Transactions of the ASME®

Transactions of the ASME, Journal of Applied Mechanics (ISSN 0021-8936) is published bimonthly (Jan., Mar., May, July, Sept., Nov.) The American Society of Mechanical Engineers, Three Park Avenue, New York, NY 10016. Periodicals postage paid at New York, NY and additional mailing office. POSTMASTER: Send address changes to Transactions of the ASME, Journal of Applied Mechanics, c/o THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 22 Law Drive, Box 2300, Fairfield, NJ 07007-2300. **CHANGES OF ADDRESS** must be received at Society headquarters seven weeks before they are to be effective. Please send old label and new address. **STATEMENT from By-Laws.** The Society shall not be responsible for statements or opinions advanced in papers or . . . printed in its publications (B7.1, Para. 3).
 INDEXED by Applied Mechanics Reviews and Engineering Information, Inc. Canadian Goods & Services Tax Registration #126148048.

Journal of Applied Mechanics

Published Quarterly by The American Society of Mechanical Engineers

VOLUME 68 • NUMBER 1 • JANUARY 2001

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 68, Number 1 JANUARY 2001

BRIEF NOTES

BOOK REVIEW

ANNOUNCEMENTS AND SPECIAL NOTES

E. Turan Onat
1925–2000

The mechanics community lost one of its distinguished members with the passing away of E. Turan Onat in New Haven, Connecticut, on July 4, 2000 due to an inoperable cancer.

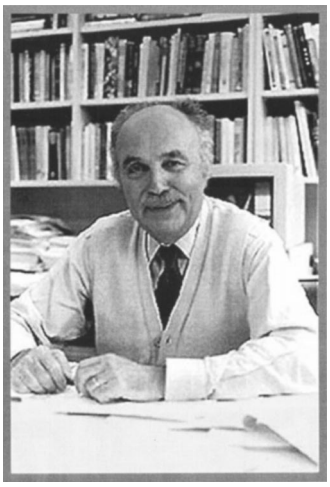
Onat was a native of Turkey and came from a prominent family. His father was a general in the judicial branch of the army, and his brother and sister live in Istanbul. He received the degrees of Dipl. Ing. in 1948 and the Doctor of Science in 1951 from the Istanbul Technical University. As part of his doctoral studies, he served as a Research Fellow at the Paris Academy in 1950.

He came to the U.S. in 1951 as a Post-Doctoral Fellow at Brown University. There he engaged in the research on plasticity that was in full development at Brown during the 1950's. His research in that period is characterized by the rigorous treatment of problems that clarify and illustrate the underlying principles of plasticity theory. These features are exemplified in work with R. T. Shield on combined bending and twisting of tubes in the plastic range; with W. Prager on limit analysis of arches, on plane strain necking in tension, and collapse and limit load analysis of various structures; and with D. C. Drucker on the stability of inelastic systems.

In 1954, Onat returned to Turkey for his compulsory military duty and served as a lieutenant attached to the scientific advisory board of the Turkish general staff. He came back to Brown University in 1957 as an Associate Professor of Engineering and was promoted to full Professor in 1960. He was awarded a Guggenheim Fellowship for the academic year 1963–1964, during which he served as Visiting Professor at Cambridge University and at the Istanbul Technical University.

In 1965, he joined the engineering faculty of Yale University, where he remained. At Yale, he participated fully in the intellectual, artistic, and social life of the university. He developed teaching and research collaborations with Yale colleagues in the departments of geology, mathematics, and the medical school. He was a Fellow of Ezra Stiles College at Yale from 1965 to 1990. Together with his wife, Etta, who taught in the English Department and served as an Associate Dean of Humanities in Yale's Graduate School, he served as interim resident Master of Ezra Stiles College in 1990, and later, from 1990–1995, as Master of Calhoun College. In these roles, Turan and Etta looked after the personal needs of several hundred undergraduates in the residential college. They were frequently in attendance at the numerous student musical and theatrical performances at Yale, and at gallery exhibitions. He was loved by the students, and during one of my visits, we were serenaded by a group of young lady students at the university restaurant.

