equilibrium position where the stress equals 4 GPa. The plastic work then for this excursion is

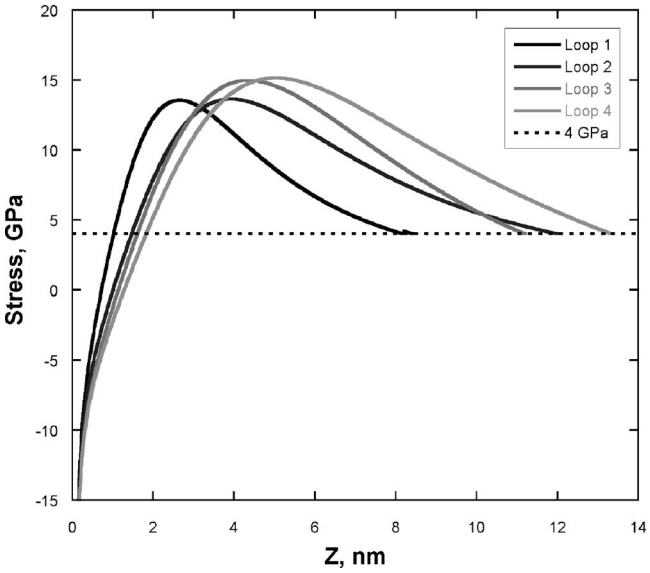


Fig. 4 Stress-depth curves for the four dislocation loops formed when compressing a 38.6 nm silicon nanosphere. The initial work is governed by the crossing point with a Peierls stress of 4 GPa which also defines the arrest position.

$$\delta W_{\mathrm{p}} = \tau b \cdot 2\pi a \cdot z. \tag{2}$$

Energy can also be absorbed by new surface creation. Here we take the contact area before and after the excursion, giving

$$\delta S = \gamma_s A|_{\mathrm{after}} - \gamma_s A|_{\mathrm{before}} \tag{3}$$

and use $\gamma_s = 1.56$ J/m^2 for Si. Using Eq. (3) requires an area of $2\pi a^2$ since there is a contact both at the top and bottom of the sphere. In addition to this energy absorbed, some elastic energy remains within the created loop, given as

$$\delta U_\ell = \pi a\left[1 + \frac{1}{1-\nu}\right]\left[\frac{\mu b^2}{4\pi}\ln\left(\frac{ae^2}{b}\right)\right] \tag{4}$$

where the first term in brackets is for a dislocation of half screw and half edge character, and the second term is the energy per unit length for a screw dislocation which includes a continuum description of the core energy contribution. The reason we pick a mixed dislocation representation is that it is probably more likely that shear loops are created as opposed to prismatic loops. In either case for the final summation of energy this makes a relatively small difference since the stored elastic energy is less than 20% of the total and pure screw versus half edge and half screw only varies by 14% making a maximum difference in the total energy of less than 3%. This energy that is absorbed or remaining is balanced by the external work such that

$$P\mathrm{d}\delta_{\mathrm{exc}} = \delta W_p + \delta S + \delta U_\ell \tag{5}$$

where the external work is $P\mathrm{d}\delta_{\mathrm{exc}}$. As shown in Table 1 these three terms represent $93 \pm 20\%$ of the work in the yield excursion.

To examine whether the critical distances, z, below the indenter from Eq. (1) and used in Eq. (2) were realistic, we additionally checked an atomistic simulation by van Vliet et al. [20] conducted using a 13 nm radius indenter penetrating a defect-free Al (001) crystal. The load-displacement curve under predominantly displacement control exhibited the load drops shown in Fig. 5(*a*). From reviewing their on-line simulations (Ref. 40 in [20]), it appeared that all load drops except the second one were associated with prismatic dislocation loop nucleation. With six dislocations nucleated, the discretized dislocation approach represented by Eq. (1) was simulated using $\mu = 26.3$ GPa, $\nu = 0.33$, $b = 0.2865$ nm and a friction stress of 0.35 GPa. After the first dislocation exceeded the friction stress at $z = 1.62$ nm, it would glide free and arrest at

Table 1 External work and energy balance in Si experiments [22] and Al simulations of Van Vliet et al. [20]. The first two columns are from Refs. 22 and 20 with the remainder being calculations from the present model based on Eqs. (1)–(10), (16), and (17). The units are multiplied by 10^{16} for both J and m^2.

	P μN	$d\delta_{exc}$ nm	W_{ext} J	δA m^2	δW_s J	z^* nm	δW_p J	δU_ℓ J	W_{int}
Si exp.	0.92	0.155	1.43	0.187	0.291	1.02	1.05	0.27	1.61
Si exp.	2.2	0.19	4.1	0.228	0.356	1.49	2.33	0.46	3.15
Si exp.	3.2	0.16	5.12	0.183	0.285	1.64	3.3	0.53	4.11
Si exp.	3.4	0.12	5.28	0.142	0.221	1.88	4.53	0.62	5.37
	δ, nm	dP, nN							
Al sim.	0.703	45	0.316	0.012	0.0133	1.62	0.309	0.2	0.522
Al sim.	0.984	62	0.61	0.0209	0.023	2.19	0.419	0.24	0.682
Al sim.	1.23	74	0.909	0.0308	0.0338	2.58	0.562	0.27	0.866
Al sim.	1.53	64	0.979	0.0125	0.0137	2.86	0.748	0.31	1.07
Al sim.	1.73	92	1.59	0.0282	0.0311	3.73	0.874	0.33	1.24
Al sim.	1.92	75	1.44	0.0157	0.0173	4.85	1.14	0.35	1.51

12.4 nm, the highest curve in Fig. 5(*b*). The second dislocation could then be emitted until at 2.19 nm below the surface it could freely glide to 10.2 nm. In doing this the stress on the first dislocation rose to 0.49 GPa and then decayed back to 0.35 GPa at a new position of 15 nm. We know this is not completely accurate since the computer simulation cell is only 20 nm deep and we have assumed no image effect at the back free surface. When the original stress exceeds the friction stress at values of $z=z^*$ ranging from 1.6 to 4.8 nm, 6 loops are sequentially emitted under increasing load. These result in the load drops indicated in Fig. 5(*a*), the loads and $z=z^*$ values being indicated in Table 1. As summarized in Table 1, the proposed model determination of the dislocation work, surface work and stored elastic energy in the dislocation agrees well with the external work done except for the first dislocation, i.e., W_{ext} compares well to W_{int} in Eq. (5). For all of these cases, it is clear that this model appears to account for the external work done in the early stages of dislocation plasticity in small volume contacts. This also gives a reasonable basis for using an energy balance criterion when large numbers of dislocations are involved in an avalanche during a yield excursion.

Multiple Dislocation Avalanche Events

Multiple dislocations released in displacement excursions like those in Fig. 1(*a*) might contain 20 or more dislocations. Here, the assessment is slightly different than that above for three reasons. First, it is too difficult to do the bookkeeping with the type of simulation represented by Eq. (1). Second, many of the dislocations may have escaped. The escape of dislocations which we know happens at indents with pileup implies either slip steps or oxide film cracking with new surfaces created. Recently, we have treated this with an energy balance criterion for somewhat larger plasticity in small volumes of various shapes such as shallow contacts of single crystals, nanospheres, and nanoboxes [26,33].

Constant Load Excursion, $dP=0$. For this larger scale plasticity we consider the energy absorbed by plastic work, W_p, and surface energy, U_s, to be balanced by the driving force, the external work, W_{ext} minus the elastic energy change, U_E. At constant load this is

$$W_p + U_s - (W_{ext} - U_E) = 0. \tag{6}$$

Consider how this might apply to the relatively larger yield excursions shown in Fig. 1(*a*). These excursions in the load-displacement curves occur rapidly as dislocations are created and move faster than the indenter tip. We can use such a dislocation avalanche to evaluate the plastic energy absorbed. For a group of dislocations the energy absorbed plastically is the force per unit length of each prismatic loop multiplied by their length, their average distance moved and the number moved, giving

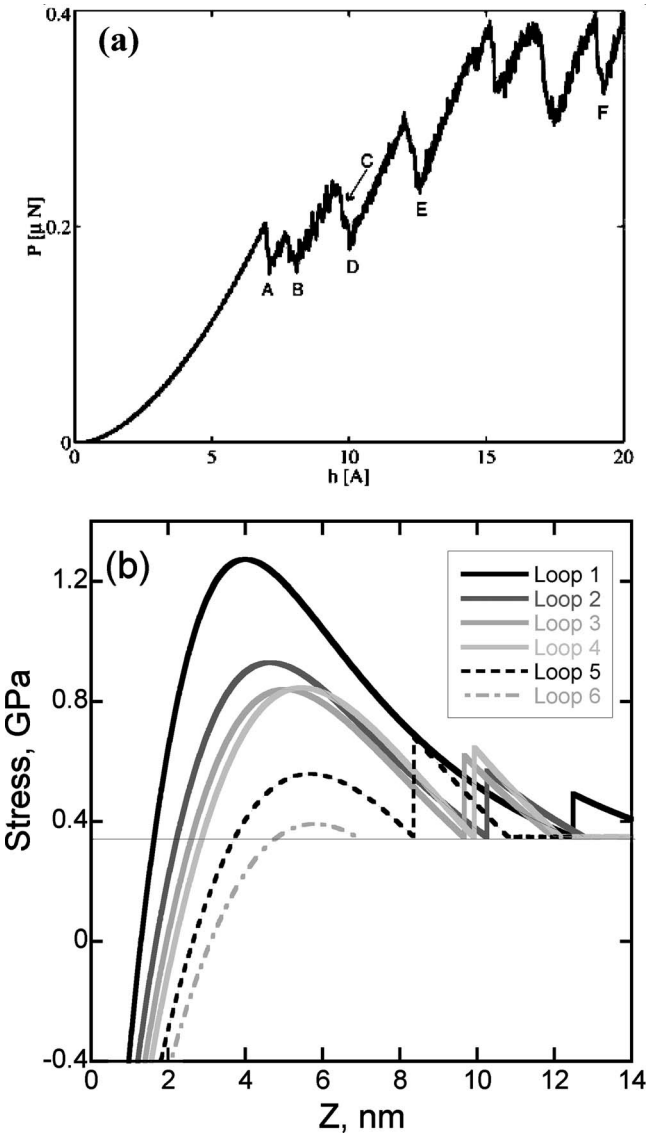


Fig. 5 (*a*) From Van Vliet et al. [20], the simulated force-displacement curve for an indentation into aluminum crystal and (*b*) the corresponding stress-depth curves using the present model in conjunction with their data for six loops. The second load-drop at B was not considered to correspond to a dislocation loop formation.

Table 2 Displacement excursion data under load control for various oxide thicknesses on Fe-3% Si (001) crystals (indenter tip was 80 nm radius)

Film, nm	δ_0, nm	δ_f, nm	δ_{exc}, nm	P, μN	Film, nm	δ_0, nm	δ_f, nm	δ_{exc}, nm	P, μN
3.6	7.1	17.7	10.6	59.6	9	12.3	28	15.7	110
	6.2	18.5	12.3	72.8		12.3	39.4	27.1	114
	5.7	10.4	4.7	46.6		12	39.8	27.8	115
	6.6	18	11.4	62.3		12.4	39.4	27	117
	6.3	13.3	7	54.8		12.3	39.1	26.8	118
	6.5	12.1	5.6	48.9		13	38.3	25.3	111
	5.4	16.2	10.8	60.6		14.2	40.4	26.2	114
	5.5	9.5	4	37.2		13.5	38	24.5	112
5.6	14.4	40.2	25.8	117	17.4	12.4	60.2	47.8	196
	9.2	25.2	16	83.1		11	47.1	36.1	160
	7.9	30.3	22.4	99.4		12.4	59.5	47.1	191
	9.9	30.6	20.7	98.6		11.3	53.5	42.2	171
	10.6	30.7	20.1	91.7		8	45.5	37.5	151
	11.4	38	26.6	117		11	58.4	47.4	197
	8.9	28.9	20	98.4		11.6	58.4	46.8	191
	10.6	38	27.4	123		14.3	55.9	41.6	176

$$W_p \cong \tau_{\text{flow}}^{\text{local}} b \cdot \eta\pi a \cdot a \cdot \frac{\delta}{b} \approx \eta\pi a^2 \tau_{ys}\delta; \quad \eta > 1$$

$$\frac{\text{force}}{\text{unit length}} \cdot \text{ave. length} \cdot \text{dist.} \cdot \text{number} \tag{7}$$

where η is a constant and δ is the depth of the indenter penetration. This gives the same result used previously [34] if c, the plastic zone radius, is equal to $2a$, twice the contact radius. With plasticity the contact radius is best characterized by a geometric contact as given by $a^2=2\delta R-\delta^2$. For $\delta \ll R$, this is given by $\delta = a^2/2R$. Taking the shear flow stress to be two-thirds the compressive flow stress with the latter being one-third the hardness, one finds [26]

$$W_p \cong \frac{\eta P a^2}{9R} = \frac{2\eta P\delta}{9}. \tag{8}$$

Besides plastic energy absorption, there is a surface energy absorption for indentation, again using $\delta=a^2/2R$ given by,

$$U_s = \alpha\gamma_s \pi a^2 = 2\pi\alpha\gamma_s R\delta; \quad \alpha > 1. \tag{9}$$

The value of $\alpha>1$ used here and elsewhere [26] is a recognition that a greater proportion of slip step emergence and/or oxide film fracture around the contact area represents new area greater than just the contact.

For the stored elastic energy, U_E, there are the number of dislocations in the avalanche, δ/b with an average length $\eta\pi a$. Parallel to Eq. (4), this term becomes

$$U_E = \frac{\delta}{b} \cdot \eta R \cdot \frac{\mu b^2}{8\pi}\left[1+\frac{1}{1-\nu}\right]\left[\ln\left(\frac{ae^2}{b}\right)\right] \simeq \eta\mu b\delta R. \tag{10}$$

The latter approximation comes from the last two terms in brackets and the average length scaling with the tip radius rather than the contact radius as used in Eq. (7). This difference is in recognition that the final configurations of the dislocation would be larger than the initial if these are shear loops. Using Eqs. (7), (9), and (10) with Eq. (6) allows an energy balance similar to that for Eq. (5). From Eq. (7), a^2 is used as $2\delta R$, and Eqs. (9) and (10) remain the same. Writing Eq. (6) in terms of δ gives

$$\eta\pi 2R\delta^2\tau_{\text{flow}} + 2\pi\alpha\gamma_s R\delta - (P\delta - \eta\mu b R\delta) = 0. \tag{11a}$$

One can see this reduces to

$$\delta = \frac{P-\eta_o}{2\eta\pi R\tau_{\text{flow}}^{\text{local}}}; \quad \eta_o = 2\pi\alpha\gamma_s R + 4\eta\mu b R \tag{11b}$$

where η_o is constant for a given material being indented with a tip of radius R, and α, η are greater than one. As before, we take the local shear flow stress to be two-thirds the compressive flow stress which in turn is one third the hardness or mean pressure. With $a^2=2\delta R$, this gives

$$\tau_{\text{flow}}^{\text{local}} = \frac{2}{9}\cdot\frac{P}{\pi a^2} = \frac{P}{9\pi\delta R}. \tag{12}$$

For a yield excursion as in Fig. 1(*a*), $dP=0$. With Eq. (12), this instability is represented by

$$-d\tau_{\text{flow}}^{\text{local}} \approx \frac{\tau_{\text{flow}}^{\text{local}}}{\delta} d\delta. \tag{13}$$

This incremental decrease in flow stress is balanced by an incremental increase in displacement which increases contact area. One can show for the experiments to be described later that $P \gg \eta_o$ such that Eqs. (11) and (13) can describe an excursion from δ_o to δ_1 by

$$\int_{\delta_o}^{\delta_1} \frac{d\delta}{\delta^2} = -\frac{2\eta\pi R}{P}\int_{\tau_o}^{\tau_1} d\tau_{\text{flow}}. \tag{14}$$

For a displacement excursion, δ_{exc}, represented by $\delta_1-\delta_o$, this gives

$$\delta_{exc} = \frac{2\eta\delta_f}{9\tau_{\text{flow}}^{\text{local}}}(\tau_o-\tau_1)^{\text{local}}. \tag{15}$$

It is seen directly that Eq. (15) is the incremental form of Eq. (13).