Onat had an abiding interest in structures and structural mechanics. Research at Brown dealt with critical aspects of the plastic behavior of beam, frame, arch, plate, shell, and membrane structures, including fundamentals of plastic stability (with D. C. Drucker), the influence of large displacements and resulting geometry change on continuing fully plastic deformation (with R. M. Haythornthwaite), and effects of coupled axial and shear loading on bending response (with W. Prager; R. T. Shield). R. H. Lance collaborated with him in plastic analysis of conical shells. Along with his technical interests in structures, his artistic and aesthetic appreciation of architecture led Onat to develop highly popular structures and structural mechanics courses. He was especially fond of his collection of model structures designed and assembled by students in these courses.



Two studies of plastic necking (with Prager) and buckling (with G. R. Cowper) were particularly important applications of plastic stability and continuing limit-load behavior. The necking solution with Prager featured a pair of centrally intersecting slip-lines separating four rigid blocks in relative motion. A physical realization of the idealization can be found in tension of single crystals oriented for symmetric double slip. R. J. Asaro once showed Onat an aluminum alloy crystal that had been deformed in his manner, and the crystal contained a small-diameter hole where the shear-bands crossed. Delighted with the correspondence of the crystal with his solution, Onat (unsuccessfully) asked Asaro to give him the specimen so he could make a tie-clasp of it.

In the early 1960s, Onat also investigated plastic wave propagation in membranes due to transverse impact, and the creep of metals subjected to incremental changes in loading. He was also concerned with fundamental issues in linear viscoelasticity, such as uniqueness, and published papers with S. Breuer on the subject.

One of the major scientific contributions of Onat was his rigorous treatment of the role of internal state variables in constitutive theories. During the mid 1960s, he advocated the now nearly universally adopted constitutive formalism based on the evolution of internal state variables, in distinction to a competing hereditary integral equation constitutive formalism inherited from linear viscoelasticity. He emphasized that the more nonlinear the behavior, the more terms (kernels) were required in the hereditary integral representation, while systems of first-order nonlinear differential equations representing evolution of state parameters offered great flexibility within a unifying framework. His work subsequently focused on unified constitutive equations for creep and plasticity which essentially depend on variables whose current values encompass all previous loading histories. Particular internal state variables used in such equations are intended to represent the various hardening effects, and continuum damage which can lead to tertiary creep. His studies on the identification and properties of those variables have served to guide the formulation of appropriate constitutive theories for elastic-viscoplastic media.

Onat's later work was directed to developing internal variable approaches to large deformation viscoplastic-

ity, and on approaches based on group theory for the representation of tensorial symmetries present in the internal state variables. Major collaborators in these endeavors included F. A. Leckie (creep damage), J. P. Boehler (elastic symmetries), E. H. Lee (large deformation plasticity), and B. L. Adams (representation of polycrystal microstructure).

Onat was much appreciated by the mechanics community for his special personal qualities, as well as for his scientific influence. His charming personality and gracious manner endeared him to his colleagues and associates, some of whom became close friends. He was an outdoorsman and, in Providence, shared a sailboat (''Tresca'') with Dick Shield to sail on Narragansett Bay, with myself as an occasional passenger. His Boston Whaler helped him catch many bluefish in Long Island Sound. At Yale, Onat obtained his pilot's license for a light plane, and became a proficient and serious pilot. He enjoyed drawing and was almost never without his sketchbook. Other avocations included gardening and long walks in the woods close to his home in Woodbridge, near New Haven.

As part of his scholarly and cultural interests, Onat traveled to a number of places and served as Visiting Professor in Cambridge and Oxford Universities, and worked with collaborators at the National Polytechnic Institute of Grenoble, the University of Illinois at Urbana-Champaign, the Istanbul Technical University where he received a special honor in 1999, Ruhr-Bochum University in Germany, Tianjin University in China, and the University of Rio Grande do Sul in Porto Alegre, Brazil. He delivered invited lectures in India, Israel, Italy, Poland, Spain, and Switzerland. He won the Senior Scientist Award of the Alexander von Humbolt Foundation in 1989. He was a member of ASME, the American Mathematical Society, the Society for Natural Philosophy, Sigma Xi, and the American Academy of Mechanics, of which he was elected Fellow in 1980.

Onat's final months were difficult, both for him and for those near him, but he retained his smile, his good humor, and his love of life and the life of the mind. He will be missed.