Fixed Displacement Excursion, $d\delta=0$. Here we use the equivalent energy changes for small excursions,

$$P d\delta = \delta dP. \tag{16}$$

Combining Eqs. (15) and (16) we find

$$P_{exc} = 2\pi\eta R\delta_f d\tau_{\text{flow}}^{\text{local}}. \tag{17}$$

Noting that $d\tau_{\text{flow}}^{\text{local}}$ is a decrease in flow stress, it is seen that P_{exc} is a load drop. This represents the incremental form of Eq. (11*b*) for displacement control and η_o a constant. In the next section we compare these displacement and load excursion estimates to indentation observations like those in Fig. 1 but specific to instabilities where oxide film effects were known to be important.

Table 3 Load excursion data under displacement control for an oxide film of approximately 17.4 nm (indenter tip was 66 nm radius)

P_0, μN	P_f, μN	P_{exc}, μN	δ, nm
70	35	35	8
112	55	57	10
100	53	47	16
118	70	48	21
148	48	100	19.5
190	120	70	22.5
225	150	75	27
230	90	140	26
220	115	105	30
300	150	150	34

Discussion

Fortunately, a large number of yield instabilities had previously been observed under both displacement control producing load drops [34], and load control producing yield excursions [6] but not all of the details had been presented. Details are presented in Tables 2 and 3 for load and displacement control, respectively. A micromechanical test apparatus with a diamond indenter tip radius of 66 nm was used for the displacement control experiments while a Hysitron Triboscope with a Digital Instruments AFM with an indenter tip radius of 80 nm was used for the load control experiments. Two points are clear here. First, the severity of the displacement or load excursion increases as the indenter displacement increases. Second, it is seen for the load control experiments that there tends to be a clustering of data points with increasing displacement for the instabilities corresponding to an increase in oxide film thickness. As measured by ellipsometry, oxide thicknesses increased by thermal oxidation at different temperatures from 3.6 to 17.4 nm [6,13]. The 40 instability points from Tables 2 and 3 are shown in Fig. 6 as a function of the final indenter displacement, δ_f, after the excursion or load drop arrested. In both cases it is seen that there is a nearly linear relationship between the excursions and the final displacement at the end of an excursion, δ_f.

For comparison to the model, one can see from Eqs. (15) and (17) for δ_{exc} and P_{exc} that by selecting $d\tau/\tau$ equal to 0.75 and $\eta = 4$, the two instabilities are given by

$$\delta_{exc} = \frac{2}{3}\delta_f; \quad P_{exc} = 6\pi R\delta_f\tau_{flow} = 6000\ \text{N/m} \cdot \delta_f. \qquad (18)$$

For the load drops, τ_{flow} was used as 4.0 GPa. The models of Eq. (18) are shown as the solid and dashed curves in Fig. 6. While the fit is good for the load control case, the displacement control case appears to overpredict the magnitude of the load drops. This is due to the micromechanical test facility used in those early experiments not being fully fixed displacement during the excursion. For example, the IBM designed nanoindentation system used in those early experiments [34] exhibited both load drops and displacement excursions in the same event. A typical case where the load drop was 115 μN at a displacement of 30 nm gave a displacement excursion of 13 nm at the same time. If such an excursion had not occurred, the load drop would have been greater. The only other slight deviation are the greater than predicted displacement excursions associated with the crystal covered with the thickest oxide film. For the 17.4 nm oxide film in Fig. 6, the excursions were underpredicted by about 20%. We strongly suggest there is an oxide film thickness effect not fully described by the present model. Finally, we checked to determine how appropriate our fitting parameters were for obtaining Eq. (18). In Table 4, calculations of the incremental flow stress changes for the four different oxide film thicknesses are shown. These are based on the load control tests where the constant load was divided by the contact area at the start and end of the displacement excursion. The average value of $-d\tau/\tau$ is 0.7 which is reasonably consistent with the value of 0.75 used for the models in Fig. 6. Additionally, the value of 4.0 GPa for τ_{flow} used is realistic for the three smallest oxide film thicknesses but not the largest. This again suggests a shortcoming of the present model regarding the exact role of oxide film thickness in these instabilities. Such approaches are under current investigation, one aspect being discussed elsewhere in terms of a surface energy argument [26]. Finally, while the modeling has considered prismatic loops, we have mentioned the possibility of shear loops as well. Identifying which type is responsible is beyond the scope of the present investigation. For the silicon of Figs. 2–4, the nanosphere orientation is unknown. Even if known, it is not clear at these very high pressures that undissociated partial dislocations would be involved [35]. For the Fe-3% Si, this body-centered cubic material could have slip on the {111} planes but the directions could be any of the $\langle 1\bar{1}1\rangle$, $\langle 1\bar{1}2\rangle$ or $\langle 12\bar{3}\rangle$ vectors. Not until improved in situ nanoindentation devices incorporated into transmission electron microscopes are producing data

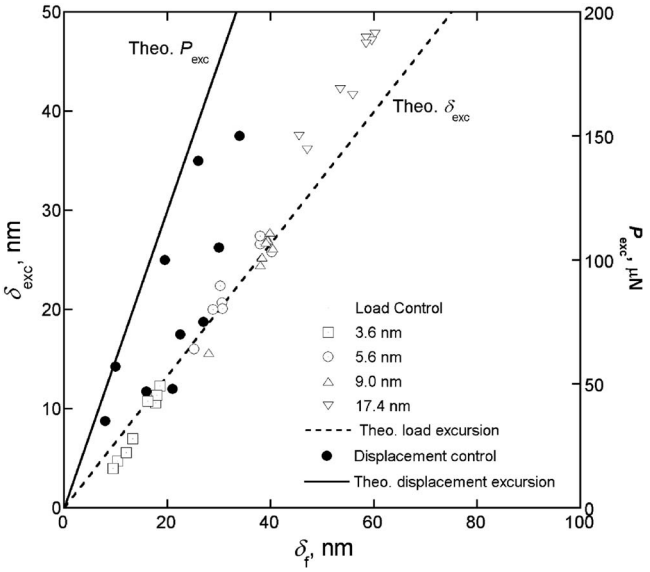


Fig. 6 Comparison of instability magnitudes from Eq. (18) for displacement control tests with one film thickness and for load control tests with four film thicknesses

Table 4 Determination of $d\tau_{flow}$ from displacement increases at constant load

Oxide film (averages)	δ_0, nm	δ_f, nm	a_0^2, nm^2	a_1^2, nm^2	P, μN	$d\tau_{flow}$, GPa	Initially observed τ_{flow}, GPa
3.6 nm	6.2	14.5	992	2320	55.4	−2.7	4.0
5.6 nm	10.3	32.7	1648	5232	102.5	−3.0	4.4
9.0 nm	12.8	37.8	2048	6048	113.5	−2.6	3.9
17.4 nm	11.5	54.8	1840	8768	179	−5.4	6.9

[25] will such definition be possible for small penetration depths or small spheres.

Summary

Two plastic instability events that can occur during nanoindentation have been addressed. These involve single loop nucleation in the absence of oxide film fracture and multiple dislocation avalanches in the presence of oxide film fracture. The external work done for a single dislocation event is proposed to be mostly accounted for by dislocation work, surface work, and residual stored elastic energy. For multiple dislocations released during a film fracture event, it is shown that the instability condition can be highly sensitive to the oxide film thickness. These events are modeled by an energy balance criterion involving plastic energy absorbed and surface energy.

Acknowledgment

The authors would like to acknowledge discussions with Stan Veprek, Eric Stach, Alan Needleman, and Sid Yip. Support from a NSF-IGERT through Grant No. DGE-0114372, a NIRT through Grant No. DMI0103619 and a current NSF Grant No. CMS-0322436 are gratefully acknowledged.

References

[1] Gane, N., and Bowden, F. P., 1968, "Microdeformation of Solids," J. Appl. Phys., **39**, pp. 1432–1435.
[2] Gane, N., 1970, "The Direct Measurement of the Strength of Metals on a Sub-Micrometer Scale," Proc. R. Soc. London, Ser. A, **317**, pp. 367–391.
[3] Gane, N., and Cox, J. M., 1970, "The Micro-Hardness of Metals at Very Low Loads," Philos. Mag., **22**, pp. 881–891.
[4] Corcoran, S. G., Colton, R. J., Lilleoddon, E. T., and Gerberich, W. W., 1997, "Anomalous Plastic Deformation at Surfaces: Nanoindentation of Gold Single Crystals," Phys. Rev. B, **55**, pp. R16057–R16060.
[5] Miller, R. E., Shilkrot, L. E., and Curtin, W. A., 2004, "A Coupled Atomistics and Discrete Dislocation Plasticity Simulation of Nanoindentation into Single Crystal Thin Films," Acta Mater., **52**, pp. 271–284.
[6] Bahr, D. F., Kramer, D. E., and Gerberich, W. W., 1998, "Nonlinear Deformation Mechanisms During Nanoindentation," Acta Mater., **46**, pp. 3605–3617.
[7] Pethica, J. B., and Tabor, D., 1979, "Contact of Characterized Metal Surfaces at Very Low Loads: Deformation and Adhesion," Surf. Sci., **89**, pp. 182–190.
[8] Pethica, J. B., 1981, "The Effect of Oxygen on the Adhesion of Clean Metal Surfaces," Philos. Mag. A, **43**, pp. 731–738.
[9] Syed Asif, S. A., and Pethica, J. B., 1997, "Nanoindentation Creep of Single-Crystal Tungsten and Gallium Arsenide," Philos. Mag. A, **76**, pp. 1105–1118.
[10] Mann, A. B., and Pethica, J. B., 1996, "Nanoindentation Studies in a Liquid Environment," Langmuir, **12**, pp. 4583–4586.
[11] Mann, A. B., and Pethica, J. B., 1999, "The Effect of Tip Momentum on the Contact Stiffness and Yielding During Nanoindentation Testing," Philos. Mag. A, **79**, pp. 577–592.
[12] Gerberich, W. W., Venkataraman, S. K., Huang, H., Harvey, S. E., and Kohlstedt, D. L., 1995, "The Injection of Plasticity by Millinewton Contacts," Acta Metall. Mater., **43**, pp. 1569–1576.
[13] Kramer, D. E., Yoder, K. B., and Gerberich, W. W., 2001, "Surface Constrained Plasticity: Oxide Rupture and the Yield Point Process," Philos. Mag. A, **81**, pp. 2033–2058.
[14] Page, T. F., Oliver, W. C., and Hargue, C. J., 1992, "The Deformation Behavior of Ceramic Crystals Subjected to Very Low Load (Nano)indentations," J. Mater. Res., **7**, pp. 450–473.
[15] Lilleodden, E. T., Bonin, W., Nelson, J., Wyrobek, J. T., and Gerberich, W. W., 1995, "*In situ* imaging of μN load indents into GaAs," J. Mater. Res., **10**, pp. 2162–2165.
[16] Kiely, J. B., and Houston, J. E., 1998, "Nanomechanical Properties of Au (111), (001), and (110) Surfaces," Phys. Rev. B, **57**, pp. 12588–12594.
[17] Michalske, T. A., and Houston, J. E., 1998, "Dislocation Nucleation at Nano-Scale Mechanical Contacts," Acta Mater., **46**, pp. 391–396.
[18] Kelchner, C. L., Plimpton, S. J., and Hamilton, J. C., 1998, "Dislocation Nucleation and Defect Structure During Surface Indentation," Phys. Rev. B, **58**, pp. 11085–11088.
[19] Christopher, D., Smith, R., and Richter, A., 2001, "Atomistic Modeling of Nanoindentation in Iron and Silver," Nanotechnology, **12**, pp. 372–383.
[20] Van Vliet, K. J., Li, J., Zhu, T., Yip, S., and Suresh, S., 2003, "Quantifying the Early Stages of Plasticity Through Nanoscale Experiments and Simulations," Phys. Rev. B, **67**, p. 104105.
[21] Zhu, T., Li, J., Van Vliet, K. J., Ogata, S., Yip, S., and Suresh, S., 2004, "Predictive Modeling of Nanoindentation-Induced Homogeneous Dislocation Nucleation in Copper," J. Mech. Phys. Solids, **52**, pp. 691–724.
[22] Gerberich, W. W., Mook, W. M., Perrey, C. R., Carter, C. B., Baskes, M. I., Mukherjee, R., Gidwani, A., Herberlein, J., McMurry, P. H., and Girshick, S. L., 2003, "Superhard silicon nanospheres," J. Mech. Phys. Solids, **51**, pp. 979–992.
[23] Gerberich, W. W., Mook, W. M., Cordill, M. J., Carter, C. B., Perrey, C. R., Heberlein, J., and Girshick, S. L., 2005, "Reverse Plasticity in Single Crystal Nanospheres," International Journal of Plasticity, **21**, pp. 2341–2405.
[24] Perrey, C. R., 2004, Ph.D. thesis, "Characterizing Nanoparticles," University of Minnesota.
[25] Minor, A., Lilleodden, E. T., Jin, M., Stach, E. A., Chirzan, D. C., and Morris Jr., S. W., 2005, "Room Temperature Dislocation Plasticity in Silicon," Philos. Mag., **85**, pp. 323–330.
[26] Gerberich, W. W., Cordill, M. J., Mook, W. M., Moody, N. R., Perrey, C. R., Carter, C. B., Mukherjee, R., and Girshick, S. L., 2005, "A Boundary Constraint Energy Balance Criterion for Small Volume Deformation," Acta Mater., **53**, pp. 2215–2229.
[27] Willis, J. R., Bullough, R., and Stoneham, A. M., 1983, "The Effect of Dislocation Loops on the Lattice Parameter Determined by X-Ray Diffraction," Philos. Mag. A, **48**, pp. 95–107.
[28] Johnson, K. L., 1985, *Contact Mechanics*, Cambridge University Press, Cambridge, MA.
[29] Zhou, S. J., and Thomson, R., 1991, "Shielding of Cracks in a Plastically Polarizable Material," J. Mater. Res., **6**, pp. 1763–1772.
[30] Rodriguez de la Fuente, O., Zimmerman, J. A., Gonzales, M. A., de la Figuera, J., Hamilton, J. C., Pai, W. W., and Rojo, J. M., 2002, "Dislocation Emission Around Nanoindentations on a (001) fcc Metal Surface Studied by Scanning Tunneling Microscopy and Atomistic Simulations," Phys. Rev. Lett., **88**, pp. 036101/1–4.
[31] Cheng, L. S., 1996, Ph.D. thesis, "Numerical Modeling of Indentation and Scratch Problems," University of Minnesota.
[32] Gerberich, W. W., Tymiak, N. I., Grunlan, J. C., Horstemeyer, M. F., and Baskes, M. I., 2002, "Interpretations of Indentation Size Effects," J. Appl. Mech., **69**, pp. 433–442.
[33] Mook, W. M., Jungk, J. M., Cordill, M. J., Moody, N. R., Sun, Y., Xia, Y., and Gerberich, W. W., 2004, "Geometry and Surface State Effects on the Mechanical Response of Au Nanostructures," Z. Metallkd., **95**, pp. 416–424.
[34] Gerberich, W. W., Nelson, J. C., Lilleodden, E. T., Anderson, P., and Wyrobek, J. T., 1996, "Indentation Induced Dislocation Nucleation: The Initial Yield Point," Acta Mater., **44**, pp. 3585–3598.
[35] Rabier, J., Cordier, P., Demenet, J. L., and Garem, H., 2001, "Plastic Deformation of Si at Low Temperature Under High Confining Pressure," Mater. Sci. Eng., A, **309–310**, pp. 74–77.

Journal of Applied Mechanics

Technical Briefs

Examination of the Stress Distribution in the Tip of a Pencil

E. Pogozelski
e-mail: pogo@geneseo.edu

D. Cole

M. Wesley

State University of New York College at Geneseo,
One College Center, Geneseo, NY 14454

The stresses within the tip of a pencil are examined theoretically, numerically, and experimentally to determine the position and orientation of the fracture surface. The von Mises stress is used to evaluate the impact of the normal and shear stresses due to compression, bending, torsion, and shear. The worst-case stress is shown to occur along the top edge of the inclined pencil point, where the normal stress is compressive. The resulting crack propagates diagonally downwards and towards the tip from this initial position, and is frequently observed to contain a cusp. [DOI: 10.1115/1.2062847]

Introduction

In 1978 Cronquist [1] analyzed normal stresses in pencil points to determine why they usually fracture at the same relative distance from the tip. Cronquist focused on pure bending, ignoring shear and axial forces. The geometry of the pencil and a complete set of the external forces acting on it are shown in Fig. 1, neglecting the weight. His analysis of the bending stress concluded that the onset of fracture occurs along the lower edge of the pencil cone, beginning at an axial distance $x^*=x/\ell=1.5$ from the point of the cone. He also observed that the fracture was largely planar, inclined up and away from the point. In 1983 Cowin [2] expanded Cronquist's analysis by including the axial components of the normal force and of the frictional force $\mu_1 N$. Both authors neglected the impact of shear stress, limiting their analyses to failure due to normal stress.

Analysis

As seen in Fig. 1, the tip of the sharpened cone is truncated by an amount ℓ. The external forces can be resolved into the following internal components of forces and moments [3]

$$\text{Axial force: } F=-N(\sin\theta-\mu_1\cos\theta) \tag{1}$$

$$\text{Shear force: } V_y=-N(\cos\theta+\mu_1\sin\theta) \tag{2}$$

$$\text{Shear force: } V_z=-N\mu_2 \tag{3}$$

$$\text{Torque: } T=\mu_2 N\ell\tan\alpha \tag{4}$$

$$\text{Bending moment: } M_y(x)=\mu_2 N(x-\ell) \tag{5}$$

$$\text{Bending moment: } M_z(x)=N(x-\ell)(\mu_1\sin\theta+\cos\theta)-N(\ell\tan\alpha)\times(\sin\theta-\mu_1\cos\theta) \tag{6}$$

Equations (1)–(6) neglect the impact of elastic deformation of the pencil geometry. The coefficients μ_1 and μ_2 are proportionality constants similar to coefficients of kinetic friction. For this study, $\mu_1=\mu_2$. However, in practice, either constant may lie between zero and the actual coefficient of kinetic friction, depending on the pencil orientation and the symbol being written.

Computation of Stress Fields. The stress components σ_{yy}, σ_{zz}, and τ_{yz} are uniformly zero; the components σ_{xx}, τ_{xy}, and τ_{xz} are non-zero. The normal stress σ_{xx} is a straightforward combination of uniform compression and bending

$$\sigma_{xx}(x,y,z)=\frac{F}{A(x)}+\frac{M_y(x)z}{I(x)}-\frac{M_z(x)y}{I(x)} \tag{7}$$

Here, $A(x)$ is the cross-sectional area of the pencil point, and $I(x)$ is the second moment of the cross-sectional area. T, V_y, and V_z all contribute to both the y and z components of shear in the xy plane. The shear due to T is circumferential. The shear flow due to V_y is primarily vertical, but as shown by Gere and Timoshenko [4], there must also be a horizontal component. V_z is similar to V_y. The net resulting shear stresses are

$$\tau_{xy}=-\frac{Tz}{J(x)}+\frac{V_y(R(x)^2-y^2)-V_z zy}{3I(x)} \tag{8}$$

$$\tau_{xz}=+\frac{Ty}{J(x)}+\frac{V_z(R(x)^2-z^2)-V_y zy}{3I(x)} \tag{9}$$

Here, $J(x)$ is the polar moment of the cross-sectional area. From these stresses, the resulting principle stresses are

$$\sigma_1=\frac{\sigma_{xx}}{2}+\frac{1}{2}(\sigma_{xx}^2+4(\tau_{xy}^2+\tau_{xz}^2))^{1/2} \tag{10a}$$

$$\sigma_2=\frac{\sigma_{xx}}{2}-\frac{1}{2}(\sigma_{xx}^2+4(\tau_{xy}^2+\tau_{xz}^2))^{1/2} \tag{10b}$$

$$\sigma_3=0 \tag{10c}$$

The von Mises stress is computed from the principle stresses, and simplifies to

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received July 20, 2004; final manuscript received May 24, 2005. Review conducted by R. M. McMeeking.

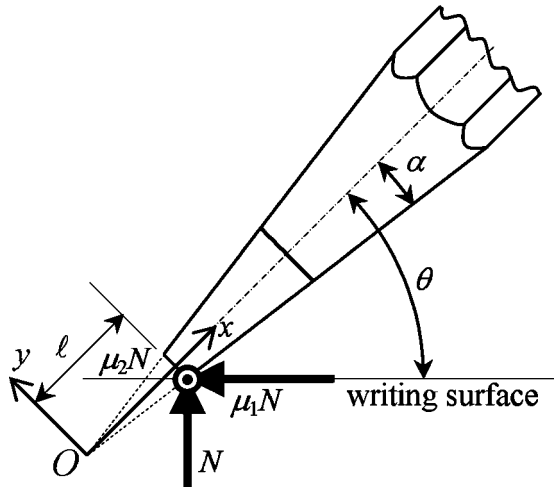


Fig. 1 Forces acting on the pencil tip. Note that the frictional force $\mu_2 N$ points out of the page.

$$\sigma_V = (\sigma_{xx}^2 + 3\tau_{xy}^2 + 3\tau_{xz}^2)^{1/2} \tag{11}$$

Failure is assumed to begin at the material point where the von Mises stress is maximum. Figure 2(*a*) shows the distribution of the von Mises stress in an axially oriented vertical plane in the tip.

Examination of Fig. 2(*a*) reveals three local stress maxima. The largest and most obvious maximum occurs in a small region near the tip at the point of application of the normal force N. The "failure" resulting from this high stress corresponds to the process of writing. The remaining two maxima occur on the outer surface of the pencil cone near the upper and lower edges. The lower maxima corresponds to a point of high tension in the tip, and the upper to a point of high compression. For typical values of α and θ, the magnitude of stress at the upper maxima is about 30% higher than that of the lower maxima due to the contribution of the compressive force F.

To verify the model above, the stresses were also computed using a commercial finite-element stress analysis package. Figure 2(*b*) shows the stress distribution for purposes of comparison with Fig. 2(*a*), and shows excellent agreement.

Fracture Initiation Position and Direction. Cronquist's primary conclusion was that fracture begins at $x^* = 1.5$ for the case without friction or axial forces (whether failure began on either the upper or lower edge). Both Cronquist and Cowin assumed that since the pencil lead material (a composite of clay and powdered carbon) is brittle, it is likely to fail at a point of tension rather than compression. This led them to conclude that crack propagation began at the lower maxima, propagating up and away from the tip. However, it is unlikely that a crack would propagate away from the pencil tip, because this would maximize the fracture area.

Even if the failure is assumed to be brittle, plots of the magnitude of the maximum principle stress show essentially the same distribution as the von Mises stress. Although pencil leads are brittle, digital photographs of used pencils show evidence of some plasticity. Also, brittle composite materials frequently have similar strengths in tension and compression.

Furthermore, the positions of the experimental fractures presented in the following much more strongly agree with the hypothesis that they initiate on the upper surface, where the stress is significantly larger. Figure 3 shows the computed positions of the upper stress maxima with the experimental positions of the fracture surface at its intersection with the upper surface of the cone. The upper line indicates the case of no friction, and when $\theta = 0°$, it also corresponds to the case of no axial force (yielding Cronquist's $x^* = 1.5$). As friction increases, the breaking position moves slightly towards the pencil point. Pencils are typically used at an angle $\theta = 56$ deg ± 7 deg, for which the breaking point is near $x^* = 1.7$. The experimental data fit well with the predicted positions; similar plots using the lower stress maxima show no such agreement.

Our experiments reveal that the fracture surface is not usually planar. Cracks in isotropic materials propagate in a direction perpendicular to the principle directions corresponding to the maximum principle stress at each point (see, for example, Anderson [5]). The principle directions are computed from the stress distribution

$$\vec{E}_1 = 1\hat{x} + \frac{\tau_{xy}}{\sigma_1}\hat{y} + \frac{\tau_{xz}}{\sigma_1}\hat{z} \tag{12a}$$

$$\vec{E}_2 = 1\hat{x} + \frac{\tau_{xy}}{\sigma_2}\hat{y} + \frac{\tau_{xz}}{\sigma_2}\hat{z} \tag{12b}$$

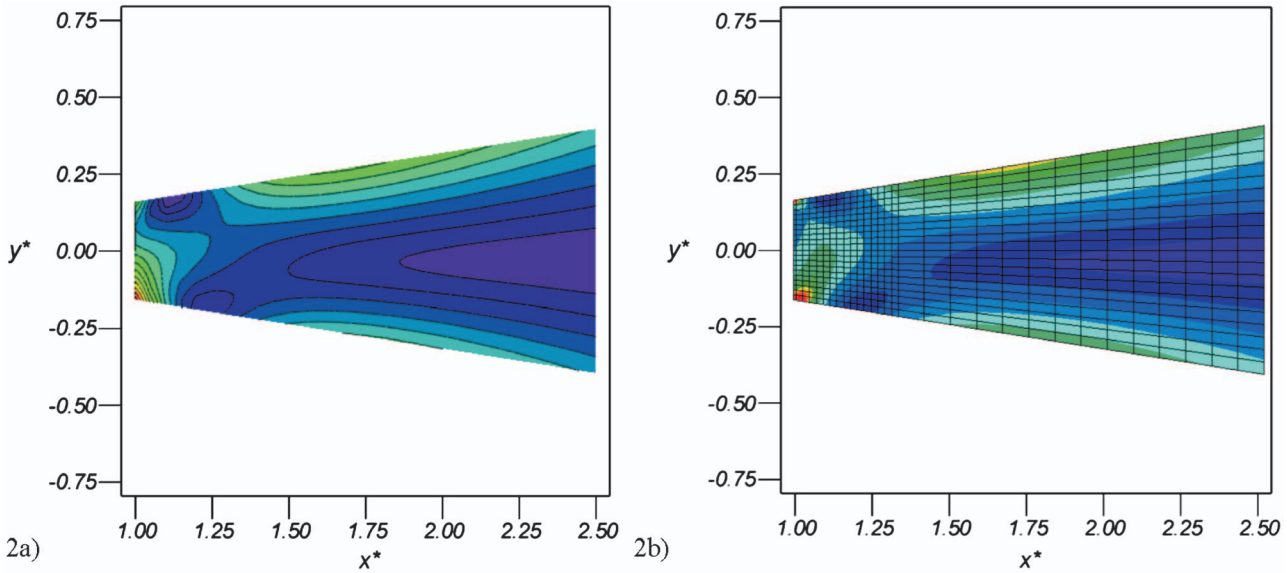


Fig. 2 von Mises Stress distribution in the pencil tip for $\alpha = 9°$, $\theta = 55°$, $\mu = 0.15$. (*a*) Vertical axial section of the pencil tip; (*b*) finite element results for the vertical axial section of the tip.

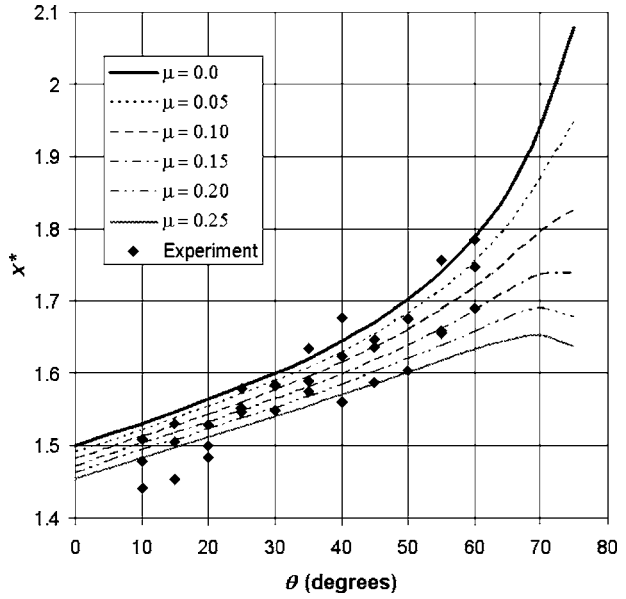


Fig. 3 Impact of friction and writing angle θ on the axial breaking position x^* (for α=9°)

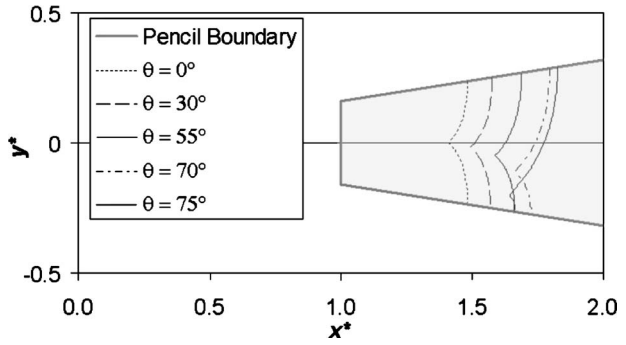


Fig. 4 Computed fracture lines in the z=0 plane as a function of θ (for α=9° and μ=0.1)

$$\vec{E}_3 = 0\hat{x} + 1\hat{y} - \frac{\tau_{xy}}{\tau_{xz}}\hat{z} \tag{12c}$$

At the point of failure, the largest principle stress is always σ_2, and is always negative. Depending on θ, $|\sigma_1|$ is sometimes larger than $|\sigma_2|$ at points near the lower edge of the pencil. Figure 4 shows the computed fracture lines in the z=0 plane for various θ. The cusp in the fracture surfaces corresponds to the point where $|\sigma_1|$ becomes larger than $|\sigma_2|$. Whether failure initiates on the upper or lower edge, the fracture is seen to always propagate towards the tip.