He is survived by his wife, Etta Onat, of Woodbridge, Connecticut, his son, Yasar Onat, of New Haven; his daughter-in-law, Wendy Natter and his granddaughter, Rebecca Onat, of Guilford, Connecticut, his brother, Dogan Onat, and his sister, Meral Aras, both of Istanbul.

Sol R. Bodner
Technion-Israel Institute of Technology, Haifa
with the collaboration of
David Parks
MIT, Cambridge, Massachusetts

Elastic Fields in Double Inhomogeneity by the Equivalent Inclusion Method

H. M. Shodja
Assoc. Mem. ASME,
Shodja@sina.sharif.ac.ir

A. S. Sarvestani

Department of Civil Engineering,
Sharif University of Technology,
Tehran, Iran

Consider a double-inhomogeneity system whose microstructural configuration is composed of an ellipsoidal inhomogeneity of arbitrary elastic constants, size, and orientation encapsulated in another ellipsoidal inhomogeneity, which in turn is surrounded by an infinite medium. Each of these three constituents in general possesses elastic constants different from one another. The double-inhomogeneity system under consideration is subjected to far-field strain (stress). Using the equivalent inclusion method (EIM), the double inhomogeneity is replaced by an equivalent double-inclusion (EDI) problem with proper polynomial eigenstrains. The double inclusion is subsequently broken down to single-inclusion problems by means of superposition. The present theory is the first to obtain the actual distribution rather than the averages of the field quantities over the double inhomogeneity using Eshelby's EIM. The present method is precise and is valid for thin as well as thick layers of coatings, and accommodates eccentric heterogeneity of arbitrary size and orientation. To establish the accuracy and robustness of the present method and for the sake of comparison, results on some of the previously reported problems, which are special cases encompassed by the present theory, will be re-examined. The formulations are easily extended to treat multi-inhomogeneity cases, where an inhomogeneity is surrounded by many layers of coatings. Employing an averaging scheme to the present theory, the average consistency conditions reported by Hori and Nemat-Nasser for the evaluation of average strains and stresses are recovered. [DOI: 10.1115/1.1346680]

1 Introduction

Often, during the processing of composites due to chemical interactions, an undesirable phase called interphase forms between the fiber and the matrix. In other situations, coating technology is employed to reduce the large residual tensile stresses between the fiber and the matrix, and consequently to prevent matrix cracking in cool-down processes. Also, coating is used to improve electrical conductivity of composites, or to provide a protective layer against aggressive corrosive agents. Existence of high stress concentration just outside a coated/uncoated fiber is a well-known phenomenon. The overall behavior of composite materials is greatly altered by their microconstituents such as interface layers, as well as geometry and distribution of phases. It should be emphasized that even though coatings are usually very thin, they play an important role in controlling the failure mechanisms and fracture toughness of a material.

In the present study inclusions and inhomogeneities are differentiated as follows: An inclusion is a finite domain with eigenstrain ϵ^*_{ij} whose elastic moduli are the same as those of its surrounding matrix, whereas an inhomogeneity is a finite domain whose elastic moduli differ from those of its surrounding matrix. The problems of single and multi-inhomogeneity of arbitrary elastic constants surrounded by a matrix of different elastic properties have been addressed by several investigators. A brief summary of some relevant theoretical treatments on the subject are presented in this section. A more complete and fairly updated review is given by Mura et al. [1]. One of the well-known and fundamental theories is due to Eshelby's [2–4] for a single-ellipsoidal inclusion. Eshelby first showed that when an isotropic infinite domain surrounds an ellipsoidal inclusion Ω with uniform eigenstrain ϵ^*_{ij}, then the strain ϵ_{ij} is also uniform in Ω and is expressed by

$$\epsilon_{ij}=S_{ijkl}\epsilon^*_{kl}, \quad \mathbf{x}\in\Omega, \tag{1}$$

where S_{ijkl} is the Eshelby's tensor which is uniform,

$$S_{ijkl}=\int_\Omega \Gamma_{ijkl}(\mathbf{x}-\mathbf{x}')d\mathbf{x}', \tag{2a}$$

$$\Gamma_{ijkl}(\mathbf{x}-\mathbf{x}')=-\tfrac{1}{2}C_{mnkl}[G_{im,nj}(\mathbf{x}-\mathbf{x}')+G_{jm,ni}(\mathbf{x}-\mathbf{x}')], \tag{2b}$$