As the crack propagates, the stress distribution changes significantly, with corresponding changes to the principle directions, so that the above model cannot remain a valid description of the system. In actual practice, variations in the details of the fractures will result from local material impurities. Nonetheless, computation of the fracture using this method results in good agreement with experiments.

Digital Photography of Pencil Points

The experimental digital images have 8 bit accuracy and a linear magnification of 122 pixels/mm. Pencils were sharpened with an electric sharpener, and rubbed down to create a distinct ℓ ($\ell_{\text{typical}} \approx 3$ mm). In contrast to Walker [6], who examined the broken tips, this study examined the pencil cone itself, both before and after breaking. The sharpness angle was determined to be $\alpha = 9.0 \pm 0.2$ deg from the difference in the slopes of the upper and lower surfaces in these images. The value of x^* for the uppermost position on the fracture was determined by the ratio of the diameter of the pencil at that point (in pixels) divided by the diameter of the tip prior to breaking. The resulting position x^* has an uncertainty of about ±0.04. In agreement with Walker, x^* was observed to depend on θ. The measured fracture positions as seen in Fig. 3 agree well with the predictions.

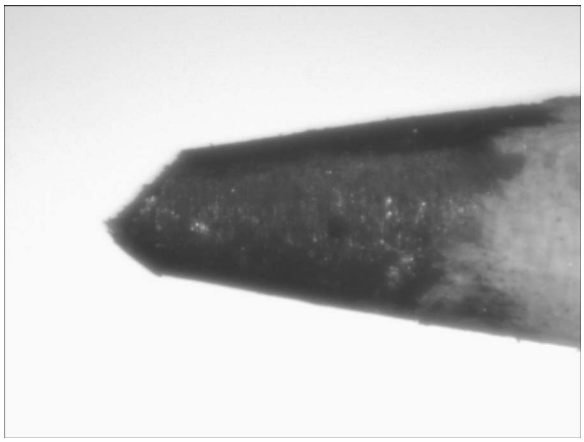


Fig. 5 Typical digital photograph of a pencil tip after breaking (θ=30°)

Figure 5 shows a typical image of a pencil tip after breaking. The formation of the cusp is evident. The cusp was observed in about two-thirds of the images examined. Images evidencing a cusp fell into two broad categories. The majority had curved surfaces on the upper and lower sides of the tip. But, several of the images containing cusps had a curved surface on the upper side of the pencil, but a short vertical fracture on the lower half. This corroborates the premise that fractures begin on the upper edge. Examination of the broken points revealed that even when the pencil had a cusp, the broken point itself usually appeared to have a primarily straight fracture surface. It is probable that the broken point itself separates into two or more pieces, the largest of which were examined by Cronquist, Cowin, and Walker.

Acknowledgment

The authors would like to thank the Department of Physics and Astronomy at SUNY Geneseo for purchasing the ALGOR software used in this study.

References

[1] Cronquist, D., 1979, "Broken Off Pencil Points," Am. J. Phys., **47**, pp. 653–655.
[2] Cowin, S. C., 1983, "A Note on Broken Pencil Points," J. Appl. Mech., **50**, pp. 453–454.
[3] Beer, F., Johnston, E., and DeWolf, J., 2002, *Mechanics of Materials*, 3rd ed., McGraw-Hill, New York, Chaps. 1–6.
[4] Gere, J. M., and Timoshenko, S. P., 1984, *Mechanics of Materials*, 2nd ed., PWS Publishers, Boston, pp. 236–238.
[5] Anderson T. L., 1995, *Fracture Mechanics: Fundamentals and Applications*, 2nd ed., CRC Press, Boca Raton, FL.
[6] Walker, J., 1979, "The Amateur Scientist," Sci. Am., **240**, pp. 158–166.

Transient Response of Transversely Isotropic Composite Plates to a Point Source

A. Mahmoud
Graduate Fellow
Mechanical and Industrial Engineering Department, University of Manitoba, Winnipeg, Manitoba R3T 5V6, Canada
e-mail: umkotbam@cc.umanitoba.ca

A. H. Shah[1]
Professor
Civil Engineering Department, University of Manitoba, Winnipeg, Manitoba R3T 5V6, Canada
e-mail: shah@cc.umanitoba.ca

S. B. Dong
Professor
Mem. ASME
Department of Civil and Environmental Engineering, University of California at Los Angles, Los Angeles, CA 90095-1593
e-mail: dong@seas.ucla.edu

In this paper, transient three-dimensional response of a transversely isotropic composite plate to a time varying point load is efficiently computed by reducing the elastodynamic equation through integral and coordinate transformations to a series of two-dimensional problems, each associated with a plane wave along a given direction in the plate. Discrete equations of a semi-analytical finite element model are solved for the thickness profile eigendata at a given frequency. Three-dimensional steady state responses in the wave number domain are formed by summing contributions from eigenmodes over propagation directions. The transient response is obtained by a numerical integration of inverse Fourier time transform of these steady state responses. Present results showed good agreement with data reported in the literature and confirmed previously observed phenomena.
[DOI: 10.1115/1.2070007]

Introduction

Recently, Mal [1] provided an overview of mechanics-based methods currently used in structural health monitoring (SHM), where wave propagation was described as one of the key methods. For laminated composite structures, where anisotropy and layered construction are involved, efficient wave propagation analysis procedures must be foremost to any improvements in effective evaluation of structural integrity of this class of structures. Herein, we present an efficient procedure for calculating transient response in a laminated composite plate caused by a time varying point source. The analysis is based on three-dimensional elasticity theory. This proposed method should advance the state-of-the-art in analysis tools by virtue of its wider range of application and greater computational efficiency.

Several techniques have been set forth for studying wave propagation in anisotropic plates under some form of forced input, where spatial Fourier transforms are employed. Transient behavior in an anisotropic laminated plate due to a point load with time step input was investigated by Liu et al. [2] using a double Fourier transform combined with one-dimensional finite elements in the thickness direction. The infinite Fourier inversion integrals were evaluated by fast Fourier transform (FFT) [3]. Mal and Lih [4] applied a multiple transform technique to compute the response of a unidirectional graphite/epoxy plate to line and concentrated surface loads. The inversion integrals in this multiple transform were evaluated numerically by a so-called adaptive surface filtering method [5] to overcome the irregular behavior in the integrand. Green and Green [6] proposed a computational method for SH motions due a point load in a nondispersive anisotropic media, which involves a single infinite integral that was evaluated by a trapezoidal rule. With regards to SHM of this class of plate structures, mention is made of Lih and Mal [7] and Banerjee et al. [8], who used a first order shear deformation plate theory. These methods provide reasonably accurate numerical results without incurring inordinate computation.

In this paper, iterated Fourier transforms are applied that convert time and the planar spatial variables into frequency and wave numbers. A subsequent coordinate rotation is then applied to recast the governing equations in the propagation direction. The three-dimensional problem is thus converted into a series of two-dimensional problems, each representing a plane wave. Moreover, the double Fourier inversion integrals can be transformed to a single infinite integral running radially and a finite integral over a full circumferential path $[-\pi \leqslant \alpha \leqslant +\pi]$. This approach can be contrasted with that of Liu et al. [2,3], where the Fourier inversions were effected by a double FFT, which were not considered herein because of inherent computational inefficiencies. Modal data representing the thickness dependence of the plane waves are determined by the method of Dong and Huang [9], which utilizes one-dimensional finite elements. The modal data are sought by subspace iteration. A range of wave numbers and/or frequencies can be obtained, where the modes of a prior wave number or frequency is used to start the subspace iteration of the next problem. Thus, this aspect of the procedure is very efficient as only a few iterations are needed for convergence. Steady state responses in the wave number domain are found by a double sum: (1) an inner sum over the normal modes and (2) an outer sum of the plane waves over a full circumferential sweep. Steady state Green's functions are then obtained by Fourier inversion involving a single infinite integral only that is evaluated analytically by Cauchy residue theorem, rather than by numerical integration [6]. Finally, the transient response is recovered numerically from their frequency domain by trapezoidal integration of the inverse Fourier time transform. Comparison of the present results with the results by Lih and Mal [5] evinced good agreement. The present method should also extend the range of application of those by Lih and Mal [7] and Banerjee et al. [8], which is based on plate theory. By three-dimensional theory, it will allow a more potent means to identify the size and shape features of surface breaking or buried flaws. It is mentioned that other finite element methods have been applied, see Prosser et al. [10]. Prosser et al.'s code is two-dimensional (2D)/three-dimensional (3D) in its capability and is more computationally intensive in comparison to the present method which uses one-dimensional finite elements. Moreover, these one-dimensional eigensolutions are evaluated once for a given composite ply layup over the range of frequencies and stored. They can then be retrieved for any number of modal synthesis computations. Application of the present technique to graphite/epoxy plates demonstrates its versatility as well as attractive computational advantages.

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received May 27, 2004; final manuscript received June 13, 2005. Review conducted by S. Govindjee.

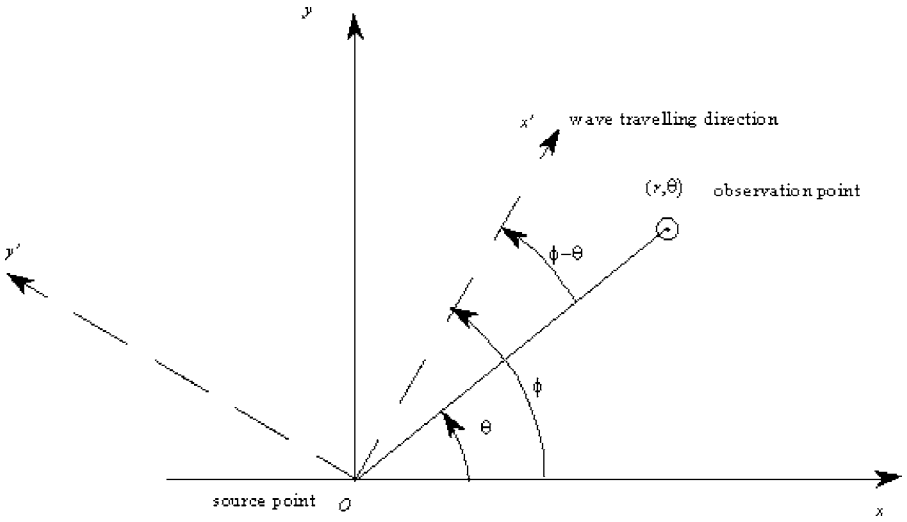


Fig. 1 Coordinate transformation into the traveling direction

Analysis

Consider a plate that is composed of linearly elastic, transversely isotropic, perfectly bonded layers. Fourier transformed displacement equation of motion in the ith layer can be written as

$$\tilde{\boldsymbol{L}}\boldsymbol{C}\tilde{\boldsymbol{L}}^T\tilde{\boldsymbol{u}} + \tilde{\boldsymbol{b}} + \rho\omega^2\tilde{\boldsymbol{u}} = \boldsymbol{0}$$

$$\tilde{\boldsymbol{L}} = \begin{bmatrix} j\xi & . & . & . & d/dz & -j\eta \\ . & -j\eta & . & d/dz & . & -j\xi \\ . & . & d/dz & -j\eta & -j\xi & . \end{bmatrix} \tag{1}$$

$\boldsymbol{C}$ is the matrix of elastic moduli, ρ is the mass density, $\boldsymbol{u} = \langle u_x\ u_y\ u_z\rangle^T$ is the displacement vector, and $\boldsymbol{b} = \langle b_x\ b_y\ b_z\rangle^T$ is the vector of body forces. An overtilde indicates spatial Fourier transformation and j is the imaginary unit, $j = \sqrt{-1}$. For a plane wave with wave number k traveling in the x' direction, where x' is oriented at an angle ϕ with the x axis, see Fig. 1, transform variables ξ and η in terms of k and ϕ are

$$\xi = k\cos\phi; \quad \eta = k\sin\phi \tag{2}$$

Transforming governing Eq. (1), into the (x', y', z) coordinates yields

$$\left(\boldsymbol{A}_1 - jk\boldsymbol{A}_2\frac{\partial}{\partial z} - k^2\boldsymbol{A}_3\frac{\partial^2}{\partial z^2} + \rho\omega^2\right)\breve{\tilde{\boldsymbol{u}}} + \breve{\tilde{\boldsymbol{b}}} = 0 \tag{3}$$

where matrices A_1, A_2, A_3, are given by

$$\boldsymbol{A}_1 = \begin{bmatrix} c'_{55} & c'_{55} & 0 \\ c'_{45} & c'_{44} & 0 \\ 0 & 0 & c'_{33} \end{bmatrix},$$

$$\boldsymbol{A}_2 = \begin{bmatrix} 0 & 0 & c'_{13} + c'_{55} \\ 0 & 0 & c'_{36} + c'_{45} \\ c'_{13} + c'_{55} & c'_{36} + c'_{45} & 0 \end{bmatrix},$$

$$\boldsymbol{A}_3 = \begin{bmatrix} c'_{11} & c'_{16} & 0 \\ c'_{16} & c'_{66} & 0 \\ 0 & 0 & c'_{55} \end{bmatrix}, \tag{4}$$

where, c'_{ij} are the elements of a $\boldsymbol{C}'$ matrix defined as:

$$\boldsymbol{C}' = \boldsymbol{\Gamma}(\phi)\boldsymbol{C}\boldsymbol{\Gamma}^T(\phi)$$

$$\boldsymbol{\Gamma}(\phi) = \begin{bmatrix} m^2 & n^2 & 0 & 0 & 0 & 2mn \\ n^2 & m^2 & 0 & 0 & 0 & -2mn \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & m & -n & 0 \\ 0 & 0 & 0 & n & m & 0 \\ -mn & mn & 0 & 0 & 0 & m^2 - n^2 \end{bmatrix}$$

$$m = \cos\phi \quad n = \sin\phi \tag{5}$$

As Eq. (3) governs a plane wave with wave number k traveling in the x' direction, the field variables are independent of y'. Thus the 3D wave propagation problem is decomposed into a series of 2D problems where all three displacement components are coupled. Using a displacement-based, Rayleigh-Ritz stiffness method [9], Eq. (3) can be recast as a system of algebraic equations, thus Eq. (3) becomes

$$(k^2\boldsymbol{K}_1 - jk\boldsymbol{K}_2 - \boldsymbol{K}_3 - \omega^2\boldsymbol{M})\breve{\breve{\boldsymbol{Q}}} = \breve{\breve{\boldsymbol{F}}} \tag{6}$$

where vector $\breve{\breve{\boldsymbol{Q}}}$ represents the nodal displacements, $\breve{\breve{\boldsymbol{F}}}$ is a consistent load vector of the body forces, and $\boldsymbol{M}$ and $\boldsymbol{K}_i$ $(i = 1, 2,)$ are mass and stiffness matrices, respectively. Note that $\boldsymbol{M}$, $\boldsymbol{K}_1$, and $\boldsymbol{K}_3$ are symmetric, whereas $\boldsymbol{K}_2$ is antisymmetric; their forms are given by Dong and Huang [9]. The matrix size of $\boldsymbol{M}$ and $\boldsymbol{K}$ is M by M, where $M = 3\times(2N_E + 1)$ with N_E as the number of elements. Concomitantly, vectors $\breve{\breve{\boldsymbol{F}}}$ and $\breve{\breve{\boldsymbol{Q}}}$ have length M.

Using the modal summation technique [11], nodal displacements can be expressed as

$$\breve{\breve{\boldsymbol{Q}}} = \sum_{m=1}^{2M} \frac{\Phi^L_{m2}\breve{\breve{\boldsymbol{F}}}\Phi^R_{m1}}{(k - k_m)B_m}, \quad B_m = \Phi^L_{m2}\boldsymbol{B}\Phi^R_{m1}$$

$$\Phi^L_m = \{\Phi^L_{m1}\ \Phi^L_{m2}\}, \quad \Phi^R_m = \{\Phi^R_{m1}\ \Phi^R_{m2}\}^T \tag{7}$$

where, Φ^L_m and Φ^R_m are the left and right eigenvectors determined by solving the eigenproblem of Eq. (6).

Consider a point load acting at $(0, 0, z_i)$, whose Fourier transform contains the components $\tilde{\boldsymbol{f}} = \{f_x\ f_y\ f_z\}^T$. In terms of (x', y', z), this load has the form

$$\breve{\bar{f}} = \{f_x \cos\phi + f_y \sin\phi \;\; -f_x \sin\phi + f_y \cos\phi \; f_z\}^T \tag{8}$$

so that the global load vector $\breve{\bar{F}}$ appears as

$$\breve{\bar{F}}^T = \{0 \;\ldots\; 0 \; \hat{\bar{f}}_x, \hat{\bar{f}}_y, \hat{\bar{f}}_z \; 0 \;\ldots\; 0\} \tag{9}$$

Then

$$\Phi^L_{m2}\hat{\bar{F}} = \bar{\psi}_1 \cos\phi + \bar{\psi}_2 \sin\phi + \bar{\psi}_3;$$

$$\bar{\psi}_1 = \psi_1 f_x + \psi_2 f_y, \quad \bar{\psi}_2 = \psi_1 f_y - \psi_2 f_x, \quad \bar{\psi}_3 = \psi_3 f_z \tag{10}$$

where, ψ_1, ψ_2, ψ_3 are corresponding elements in Φ^L_{m2}. Displacement response at any nodal point due to this load can be written from Eqs. (7) and (9) as

$$\hat{\bar{u}} = \sum_{m=1}^{2M} \frac{\bar{\psi}_1 \cos\phi + \bar{\psi}_2 \sin\phi + \bar{\psi}_3}{(k_m - k)B_m}\{U_m \; V_m \; W_m\}^T \tag{11}$$

where, U_m, V_m, W_m are elements in Φ^R_{m1}. Recasting Eq. (11) in cylindrical coordinates (r, θ, z) followed by an inverse Fourier transform and some algebraic manipulation that includes the use of Cauchy's residue and the middle point integration formula, the displacement vector for a load in the z direction can be simplified to

$$\boldsymbol{U} = \frac{1}{\pi r}\sum_{m=1}^{M} \frac{\bar{\psi}^*_3 k_m}{B^*_m}\{0 \; 0 \; W^*_m\}^T - \frac{j}{2\pi}\sum_{i=1}^{N_\alpha}\sum_{m=1}^{M} \frac{\bar{\psi}_3 k_m}{B_m}\{\bar{U}^i_m \; \bar{V}^i_m \; W_m\}^T e^{jk_m r \cos\alpha^i}\Delta\alpha \tag{12}$$

where

$$\bar{U}^i_m = U_m \cos\alpha^i - V_m \sin\alpha^i,$$

$$\bar{V}^i_m = U_m \sin\alpha^i + V_m \cos\alpha^i,$$

$$\alpha^i = i\Delta\alpha,$$

$$\alpha = \phi - \theta \tag{13}$$

and N_α is the number of sampling points used in numerical integration over α, a super asterisk indicates quantities belonging to $\alpha^i = \pi/2$.

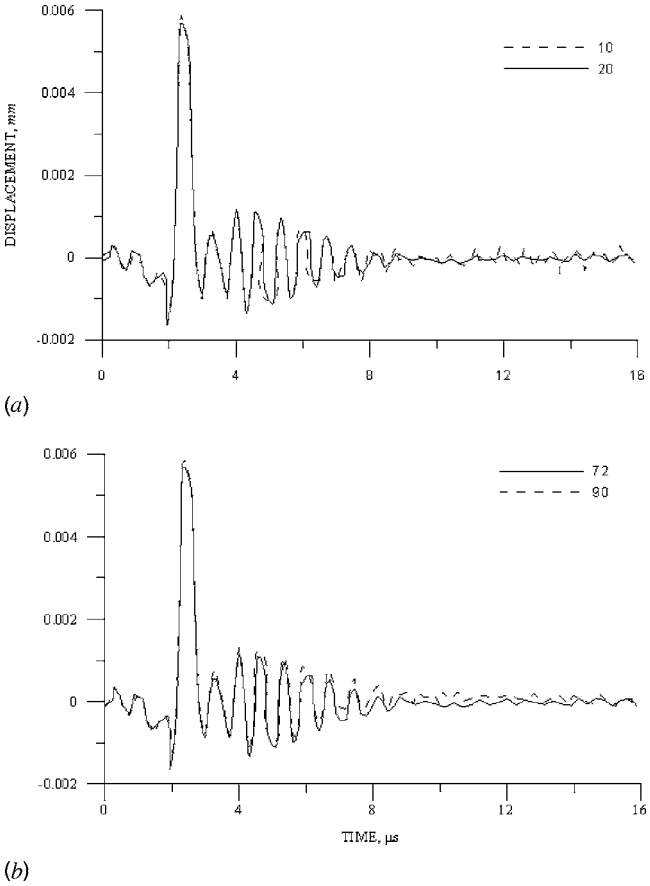


Fig. 2 Variation of displacements along the *x* direction, 4 mm from a vertical source in a unidirectional 1-mm-thick graphite/epoxy plate: (*a*) with number of sublayers (N_α=72); (*b*) with number of propagation directions (N_E=10)

Results and Discussion

Computing the transient response of a 1-mm-thick unidirectional graphite/epoxy plate to a vertical source at the origin has been reported by Lih and Mal [5]. Their results are used in this

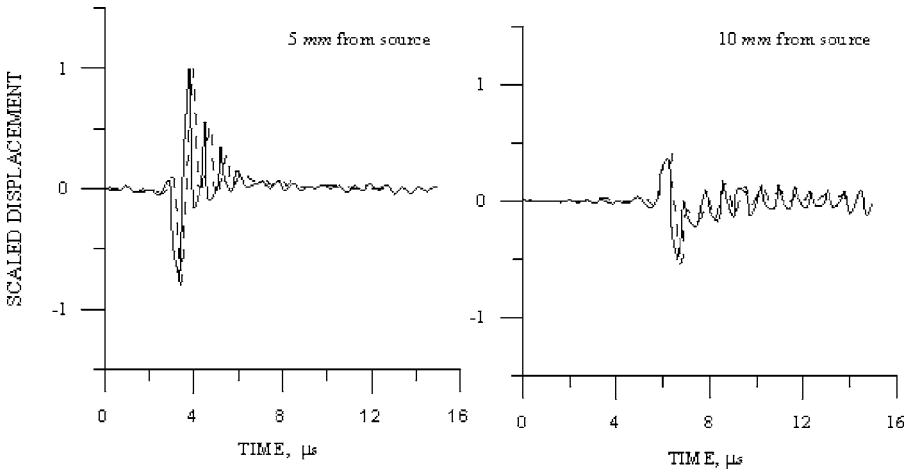


Fig. 3 Scaled normal top surface displacements along 45 deg from the *x* direction: a comparison between PWS (solid line) and Lih and Mal (dashed line)

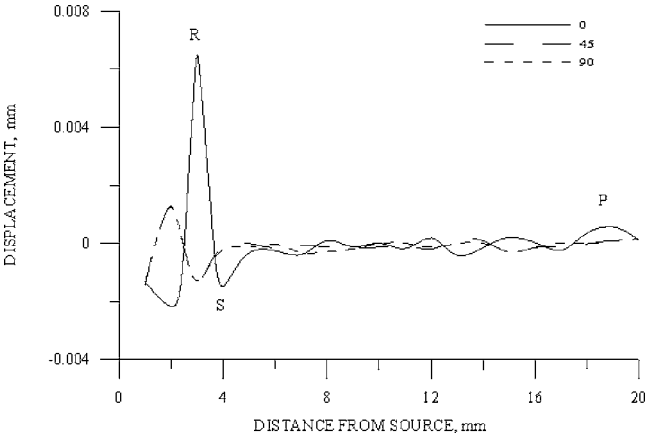


Fig. 4 Snapshots, at t=1.9 s, of top surface displacements along 45 deg, 0 deg, and 90 deg from the x direction

brief as benchmark to validate the proposed method. Properties of the graphite/epoxy as well as time history and amplitude spectrum of the source can be found therein.

First, convergence of the proposed method is illustrated by using several combinations of the number of elements, N_E, and the number of propagation directions, N_α to predict the response along x direction at 4 mm from the source. Values of N_E equal 10 and 20 at fixed N_α of 72 were used to compute the response in Fig. 2(a), the difference between the two predictions is hardly noticeable. In Fig. 2(b), with N_E fixed at 10, there is virtually no difference between predictions made by N_α=72 and N_α=90. Therefore, the values of N_E and N_α in the example to follow are prescribed to be 10 and 72, respectively. For these values, the CPU time on a 1.05 MHz Ultra Sparc III+ processor is about 67 min which is very competitive, for example, with finite element method that may take several hours or days [10].

These values of N_E and N_α are used in computing normal top surface response of a unidirectional graphite/epoxy plate 45 deg off x direction at 5 and 10 mm from the source. Figure 3 shows these predictions (scaled to the maximum absolute displacement at 5 mm) in comparison with their counterparts made by Lih and Mal's [5]. Very good agreement between the two sets of results is evident. Figure 4 shows snapshots of normal top surface displacements along 0 deg, 45 deg, and 90 deg from the x direction. It can be seen that the excitation propagates faster along the fiber direction rather than in the off fiber direction. P, S, and R indicate the arrival of pressure, shear and Rayleigh waves along x direction, respectively. A large peak corresponding to the Rayleigh wave is clear.

Concluding Remarks

A new numerical technique for determining the three-dimensional transient response of composite plates to a point load was presented, where the three-dimensional problem is decomposed into a series of two-dimensional problems, each associated with a plane wave. This technique was shown to yield excellent results when compared with existing data in the literature at a reasonably lower cost.

Acknowledgment

This work is funded by a University of Manitoba for a Graduate Fellowship granted to the first author and in part by the Natural Science and Engineering Research Council of Canada Grant No. OGP007988.

References

[1] Mal, A. K., 2004, "Structural Health Monitoring," J. Mech., **33**, pp. 11–12.
[2] Liu, G. R., Tani, J., Ohyoshi, T., and Watanabe, K., 1991, "Transient Waves in Anisotropic Laminated Plates, Part I: Theory," ASME J. Vibr. Acoust., **113**, pp. 230–234.
[3] Liu, G. R., Tani, J., Ohyoshi, T., and Watanabe, K., 1991, "Transient Waves in Anisotropic Laminated Plates, Part II: Application," ASME J. Vibr. Acoust., **113**, pp. 235–239.
[4] Mal, A. K., and Lih, S.-S., 1992, "Elastodynamic Response of Unidirectional Composite Plate to Concentrated Surface Loads: Part I," ASME J. Appl. Mech., **59**, pp. 878–886.
[5] Lih, S.-S., and Mal, A. K., 1992, "Elastodynamic Response of Unidirectional Composite Plate to Concentrated Surface Loads: Part II," ASME J. Appl. Mech., **59**, pp. 887–892.
[6] Green, E. R., and Green, W. A., 2000, "A Computational Method for Wave Propagation from a Point Load in an Anisotropic Material," Ultrasonics, **38**, pp. 262–266.
[7] Lih, S. S., and Mal, A. K., 1995, "On the Accuracy of Approximating Plate Theories for Wave Field Calculations in Composite Laminates," Wave Motion, **21**, pp. 17–34.
[8] Banerjee, S., Mal, A. K., and Prosser, W. H., 2004, "Analysis of Transient Lamb Waves Generated by Dynamic Surface Sources in Thin Composite Plates," J. Acoust. Soc. Am., **115**(5), pp. 1905–1911.
[9] Dong, S. B., and Huang, K. H., 1985, "Edge Vibrations in Laminated Composite Plates," ASME J. Appl. Mech., **52**(2), pp. 433–438.
[10] Prosser, W. H., Hamstad, M. A., Gary, J., and O'Gallagher, A., 1999, "Comparison of Finite Element and Plate Theory Methods for Modeling Acoustic Emission Waveforms," J. NDE., **18**, pp. 83–90.
[11] Liu, G. R., and Achenbach, J. D., 1995, "Strip Element Method to Analyze Wave Scattering by Cracks in Anisotropic Laminated Plates," ASME J. Appl. Mech., **62**, pp. 607–613.