where $G_{ij}(\mathbf{x}-\mathbf{x}')$ is the fundamental solution satisfying the Green's functions problem

$$C_{ijkl}G_{kp,lj}(\mathbf{x}-\mathbf{x}')+\delta_{ip}\delta(\mathbf{x}-\mathbf{x}')=\mathbf{0}, \tag{3}$$

which is associated with the equilibrium equations. $\delta(\mathbf{x}-\mathbf{x}')$ is Dirac's delta function and δ_{ip} is the Kronecker delta. Then he extended his treatment to the case of an ellipsoidal inhomogeneity by reducing it to an inclusion problem with a proper choice of eigenstrains. These eigenstrains and consequently the stress and strain fields inside the inhomogeneity are uniform if the far-field applied stress (strain) is uniform. This treatment, which is known as ''the equivalent inclusion method'' (EIM), is valid only for a single inhomogeneity. If the body contains two or more inhomogeneities which are interacting, then the method must be modified. Moschovidis and Mura [5] have extended EIM to the case of two interacting ellipsoidal inhomogeneities that are nonintersecting and occupy different points in space.

To date, unlike single-inclusion and single-inhomogeneity problems which have been studied extensively, not much attention has been given to the determination of the local elastic fields in the case of the double-inhomogeneity problem. This problem occurs when a layer of different elastic properties is added between an inhomogeneity and its surrounding matrix. In general, for such a situation the stress and strain fields inside the coating and the core inhomogeneity will no longer be uniform, even if the far-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Oct. 7, 1999; final revision, June 14, 2000. Associate Editor: D. Kouris. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

field loading is uniform; see, for example, Chen et al. [6]. Consequently, it may lead one to believe that Eshelby's method would not be applicable to such problems. As a result, researchers have employed approximations and other complicated approaches. Some of these methods are given by Walpole [7], Mikata and Taya [8], and Benveniste et al. [9]. It is interesting to note that for a double-inhomogeneity system under uniform far-field loading, there are situations when the stress and strain fields become uniform in the core inhomogeneity only. For instance, it can easily be verified from the treatments given by Christensen and Lo [10] that for isotropic three-phase spherically and cylindrically concentric solids subjected to uniform far-field loading, the stresses are uniform in the core inhomogeneity. This result can also be deduced from the paper of Chen et al. [6] on three-phase cylindrically concentric inhomogeneities, when the core inhomogeneity is transversely isotropic. On the other hand, it can be observed from the results obtained by Chen et al. that when the core inhomogeneity is not transversely isotropic, the stresses cease to be uniform there.

Walpole [7] considers an inhomogeneity with a very thin coating in an infinite domain under mechanical loading. To handle the complexities of the interface layer he assumes thin enough coating such that its existence has very little bearing on the elastic fields inside the central particle. With this simplifying assumption he makes use of Hill's theorem ([11]) to find the stress and strain components at the interface. He points out that even for thin coating, his analysis looses validity when coating is excessively soft or rigid. Mikata and Taya [8] only consider the axisymmetric problem of the short coated fiber in an infinite domain subjected to axisymmetric loading conditions. They model a short coated fiber as two confocal spheroids. They claim that Eshelby's method cannot be employed in this problem and use Boussinesq-Sadowsky stress functions in their analysis to find the stress field in and around a coated fiber. Benveniste et al. [9], on the basis of Benveniste's [12] re-examination of Mori-Tanaka's theory, presented a micromechanical model and computed approximate stress fields and overall thermomechanical properties of composites with coated inhomogeneities. To date, no general and exact solution of the double-inhomogeneity problem pertaining calculation of elastic fields inside the inhomogeneity and its coating has been given. However, an analytical treatment for the calculation of average field quantities within these domains has been given by Hori and Nemat-Nasser [13] and Nemat-Nasser and Hori [14]. They use Eshelby's method and present a more general theory than the one due to Tanaka and Mori's theorem ([15]) namely ''the double-inclusion model.'' In their treatment, they generalize the double-inclusion to a multi-inclusion model and show that the average elastic fields over a set of nested ellipsoidal regions consisting of innermost inhomogeneity and its surrounding layers of coatings can be computed exactly, provided that the prescribed eigenstrains are uniform but different within each annulus, and not necessarily uniform in the core.