Estimation of Internal Stress Profiles Using the Barkhausen Noise Effect—Application: Verification to the Homogeneity of the Shot Peening Process Performed on the Raceways of Aerospace Bearing Rings

S. Desvaux
IEMN, Département OAE (UMR CNRS 8520),
Université de Valenciennes et du Hainaut Cambrésis,
59313 Valenciennes Cedex 9,
France and SNFA,
Z.I. No. 2 Batterie 900 Rouvignies,
59309 Valenciennes, France

M. Duquennoy
e-mail: marc.duquennoy@univ-valenciennes.fr

M. Ouaftouh

IEMN, Département OAE (UMR CNRS 8520),
Université de Valenciennes et du Hainaut Cambrésis,
59313 Valenciennes, Cedex 9, France

J. Gualandri
SNFA, Z.I. No. 2 Batterie 900 Rouvignies,
59309 Valenciennes, Cedex 9, France

M. Ourak
IEMN, Département OAE (UMR CNRS 8520),
Université de Valenciennes et du Hainaut Cambrésis,
59313 Valenciennes, Cedex 9, France

In this paper, we present a nondestructive method based on the Barkhausen noise effect for estimating internal stress profiles, with the goal of controlling bearing raceways to determine the homogeneity of the shot peening process. Given the industrial imperative to characterize parts, our method concentrates on stress profiles in the first 60 μ*m under the surface, in steps of 5 or* 10 μ*m. Our results show the efficient use of Barkhausen noise to estimate the residual stress profiles of bearing raceways. A mapping of the stress profiles measured by Barkhausen noise was completed for the entire width of an outer ring raceway. This study confirmed that it is possible to use Barkhausen noise to test the uniformity of the shot peening treatment and to inspect the homogeneity of surface and subsurface stresses on the bearing rings. The proposed method is rapid, well suited to industrial imperatives connected to on-line measurement, and easily adapted to the circular geometries of the bearing rings.* [DOI: 10.1115/1.2073329]

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received August 30, 2004; final manuscript received June 1, 2005. Review conducted by K. M. Liechti.

1 Introduction

Bearings in aeronautic engines are subject to extremely severe running conditions that lead to a hertzien fatigue phenomenon. As a result of this fatigue, micro-cracks may appear on the contact area. Without intervention, these cracks can develop until they cause bearing malfunction, which may lead to engine failure. Measuring internal stress profiles after the shot peening operation is thus essential to make sure that the mechanical treatment is uniform over the entire width of the bearing raceway. Otherwise, the disparity of bearing raceway stresses could reduce the bearing fatigue strength. This paper presents a nondestructive method for identifying the internal stress profiles of the contact areas between the balls or rollers and their raceways. The method proposed is consistent enough to evaluate the residual stresses and fast enough to ensure that the cost of inspecting all the parts is not prohibitive.

The proposed method is based on the phenomenon of Barkhausen noise (BN), which is studied in different frequency analysis ranges in order to determine the internal stress gradient. The stress just under the surface is represented by a BN parameter calculated in a high-frequency analysis range. The subsurface stresses are estimated by increasing the lower bound of the analysis section to provide information on the BN coming from under the surface. Thus, the stress profile is reconstructed through successive searches for stresses at the different analysis depths, with each stress level being determined through a BN parameter that is calculated in a typical frequency analysis range for a given depth. The homogeneity of the shot peening process is verified on the bearing raceways by setting up a series of profiles for the entire width of the raceway. Given the success of our current research, we have begun to consider the possibility of transferring this method to an industrial environment.

2 Estimation of Internal Stress Profiles Using Barkhausen Noise

The phenomenon of BN and the influence of stress on the latter have been well described [1,2]. Many authors have shown that BN can be used to evaluate surface residual stresses [3,4]. Though some studies have focused on the estimation of residual stress profiles, most of these results are qualitative [5,6]. Very few attempts have been made to quantify residual stress profiles [7,8]. Several authors have shown that BN is subject to an exponential attenuation in relation to the distance crossed by the electromagnetic wave moving through the material. These authors used this attenuation to evaluate surface stress or to average stress for a given thickness [9–11]. In this article, profiles of the first 60 μm under the surface, in steps of 5 or 10 μm, can be quantified because there are reference values coming from the X-ray diffraction measurements.

In ferromagnetic materials, electromagnetic wave propagation is always subject to an attenuation that is a function of the frequency. This attenuation phenomenon is essentially due to eddy currents in a semi-infinite medium [12]. An elementary Barkhausen event, noted $e(f)$, can be expressed in terms of frequency according to the formula

$$e(f) = p(f) \times \alpha(f,d) = \frac{A(\beta - 2j\pi f)}{\beta^2 + (2\pi f)^2} \times \exp(-\alpha \cdot d) \qquad (1)$$

where $\alpha = \sqrt{\pi\mu\sigma_C f}$ is the attenuation factor depending on the frequency f, on the permeability μ, and on the electrical conductivity σ_C; $p(f)$ is the Fourier transformation of an elementary Barkhausen event; A is the amplitude of the impulse; β is the base frequency of the event corresponding to the reciprocal of the average duration of the elementary Barkhausen event (up to 37% of the amplitude); and d is the distance between the surface of the part and the broadcast zone of the Barkhausen event.

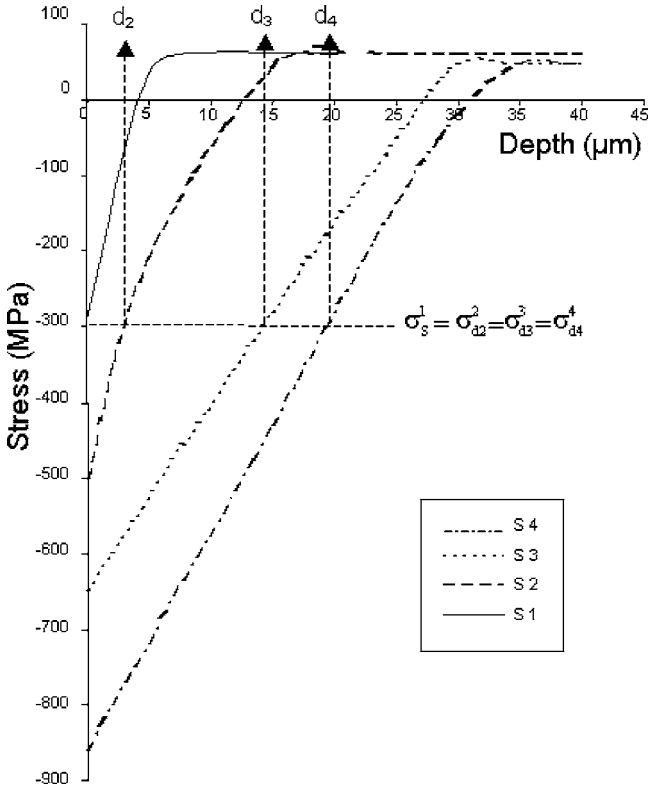


Fig. 1 Residual stress profile for the four samples of reference estimated by X-ray diffraction

Several studies have shown that the spectrum of a BN is a function of its broadcast depth [13]. On the other hand, Tiitto [9] simulated frequency spectra $P(f)$ for frequencies between f1 and f2 for different thicknesses d using the formula

$$P(f) = \left(\frac{\int_0^d g(f)\exp^{-\sqrt{\pi.\sigma_C.\mu.f}.x} dx}{\int_0^d g(f)\exp^{-\sqrt{\pi.\sigma_C.\mu.f_1}.x} dx} \right)^2 \quad (2)$$

for a random noise $g(f)=1$. Results show that the greater the penetration depth d, the narrower the frequency bandwidth. Thus, based on Titto's work [9], for a volume of material presenting a given level of residual stress, either at or under the surface, the receiving coil of the sensor situated at the surface of the material obtains a BN with a certain attenuation that will modify the content of frequency spectrum. This phenomenon makes it possible to reconstruct the stress profile. To establish a material's stress profile, it is necessary to identify the stress levels at a variety of depths under the surface. In the method that we propose, each depth is determined by the choice of a specific frequency range; the stress level is obtained from the analysis of the BN spectrum in this range.

In our experiments, four samples of 52100 steel were used. These samples had an identical microstructure and had undergone fairly harsh shot peening operations, giving them each a specific compressive stress profile. The different stress profiles were identical in form: each had a compressive stress field at the surface and a stress level that decreased (in absolute value) nearly linearly with depth before becoming tensile, showing a stress level close to zero (Fig. 1). Prior to beginning our experiments, X-ray diffraction was used to measure the stress gradient of each sample, up to the depth at which the stress became tensile. This method is considered to be semi-destructive because, given that x rays can only go through several micrometers of the material, some chemical etching is necessary to obtain information about the stress at certain depths. Samples were labeled according to their level of stress (S1 to S4). Sample S1 had the weakest surface compressive stress, while sample S4 had the strongest surface compressive stress.

The "μ scan 500-1" equipment for BN measurement consists of a magnetizing generator, a sensor, and a computer with a specific signal processing software. This software acquires and analyzes the BN signal. The S1-138 sensor, which is made of magnetizing polar parts and a receipt ferrite that is mounted on springs to ensure a good contact on varied surfaces, has a frequency bandwidth between 3 kHz and 1 MHz. The magnetic excitation is accomplished by the magnetizing generator via the application of a sinusoidal voltage whose amplitude ranges from 0 to 20 V and whose frequency ranges from 1 to 1000 Hz. In our application, the specific software allows us to process the BN signal in both the time and frequency domains on frequency analysis bands that vary from 3 kHz to 1 MHz.

Because the high-frequency components of the Barkhausen signal attenuate rapidly as the wave progresses through the material, analyzing the components with the highest frequency made it possible to concentrate the part of the Barkhausen signal that is typical of the material's condition at the surface, and therefore its surface stress. This bandwidth $[F_L;F_H]$ corresponds to that of the receiving coil. The upper bound of the section is determined by the bandwidth of the receiving coil, F_H, but the lower bound of the analysis section, F_S, had to be determined. To determine this lower bound, we took ten BN measurements for each sample and determined the amplitude of each spectrum measurement. After calculating the average amplitude for each sample, thus establishing the relationship of these measurements with the level of surface residual stress, we looked for the best regression curves (linear, exponentials,…) that would allow the evolution of the BN parameters (amplitude) to be characterized in terms of the surface stress. Finally, we shifted the lower bound of the analysis frequency range, and we repeated all these steps until we obtained the optimum correlation coefficient. A computerized search using an appropriate frequency range allowed us to demonstrate a linear relationship between the surface residual stress and the BN parameter. We validated the method on different bearing rings that had undergone various shot peening operations [14].

We repeated the steps described in the paragraphs above to obtain the stress at each given depth, placing the different stresses, depths, and amplitudes on the spectrums. Sample S1 has a surface residual stress level equal to σ_S^1, and samples S2, S3, and S4 has the same residual stress level at depths d_2, d_3, and d_4, called $\sigma_{d_2}^2$, $\sigma_{d_3}^3$, and $\sigma_{d_4}^4$, respectively. Given the attenuation caused by compressive stress and because $\sigma_S^1 < \sigma_S^2 < \sigma_S^3 < \sigma_S^3$, the depths are different and are classified as follows: $d_2 < d_3 < d_4$ (Fig. 1). α_S^1 is the amplitude of the frequency spectrum in the analysis range that characterizes the surface of S1, whose surface stress level is σ_S^1. This amplitude α_S^1 is thus characteristic of the stress level σ_S^1. Since $\sigma_S^1 = \sigma_{d_2}^2 = \sigma_{d_3}^3 = \sigma_{d_4}^4$, the BN emissions from S2, S3, and S4 all present the same amplitude: α_S^1. (The emission of the volume of material in sample S2 whose stress level is $\sigma_{d_2}^2$ at depth d_2=the emission of the volume of material in sample S3 whose stress level is $\sigma_{d_3}^3$ at depth d_3=the emission of the volume of material of the sample 4 whose stress level is $\sigma_{d_4}^4$ at depth d_4.) Because of the attenuation, decreasing the lower bound of the frequency analysis range for every spectrum of samples S2, S3, and S4 produces the amplitude shown in Fig. 2. Therefore, the frequency analysis ranges $[f_{d_2};F_H]$, $[f_{d_3};F_H]$, and $[f_{d_4};F_H]$ are characteristic of the stress levels corresponding to depths d_2, d_3, and d_4, respectively.

In order to validate our hypothesis, according to which the given stress level of a volume of material is characteristic of an amplitude in the BN spectrum, it was once again necessary to analyze the frequency ranges of the four stress profiles determined

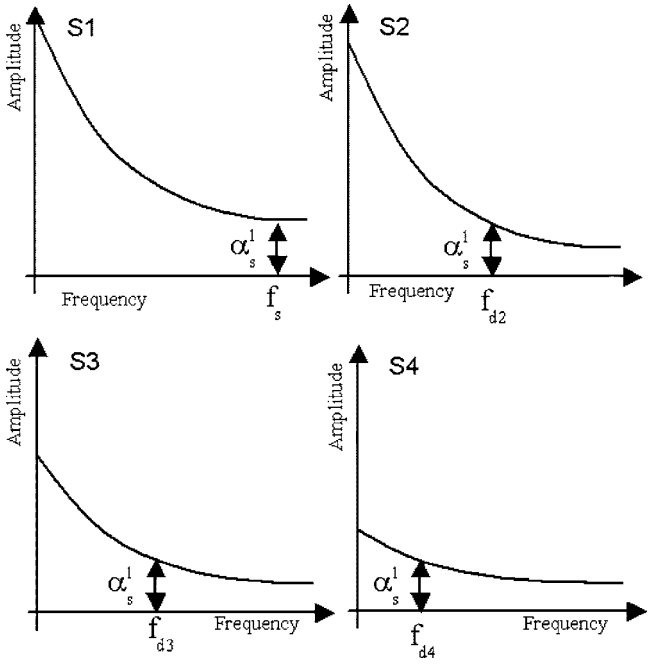


Fig. 2 Principle of determination of frequency analysis ranges on schematized BN spectrum on samples of reference

by X-ray diffraction. This time, however, we analyzed them for a certain depth value. Let σ_i^j equal stress levels, where j is the number of the sample, Sj, and i is the analysis depth, d_i. From the characteristic network depicted in Fig. 1, we calculated the stress σ_i^j of each sample for a given depth, d_i. In the example shown in Fig. 3(*a*), depth equals 5 μm. In order to determine the amplitudes α_i^j that corresponded to the different stresses σ_i^j [Fig. 3(*b*)], we referred to the curve showing amplitude in relation to stress. By connecting these different amplitudes α_i^j to the BN spectrum of each sample, it was possible to determine the low frequencies f_i^j of the analysis ranges that correspond to the amplitudes α_i^j [Fig. 3(*c*)] as well as those that, according to our hypothesis, correspond to the analysis depth *di*. We generated 30 spectrums for each sample at three different inspection depths (5, 10, or 15 μm). By comparing the low frequencies of the analysis range strip ($f_i^1, f_i^2, f_i^3, f_i^4$), we were able to verify the relevance of the analysis range in relation to the stress level at the given depths.

The low frequencies of the frequency analysis ranges (betweeen 300 kHz and 1 MHz) are specific to the four samples of 52100 steel. The entire range is not given here for reasons of confidentiality. However, the lower bounds for each depth are very close ($f_5^1 \cong f_5^2 \cong f_5^3 \cong f_5^4$; $f_{10}^1 \cong f_{10}^2 \cong f_{10}^3 \cong f_{10}^4$; $f_{15}^1 \cong f_{15}^2 \cong f_{15}^3 \cong f_{15}^4$), and the standard deviations of the lower bound are 5, 6, and 10 kHz, respectively. Thus, taking the low frequency values of the lower bounds into account, these standard deviations allow us to analyze the results and to verify the reproducibility of the method. The results show that this method is satisfactory: given four samples with different stress profiles, the same analysis range is systematically found to characterize the four different internal stress levels at the same depth. Standard deviations obtained on the lower bounds for each depth are weak for the three depths tested. The close correlation between frequency range and analysis depth was also verified for other raceways, by testing different rings made of diverse materials, and some of the results of these tests are presented in the last section of this paper. Obviously, because the ranges are specific to the magnetic microstructures of materials, the stress profile samples used to determine the analysis ranges had to be identical in shape. Since each batch of tested parts has its own "metallurgic history" (composition, thermal cycles, thermal and mechanical treatments, etc.), a calibration phase was necessary for each one in order to guarantee profile homogeneity.

There appears to be an exponential evolution of the lower bound in relation to the depth of inspection. This evolution is quite important because it allows other frequency analysis ranges, corresponding to other inspection depths (20, 30, 50, and 60 μm), to be determined through extrapolation. This extrapolation is significant for two reasons. First, in order for the method to be accepted by industry, it must be both rapid and simple to use. As stated above, the analysis ranges are determined during a calibra-

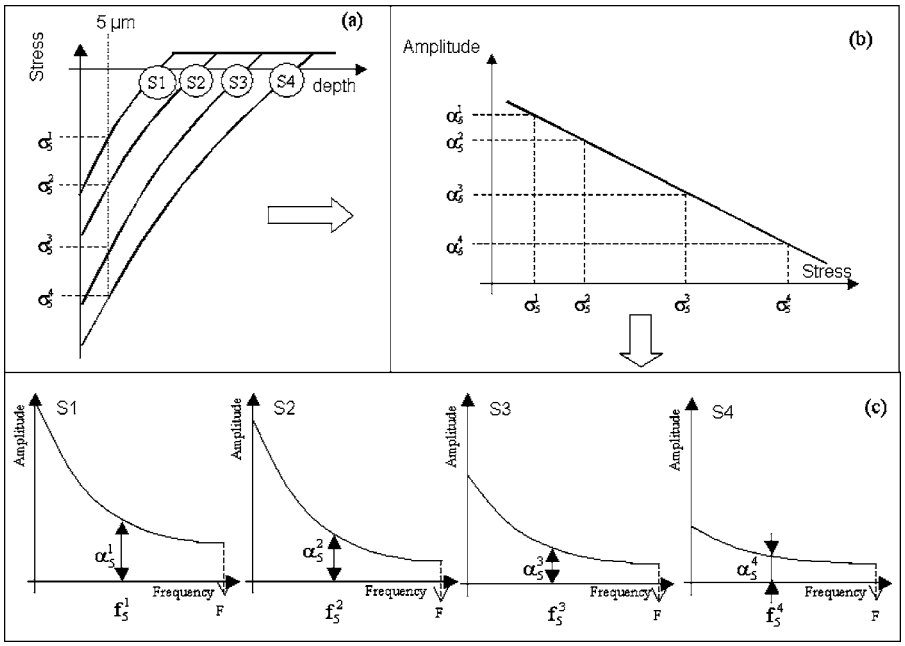


Fig. 3 Method for the determination of frequency analysis ranges with schematized profiles and the BN corresponding spectrum

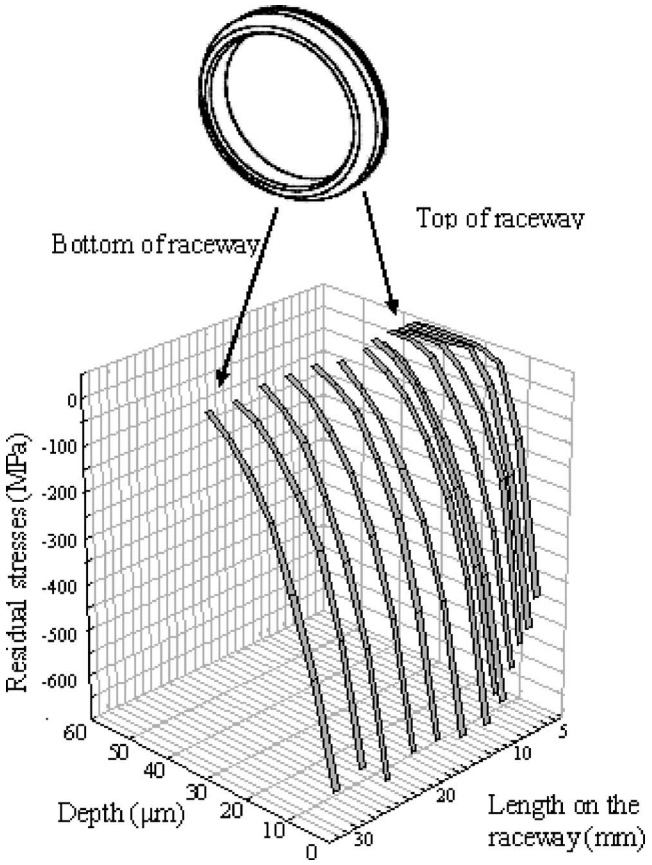


Fig. 4 Mapping of stress profiles by Barkhausen noise on the width of a bearing ring

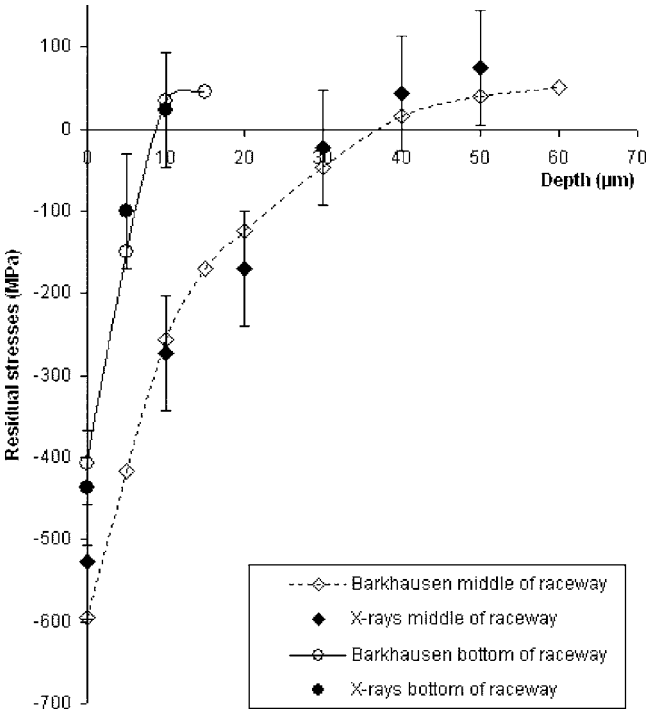


Fig. 5 Evaluation of two stress profiles by Barkhausen noise and X-ray diffraction methods

tion phase and are specific to the magnetic microstructures of the materials, and so, because of their "metallurgical history," are also specific to each batch of parts to be tested. If it is possible, given the lower bound of the first frequency, to deduce the others by extrapolating the curve of this lower bound in relation to depth, much time can be saved in the calibration phase. Second, finding the frequency corresponding to the lower bound of a BN spectrum amplitude is sometimes difficult because spectrum fluctuations are more and more striking as the frequency of the lower bound decreases. Because of this fluctuation, the standard deviation of a lower bound value is greater for f^i_{10} than for f^i_5 and for f^i_{15} than for f^i_{10}.

3 Testing of Homogeneity of Shot Peening Operation on Aeronautic Bearings

In order to characterize the homogeneity of residual stress profiles, we checked bearing rings by creating a series of profiles for the entire width of the raceway. The shot peening used to control the stress profile in raceways employs ultrasonic sounds that cause ball movement in a cavity. Depending on the parameters of this shot peening process, different stress profiles can be obtained. The microstructure effect introduced by shot peening is considered along with stress because these two phenomena are intimately linked. However, this dual aspect is ultimately unimportant since the stress estimation is compared with X-ray measurements.

For the example presented in this article, we performed 14 profiles for the area between the top and the bottom of the raceway. Figure 4 depicts the mapping of the BN-measured stress profiles for the entire width of an outer ring raceway of M50 steel. Significant deviations can be noted. Surface stresses vary from −600 to −400 MPa between the top and the bottom of the raceway; the transformation into tensile stresses occurs at approximately 35 μm under the surface on the bottom of the raceway, whereas it occurs at 8 μm under the surface at the top of the raceway. This dissimilarity was confirmed by comparing the BN-measured stress profiles of the whole batch of rings, with the profiles of the middle and top of the raceways of two rings, which were measured using X-ray diffraction (Fig. 5). As the stress profile obtained for the top of the raceway clearly shows, such divergence can be very important. In this example, the internal stresses in that part of the raceway cannot guarantee the good, long-lasting fatigue strength of the material. By showing that the stress profiles were not similar for the whole width of the raceway, our mapping revealed the nonhomogeneity of this shot peening operation, which made us question the homogeneity of the shot peening process in general. After conversation with our supplier, we identified a dissimilarity in the geometry of the shot peening cavity and were able to modify it to allow a homogeneous operation. Clearly, Barkhausen analysis offers a consistent and appropiate nondestructive method that, in this case, allowed us to correct the problem of nonhomogeneity in the shot peening process.

4 Conclusion

In this paper, we presented a nondestructive method for estimating residual stress profiles, based on the BN effect. This method was validated on several batches of bearings. Given the industrial imperative to characterize parts, we concentrated our efforts on stress profiles in the first 60 μm under the surface. We show the efficiency of the Barkhausen noise effect for estimating these residual stress profiles in bearing raceways. This study clearly shows that, in the future, it will be possible to inspect the homogeneity of surface and subsurface stresses on the bearing rings at any stage of the manufacturing process, regardless of the treatments that have been effected. Given that safety standards today are not always met through a statistical control of the parts, we propose a fast magnetic method that is compatible with industrial needs associated with on-line measurement and can inspect 100% of the parts produced. Such verification is more and more necessary; in fact, it is essential in order to guarantee an extended life span for engine bearings.

References

[1] Tiitto, S., 1977, "On the Influence of Microstructure on Magnetization Transitions in Steel," Acta Polytech. Scand., Appl. Phys. Ser., **119**, pp. 0–80.

[2] Cullity, B. D., 1972, "*Introduction to Magnetic Materials*," Addison-Wesley, London.

[3] Pasley, R. L., 1969, "Barkhausen Effect—An Indication of Stress," *29th National Fall Conference of the American Society for Nondestructive Testing*, Philadelphia, PA, pp.13–16.

[4] Tiitto, S., 1988, "Magnetoelastic Testing of Uniaxial and Biaxial Stresses," *International Conference Residual Stresses*. Soc. Française de Métallurgie, Nancy.

[5] Hillmann, R., 1998, "Combined NDT Testing of Camshafts," *Conf. Proc. Barkhausen Noise and Micromagnetic Testing-I*, ICBM 1, Hannover, Germany, September, pp. 77–84.

[6] Tiitto, K., 1991, "Testing Shot Peening Stresses in the Field," *The Shot Peener*, International newsletter for shot peening-surface finishing industry, 4(4).

[7] Grum, J., Zerovnik, P., and Fefer D., 1999, "Determination of Residual Stresses After Heat Treating and Grinding by Means of the Magnetic Barkhausen Noise," *Conf. Proc. Barkhausen Noise and Micromagnetic Testing-II*, ICBM 2, Newcastle, United Kingdom, October, pp. 141–152.

[8] Friemuth, T., and Mandrysch, T., 1999, "Micromagnetic In-Process Surface Integrity Analysis of Ground Workpieces," *Conf. Proc. Barkhausen Noise and Micromagnetic Testing-II*, ICBM 2, Newcastle, United Kingdom, October, pp. 101–113.

[9] Tiitto S., and Saynajakangas S., 1975, "Spectral Damping in Barkhausen Noise," IEEE Trans. Magn., **11** (6), pp. 1666–1675.

[10] Lu, J., James, M., and Roy, G., 1996, *Handbook of Measurement or Residual Stresses*, Society for Experimental Mechanics, Inc. (Ed.), The Fairmont Press, Lilburn, GA, pp. 225–231.

[11] Theiner, W. A., 1998, "Physical Basis of Micromagnetic Methods and Sensor Systems and Their Application Areas," *Conf. Proc. Barkhausen Noise and Micromagnetic Testing-I*, ICBM 1, Hannover, Germany, September, Vol.1-2, pp. 197–218.

[12] Perdrix, M., 1990, "Contrôle Non Destructif Par Courants De Foucault, La Théorie Des Courants De Foucault," Cours du CAST, Niveau III.

[13] Bloor, D., 1963, "Domain Size in Polycristalline Iron," Proc. Phys. Soc., **81**, pp. 493–497.

[14] Desvaux, S., Duquennoy, M., Gualandri, J., and Ourak, M., 2003, "The Evaluation of Surface Residual Stress in Aeronautic Bearings Using the Barkhausen Noise Effect," NDT & E Int., **37**, pp. 9–17.