In Section 2 of this paper, following Eshelby's EIM and the method of Moschovidis and Mura [5] and Mura [16], a general treatment of double inhomogeneity is presented. In Section 3, the method is extended to multi-inhomogeneity system, where an ellipsoidal inhomogeneity is surrounded by many layers of coatings of ellipsoidal shape. In this analysis the core and its coatings can have arbitrary elastic properties, orientations, positions, and aspect ratios. In Section 4 some numerical examples are presented to verify the accuracy and robustness of the method. For this purpose, some relevant problems considered by other investigators will be reexamined by the present method. In Appendix B, it is shown that the average field quantities obtained over each annulus and core of the multi-inhomogeneity configuration will be exactly the same as those obtained by Hori and Nemat-Nasser [13].

2 Formulation of the Double-Inhomogeneity Problem

Consider a solid consisting of a double-inhomogeneity $\Sigma=\Psi\cup\Omega$ embedded in an isotropic infinite medium Φ, where Ψ and Ω have ellipsoidal shapes and arbitrary elastic properties, orientations, positions, and aspect ratios, as depicted in Fig. 1. Note that Ψ and Ω may be either isotropic or anisotropic materials. The elastic moduli of the core Ω, annulus $\Psi=\Sigma-\Omega$, and matrix $\Phi-\Sigma$ are denoted by $\mathbf{C}^1$, $\mathbf{C}^2$, and $\mathbf{C}$, respectively. The medium is under uniform far-field stress, σ^0_{ij} (strain, ϵ^0_{ij}) so that

$$\sigma^0_{ij}=C_{ijkl}\epsilon^0_{kl}. \tag{4}$$

Unlike Eshelby's single-inhomogeneity problem, in the double-inhomogeneity case, the stress and strain fields in Ω will no longer be uniform, when subjected to constant far-field stress (strain) due to the presence of region Ψ and its interaction with the inhomogeneity Ω.

Let $\mathbf{x}$ and $\bar{\mathbf{x}}$ denote the Cartesian coordinates with origins o and $\bar{o}$ located at the center of ellipsoids Σ and Ω, respectively. Moreover, the coordinate axes are taken to coincide with the principal axes of the ellipsoids, as shown in Fig. 1. The double inhomogeneity is replaced with the equivalent double inclusion (EDI) shown in Fig. 2. This equivalency holds for proper choice of homogenizing eigenstrains $\bar{\epsilon}^{*(1)}_{ij}(\bar{\mathbf{x}})$ and $\epsilon^{*(2)}_{ij}(\mathbf{x})$ defined over the regions Ω and Ψ, respectively. As mentioned earlier, the eigenstrains $\bar{\epsilon}^{*(1)}_{ij}(\bar{\mathbf{x}})$, and $\epsilon^{*(2)}_{ij}(\mathbf{x})$ are not uniform, and indeed can be expanded in terms of space variables

$$\bar{\epsilon}^{*(1)}_{ij}(\bar{\mathbf{x}})=\alpha^*_{ij}+\alpha^*_{ijk}\bar{x}_k+\alpha^*_{ijkl}\bar{x}_k\bar{x}_l+\ \ldots\ , \quad \bar{\mathbf{x}}\in\Omega, \tag{5a}$$

$$\epsilon^{*(2)}_{ij}(\mathbf{x})=\beta^*_{ij}+\beta^*_{ijk}x_k+\beta^*_{ijkl}x_kx_l+\ \ldots\ , \quad \mathbf{x}\in\Psi, \tag{5b}$$

where $\bar{\epsilon}^{*(1)}_{ij}(\bar{\mathbf{x}})$ and $\epsilon^{*(2)}_{ij}(\mathbf{x})$ are continuous, differentiable, and symmetric with respect to indices i and j. Thus the constants α^*_{ij}, α^*_{ijk}, ... are symmetric with respect to the free indices and also $\alpha^*_{ijkl}=\alpha^*_{ijlk}$, $\alpha^*_{ijklm}=\alpha^*_{ijkml}$, etc. Similar symmetries hold for $\boldsymbol{\beta}^*$s. The tensors referring to the $\bar{\mathbf{x}}$-coordinate are indicated by bars. The components of a Cartesian tensor $\mathbf{A}$ of order n associated with the $\mathbf{x}$-coordinate are transformed to the components of the tensor $\mathbf{A}$ according to the following relation:

$$\bar{A}_{ij\ldots k}=q_{il}q_{jm}\ldots q_{kn}A_{lm\ldots n}, \tag{6}$$

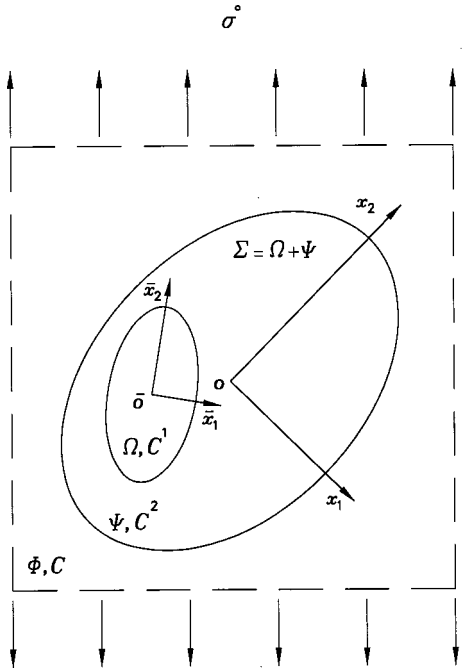


Fig. 1 Double inhomogeneity $\Sigma=\Psi\cup\Omega$ embedded in an infinite medium Φ. Σ and Ω have arbitrary orientations, and C^1, C^2 and C are distinct

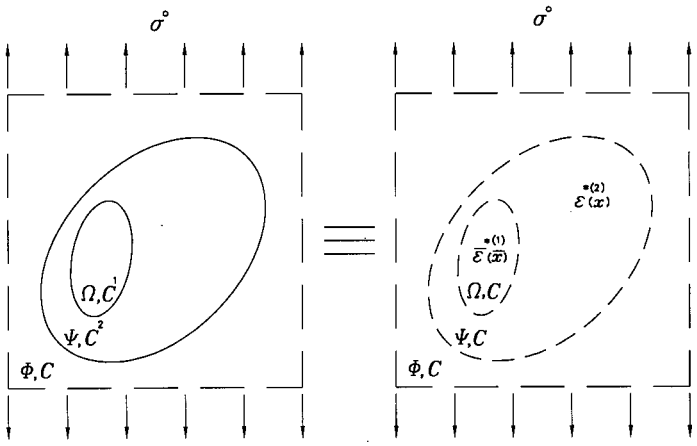


Fig. 2 Double inhomogeneity is replaced by an EDI with proper homogenizing polynomial eigenstrains

where q_{ij} is the transformation tensor. The eigenstrains give rise to a disturbance in the strain fields over the regions Ω and Ψ denoted by $\bar{\epsilon}_{ij}^{d(1)}(\bar{\mathbf{x}})$ and $\epsilon_{ij}^{d(2)}(\mathbf{x})$, respectively, which can be expanded as follows:

$$\bar{\epsilon}_{ij}^{d(1)}(\bar{\mathbf{x}}) = \xi_{ij}^{d} + \xi_{ijk}^{d}\bar{x}_k + \xi_{ijkl}^{d}\bar{x}_k\bar{x}_l + \ldots, \quad \bar{\mathbf{x}} \in \Omega, \tag{7a}$$

$$\epsilon_{ij}^{d(2)}(\mathbf{x}) = \eta_{ij}^{d} + \eta_{ijk}^{d}x_k + \eta_{ijkl}^{d}x_k x_l + \ldots, \quad \mathbf{x} \in \Psi, \tag{7b}$$

where $\boldsymbol{\xi}^d$s and $\boldsymbol{\eta}^d$s have symmetry properties similar to those of $\acute{\mathbf{a}}^*$s and $\boldsymbol{\beta}^*$s. It should be emphasized that the interaction between the inhomogeneities Ω and Ψ is inherent in the disturbance strains $\bar{\epsilon}_{ij}^{d(1)}(\bar{\mathbf{x}})$ and $\epsilon_{ij}^{d(2)}(\mathbf{x})$.