Journal of Applied Mechanics

Discussion

Discussion: "Damage Modeling in Random Short Glass Fiber Reinforced Composites Including Permanent Strain and Unilateral Effect" (Mir, H., Fafard, M., Bissonnette, B., and Dano, M. L., 2005, ASME J. Appl. Mech., 72, pp. 249–258)

Noël Challamel
e-mail: noel.challamel@insa-rennes.fr

Christophe Lanos
e-mail: christophe.lanos@insa-rennes.fr

Charles Casandjian
e-mail: charles.casandjian@insa-rennes.fr

Laboratoire de Génie Civil et Génie Mécanique (LGCGM),
INSA de Rennes,
20, avenue des Buttes de Coësmes,
35043 Rennes cedex, France

*The paper of Mir et al. [ASME J. Appl. Mech., **72**, pp. 249–258 (2005)] analyzes the deformation behavior of random short glass fiber composites using a continuum damage mechanics model incorporating permanent strain and unilateral effect. Unilateral effect is a difficult topic [see, for instance, J. L. Chaboche, Int. J. Damage Mech., **2**, pp. 311–329 (1993)] and it seems to us that the model of Mir et al. still raises some questions.*
[DOI: 10.1115/1.2126694]

1 Analysis of the Complementary Elastic Energy

The complementary elastic energy is assumed by authors to be (see Eq. (2) of [1])

$$U^e(\underline{\underline{\sigma}},\underline{\underline{D}}) = \frac{1}{2E}\,\mathrm{tr}\left[\left(\underline{\underline{\tilde{\sigma}}}\right)^{+^2}\right] + \frac{1}{2E}\,\mathrm{tr}\left[\left(\underline{\underline{\tilde{\sigma}}}\right)^{-^2}\right] + \frac{1}{2E}\left\{\mathrm{tr}\left[\left(\underline{\underline{\tilde{\sigma}}}\right)^2\right] - \tilde{\sigma}_{ii}^2\right\} - \frac{\nu}{2E}\left\{\left[\mathrm{tr}\left(\underline{\underline{\tilde{\sigma}}}\right)\right]^2 - \mathrm{tr}\left[\left(\underline{\underline{\tilde{\sigma}}}\right)^2\right]\right\}$$

The projector tensors are introduced as in Eq. (3) of [1]:

$$\underline{\underline{\sigma}}^+ = \underline{\underline{\underline{\underline{P}}}}^+\left(\underline{\underline{\sigma}},\underline{\underline{D}}\right):\underline{\underline{\sigma}} \quad \text{and} \quad \underline{\underline{\sigma}}^- = \underline{\underline{\underline{\underline{P}}}}^-\left(\underline{\underline{\sigma}},\underline{\underline{D}}\right):\underline{\underline{\sigma}} \tag{A1}$$

It can be interesting to simplify the model in order to check some properties in simple cases. One can assume, for instance, that the damage tensor and the stress tensor have the same principal directions. In this case, the projection tensor only depends on the stress tensor as

$$\underline{\underline{\sigma}}^+ = \underline{\underline{\underline{\underline{P}}}}^+\left(\underline{\underline{\sigma}}\right):\underline{\underline{\sigma}} \quad \text{and} \quad \underline{\underline{\sigma}}^- = \underline{\underline{\underline{\underline{P}}}}^-\left(\underline{\underline{\sigma}}\right):\underline{\underline{\sigma}} \tag{A2}$$

This is the classical projection tensor introduced by Ortiz [2]. One can, moreover, assume that the material in undamaged:

$$\underline{\underline{D}} = \underline{\underline{0}} \Rightarrow \underline{\underline{\underline{\underline{M}}}}\left(\underline{\underline{D}}\right) = \underline{\underline{\underline{\underline{1}}}} \Rightarrow \underline{\underline{\tilde{\sigma}}} = \underline{\underline{\sigma}}$$

The complementary elastic energy is then reduced to

$$U^e(\underline{\underline{\sigma}}) = \frac{1}{2E}\,\mathrm{tr}\left[\left(\underline{\underline{\sigma}}\right)^2\right] + \frac{1}{2E}\left\{\mathrm{tr}\left[\left(\underline{\underline{\sigma}}\right)^2\right] - \sigma_{ii}^2\right\} - \frac{\nu}{2E}\left\{\left[\mathrm{tr}\left(\underline{\underline{\sigma}}\right)\right]^2 - \mathrm{tr}\left[\left(\underline{\underline{\sigma}}\right)^2\right]\right\} \quad \text{with } \underline{\underline{\sigma}}^+ + \underline{\underline{\sigma}}^- = \underline{\underline{\sigma}}$$

It can be noticed that the function U^e is not an isotropic tensorial function. It is clear, for instance, that the function

$$f(\underline{\underline{\sigma}}) = \sigma_{ii}^2$$

is not an isotropic tensorial function. It is sufficient to show that an orthogonal tensor can be found such as [3]

$$\exists \underline{\underline{Q}} / f\left(\underline{\underline{Q}}\,\underline{\underline{\sigma}}\,\underline{\underline{Q}}^T\right) \neq f\left(\underline{\underline{\sigma}}\right)$$

In this simplified case, the elastic constitutive law is given by

$$\underline{\underline{\epsilon}} = \frac{\partial U^e\left(\underline{\underline{\sigma}}\right)}{\partial \underline{\underline{\sigma}}} \Rightarrow \underline{\underline{\epsilon}} = \frac{1+\nu}{E}\underline{\underline{\sigma}} - \frac{\nu}{E}\left(tr\underline{\underline{\sigma}}\right)\underline{\underline{1}} + \frac{1}{E}\left(\underline{\underline{\sigma}} - \sigma_{ii}\underline{e_i} \otimes \underline{e_i}\right) \tag{A3}$$

The authors assume that this relation is only true in the principal damage directions which are not necessarily the same as the one of the stress tensor (in the general case). For instance, for the undamaged material, this relation should be verified whatever the directions, and the classical elastic linear isotropic relation is not found.

2 On the Projection Derivative Operator

We think that there are some difficulties in the derivation of the projection tensor. It has been shown, for instance, that

$$\frac{\partial \underline{\underline{\sigma}}^+}{\partial \underline{\underline{\sigma}}}:\underline{\underline{\sigma}}^+ = \underline{\underline{\sigma}}^+ \text{ and } \frac{\partial \underline{\underline{\sigma}}^-}{\partial \underline{\underline{\sigma}}}:\underline{\underline{\sigma}}^- = \underline{\underline{\sigma}}^-$$

when the damage principal axes coincide with the one of the stress tensor, as it is the case in Eq. (A2) [4]. This relation is implicitly used in most stress-based anisotropic damage modeling including unilateral effects (see, for instance, [5]). However, this relation is no more true in the case considered by authors, and given by Eq. (A1), when the principal axes of the damage tensor do not coincide with the one of the stress tensor. Following the reasoning of [4], it can be shown for a two-dimensional problem that

$$\sigma_{11}{}^+ = \frac{h_I\sigma_I + h_{II}\sigma_{II}}{2} + \frac{h_I\sigma_I - h_{II}\sigma_{II}}{2}\cos 2\theta$$

$$\sigma_{22}^{+}=\frac{h_I\sigma_I+h_{II}\sigma_{II}}{2}-\frac{h_I\sigma_I-h_{II}\sigma_{II}}{2}\cos 2\theta$$

$$\tau_{12}^{+}=\left(h_I\sigma_I-h_{II}\sigma_{II}\right)\sin 2\theta$$

with

$$\theta-\theta_0=\frac{1}{2}\arctan\frac{\tau_{12}}{\sigma_{11}-\sigma_{22}}\quad\text{and}\quad\tau_{12}=\sigma_{12}+\sigma_{21}$$

θ_0 is the rotation value between the principal axes of the stress tensor and the one of the damage tensor. h_I and h_{II} are two Heaviside functions of the principal stresses:

$$h_I=H\left(\frac{\sigma_{11}+\sigma_{22}}{2}+\frac{\sigma_{11}-\sigma_{22}}{2\cos\left(2\theta-2\theta_0\right)}\right)$$

$$h_{II}=H\left(\frac{\sigma_{11}+\sigma_{22}}{2}-\frac{\sigma_{11}-\sigma_{22}}{2\cos\left(2\theta-2\theta_0\right)}\right)$$

The Jacobian matrix (J) can be introduced as

$$\left(J\left(\sigma_{11},\sigma_{22},\tau_{12}\right)\right)=\begin{pmatrix}\frac{\partial\sigma_{11}^{+}}{\partial\sigma_{11}} & \frac{\partial\sigma_{22}^{+}}{\partial\sigma_{11}} & \frac{\partial\tau_{12}^{+}}{\partial\sigma_{11}}\\ \frac{\partial\sigma_{11}^{+}}{\partial\sigma_{22}} & \frac{\partial\sigma_{22}^{+}}{\partial\sigma_{22}} & \frac{\partial\tau_{12}^{+}}{\partial\sigma_{22}}\\ \frac{\partial\sigma_{11}{}^{+}}{\partial\tau_{12}} & \frac{\partial\sigma_{22}^{+}}{\partial\tau_{12}} & \frac{\partial\tau_{12}^{+}}{\partial\tau_{12}}\end{pmatrix}$$

In this case, the Jacobian can be more easily evaluated for $\theta=\theta_0$:

$$\frac{\partial\underline{\underline{\sigma}}^{+}}{\partial\underline{\underline{\sigma}}}:\underline{\underline{\sigma}}^{+}\big|_{\theta=\theta_0}=\begin{pmatrix}\frac{\partial\sigma_{11}^{+}}{\partial\sigma_{11}} & \frac{\partial\sigma_{22}^{+}}{\partial\sigma_{11}} & \frac{\partial\tau_{12}{}^{+}}{\partial\sigma_{11}}\\ \frac{\partial\sigma_{11}{}^{+}}{\partial\sigma_{22}} & \frac{\partial\sigma_{22}^{+}}{\partial\sigma_{22}} & \frac{\partial\tau_{12}{}^{+}}{\partial\sigma_{22}}\\ \frac{\partial\sigma_{11}{}^{+}}{\partial\tau_{12}} & \frac{\partial\sigma_{22}{}^{+}}{\partial\tau_{12}} & \frac{\partial\tau_{12}{}^{+}}{\partial\tau_{12}}\end{pmatrix}_{(\theta=\theta_0)}\begin{pmatrix}\sigma_{11}^{+}\\ \sigma_{22}^{+}\\ \sigma_{12}{}^{+}\end{pmatrix}$$

It is easy to calculate the first line:

$$\frac{\partial\sigma_{11}{}^{+}}{\partial\sigma_{11}}_{(\theta=\theta_0)}=h_I+\frac{h_I}{2}\left(1+\cos2\theta_0\right);\ \frac{\partial\sigma_{22}^{+}}{\partial\sigma_{11}}_{(\theta=\theta_0)}=\frac{h_I}{2}\left(1+\cos2\theta_0\right)$$

and

$$\frac{\partial\tau_{12}^{+}}{\partial\sigma_{11}}_{(\theta=\theta_0)}=h_I\sin2\theta_0$$

It appears that

$$\frac{\partial\sigma_{11}^{+}}{\partial\sigma_{11}}\sigma_{11}^{+}+\frac{\partial\sigma_{22}^{+}}{\partial\sigma_{11}}\sigma_{22}{}^{+}+\frac{\partial\tau_{12}{}^{+}}{\partial\sigma_{11}}\sigma_{12}{}^{+}\neq\sigma_{11}{}^{+}\quad\text{except for }\theta_0=0$$

As a conclusion, the relation

$$\frac{\partial\underline{\underline{\sigma}}^{+}}{\partial\underline{\underline{\sigma}}}:\underline{\underline{\sigma}}^{+}=\underline{\underline{\sigma}}^{+}\quad\text{and}\quad\frac{\partial\underline{\underline{\sigma}}^{-}}{\partial\underline{\underline{\sigma}}}:\underline{\underline{\sigma}}^{-}=\underline{\underline{\sigma}}^{-}$$

is not true when the principal axes of the damage tensor and the ones of the stress tensor do not coincide. As a consequence, the calculation of the strain variable from the derivation of the complementary elastic energy is difficult to achieve with the presented model. This is furthermore more difficult to achieve when considering the fourth-order damage operator:

$$\frac{\partial\left[\underline{\underline{\underline{\underline{P}}}}^{+}:\underline{\underline{\underline{\underline{M}}}}:\underline{\underline{\sigma}}\right]}{\partial\underline{\underline{\sigma}}}:\underline{\underline{\tilde{\sigma}}}^{+}\neq\underline{\underline{\tilde{\sigma}}}^{+}\quad\text{and}\quad\frac{\partial\left[\underline{\underline{\underline{\underline{P}}}}^{-}:\underline{\underline{\underline{\underline{M}}}}:\underline{\underline{\sigma}}\right]}{\partial\underline{\underline{\sigma}}}:\underline{\underline{\tilde{\sigma}}}^{-}\neq\underline{\underline{\tilde{\sigma}}}^{-}$$

These remarks concern the thermodynamical background of the developed model. Similar remarks can be formulated for the derivation of the damage strain energy release rate as the projective tensor also depends on the damage tensor (see [6] without introducing this coupling).

References

[1] Mir, H., Fafard, M., Bissonnette, B., and Dano, M. L., 2005, "Damage Modeling in Random Short Glass Fiber Reinforced Composites Including Permanent Strain and Unilateral Effect," ASME J. Appl. Mech., **72**, pp. 249–258.

[2] Ortiz, M., 1985, "A Constitutive Theory for the Inelastic Behavior of Concrete," Mech. Mater., **4**, pp. 67–93.

[3] Truesdell, C., *Introduction à la mécanique rationnelle des milieux continus*, Masson, Paris, 1974.

[4] Challamel, N., Lanos, C., and Casandjian, C., 2005, "Strain-Based Anistropic Damage Modelling and Unilateral Effects," Int. J. Mech. Sci., **47**, pp. 459–473.

[5] Hayakawa, K., and Murakami, S., 1997, "Thermodynamical Modelling of Elastic-Plastic Damage and Experimental Validation of Damage Potential," Int. J. Damage Mech., **6**, pp. 333–363.

[6] Chen, X. F., and Chow, C. L., 1995, "On Damage Strain Energy Release Rate Y," Int. J. Damage Mech., **4**, pp. 251–263.

Closure to "Discussion of 'Damage Modeling in Random Short Glass Fiber Reinforced Composites Including Permanent Strain and Unilateral Effect' " (2006, ASME J. Appl. Mech., 73, pp. 347–348)

Hicham Mir

Mario Fafard

Benoit Bissonnette

Marie-Laure Dano

Department of Civil Engineering,
Universite Laval,
Quebec City, Quebec G1K 7P4 Canada

1 Analysis on the Complementary Elastic Energy

If we compare the first equation of the Discussion to Eq. (2) of the paper, we can remark that the authors of the Discussion have misunderstood the former equation,

$$\frac{1}{2E}\tilde{\sigma}_{ii}^{+}\tilde{\sigma}_{ii}^{+} \neq \frac{1}{2E}\,\mathrm{tr}[(\tilde{\sigma}^{+})^2] \quad \frac{1}{2E}\sigma_{ii}^{-}\sigma_{ii}^{-} \neq \frac{1}{2E}\,\mathrm{tr}[(\tilde{\sigma}^{-})^2]$$

Considering this, their remark is irrelevant. Nevertheless, we can observe easily that in the particular case $[D]=0$, Eq. (2) is simplified to Eq. (1) of the paper,

$$U^{e}(\sigma,[D]=0) = \frac{1}{2E}\left(\sigma_{ii}^{+}\sigma_{ii}^{+} + \sigma_{ii}^{-}\sigma_{ii}^{-} + \sigma_{ij}\sigma_{ij}\big|_{i\neq j}\right) - \frac{\upsilon}{E}(\sigma_{ii}\sigma_{jj} - \sigma_{ij}\sigma_{ij})$$

$$= \frac{1}{2E}[\sigma]:[\sigma] - \frac{\upsilon}{E}((\mathrm{tr}[\sigma])^2 - \mathrm{tr}([\sigma]^2))$$

Furthermore, we can observe that if we put $[D]=0$, the fourth-order damage operator defined in Eq. (6) becomes $[M]=[I_4]$. Thus, in this particular case, Eq. (9) is exactly equal to the classical complementary energy presented at Eq. (1).

2 On the Projection Derivative Operator

The authors of the Discussion mentioned that "when the principal axes of the damage tensor do not coincide with the one of the stress tensor," some relations presented in the paper are not applicable. We agree with this affirmation because we mentioned it in Sec. 2.3 of the paper and in the Conclusion. When we spoke about proportional loading, this means that the principal stress directions, and thus, those of damage tensor, are identical and do not change during loading. We apologize not having mentioned it more clearly at the beginning of the paper.