Employing the EIM, the following consistency conditions are obtained:

$$\bar{C}_{ijkl}^{1}(\bar{\epsilon}_{kl}^{0} + \bar{\epsilon}_{kl}^{d(1)}(\bar{\mathbf{x}})) = C_{ijkl}(\bar{\epsilon}_{kl}^{0} + \bar{\epsilon}_{kl}^{d(1)}(\bar{\mathbf{x}}) - \bar{\epsilon}_{kl}^{*(1)}(\bar{\mathbf{x}})), \quad \bar{\mathbf{x}} \in \Omega, \tag{8a}$$

$$C_{ijkl}^{2}(\epsilon_{kl}^{0} + \epsilon_{kl}^{d(2)}(\mathbf{x})) = C_{ijkl}(\epsilon_{kl}^{0} + \epsilon_{kl}^{d(2)}(\mathbf{x}) - \epsilon_{kl}^{*(2)}(\mathbf{x})), \quad \mathbf{x} \in \Psi. \tag{8b}$$

Suppose that the polynomials given by expressions (5) and (7) for the eigenstrains and the disturbance strains are of degree m. Let the corresponding number of coefficients in each expression be equal to m'. Equating the coefficients of like powers on both sides of Eqs. (8) result in $12m'$ equations with $24m'$ unknown coefficients, $\boldsymbol{\alpha}^*$s, $\boldsymbol{\beta}^*$s, $\boldsymbol{\xi}^d$s, and $\boldsymbol{\eta}^d$s. In the remainder of this section, the remedy to obtain the necessary additional equations is outlined.

With the aid of superposition we decompose the EDI shown in Fig. 2 to the algebraic sum of an infinite domain Φ under uniform far-field stress, σ_{ij}^0, an inclusion Σ with eigenstrain $\epsilon_{ij}^{*(2)}(\mathbf{x})$ surrounded by the infinite medium $\Phi - \Sigma$, and an inclusion Ω surrounded by the infinite medium $\Phi - \Omega$ with eigenstrains $\epsilon_{ij}^{*(1)}(\mathbf{x})$ and $\epsilon_{ij}^{*(2)}(\mathbf{x})$, respectively. The schematic representation of such a superposition is illustrated in Fig. 3.

According to the superposition shown in Fig. 3, the disturbances in strain fields $\bar{\epsilon}_{ij}^{d(1)}(\bar{\mathbf{x}})$ and $\epsilon_{ij}^{d(2)}(\mathbf{x})$ are given by

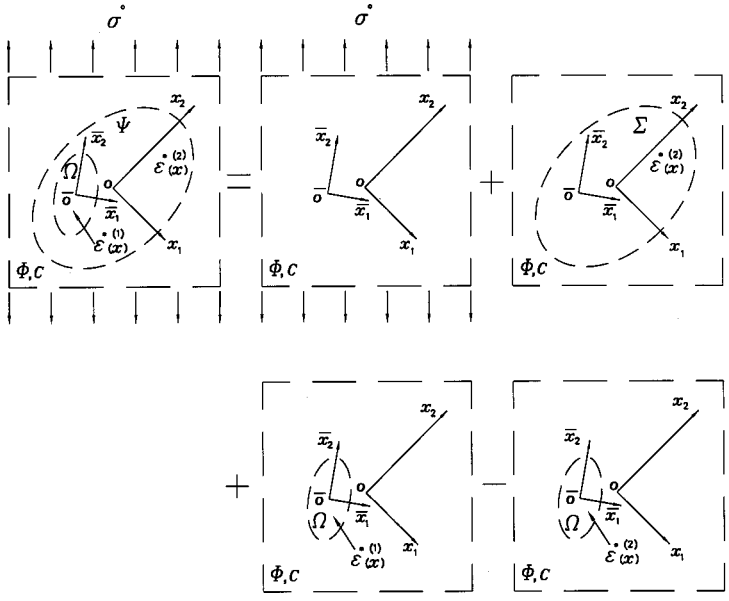


Fig. 3 Decomposition of the EDI problem to a domain under uniform far-field stress and three single-inclusion problems with proper polynomial eigenstrains

$$\bar{\epsilon}_{ij}^{d(1)}(\bar{\mathbf{x}}) = \bar{\epsilon}_{ij}^{\Sigma}(\bar{\mathbf{x}};\bar{\boldsymbol{\epsilon}}^{*(2)}) + \bar{\epsilon}_{ij}^{\Omega}(\bar{\mathbf{x}};\bar{\boldsymbol{\epsilon}}^{*(1)}) - \bar{\epsilon}_{ij}^{\Omega}(\bar{\mathbf{x}};\bar{\boldsymbol{\epsilon}}^{*(2)}), \quad \bar{\mathbf{x}} \in \Omega, \tag{9a}$$

$$\epsilon_{ij}^{d(2)}(\mathbf{x}) = \epsilon_{ij}^{\Sigma}(\mathbf{x};\boldsymbol{\epsilon}^{*(2)}) + \epsilon_{ij}^{\Omega}(\mathbf{x};\boldsymbol{\epsilon}^{*(1)}) - \epsilon_{ij}^{\Omega}(\mathbf{x};\boldsymbol{\epsilon}^{*(2)}), \quad \mathbf{x} \in \Psi, \tag{9b}$$

where $\epsilon_{ij}^{\alpha}(\mathbf{x};\boldsymbol{\epsilon}^{*(i)})$ is the disturbance strain in region α (α being domains Σ or Ω) at point $\mathbf{x}$ due to the eigenstrain $\boldsymbol{\epsilon}^{*(i)}$ (i being 1 or 2 correspond to the homogenizing eigenstrain defined over the regions Ω or Ψ, respectively). The disturbance strain $\epsilon_{ij}^{\alpha}(\mathbf{x};\boldsymbol{\epsilon}^{*(i)})$ is given by

$$\epsilon_{ij}^{\alpha}(\mathbf{x};\boldsymbol{\epsilon}^{*(i)}) = \int_{\alpha} \Gamma_{ijmn}(\mathbf{x}-\mathbf{x}')\epsilon_{mn}^{*(i)}(\mathbf{x}')d\mathbf{x}', \tag{10}$$

where $\Gamma_{ijmn}(\mathbf{x}-\mathbf{x}')$ is defined by the expression (2*b*). When α is an inclusion in an isotropic infinite body, where the eigenstrain field in α is in the form of polynomials of coordinates

$$\epsilon_{ij}^{*}(\mathbf{x}) = \zeta_{ij}^{*} + \zeta_{ijk}^{*}x_k + \zeta_{ijkl}^{*}x_k x_l + \ldots, \tag{11}$$

then the disturbance in strain field is given by

$$\epsilon_{ij}^{\alpha}(\mathbf{x};\boldsymbol{\epsilon}^{*(i)}) = \gamma_{ijkl}(\mathbf{x})\zeta_{kl}^{*} + \gamma_{ijklq}(\mathbf{x})\zeta_{klq}^{*} + \gamma_{ijklqr}(\mathbf{x})\zeta_{klqr}^{*} + \ldots, \tag{12}$$

where the tensors $\gamma_{ijkl\ldots}$ are Eshelby's tensors ([4]). We select two distinct points $\bar{\mathbf{x}} \in \Omega$ and $\mathbf{x} \in \Psi$ at which the evaluation of elastic fields are desirable. It should be emphasized that $\bar{\mathbf{x}}$ is an interior point of domains Ω and Ψ but $\mathbf{x}$, which is an interior point of domain Ψ, is an exterior point with respect to region Ω. Eshelby showed that, if $\mathbf{x}$ is an interior point of the ellipsoidal inclusion α, and the eigenstrain in the inclusion is an mth order polynomial of the form given by the expression (11), then according to (12) $\epsilon_{ij}^{\alpha}(\mathbf{x};\boldsymbol{\epsilon}^{*(i)})$ will be an inhomogeneous polynomial in $\mathbf{x}$ whose terms are of degree $m,(m-2),(m-4),\ldots$ (see Appendix A). Asaro and Barnett [17] obtained similar results for a single anisotropic ellipsoidal inclusion. On the other hand, if $\mathbf{x}$ is an exterior point of the inclusion α, then Taylor's expansion of the tensors $\gamma_{ijkl\ldots}$ about point $\mathbf{x}$ is used (see Appendix A). In view of the above discussions, an appropriate substitution for $\gamma_{ijkl\ldots}$ into (12) leads to a new expression for $\epsilon_{ij}^{\alpha}(\mathbf{x};\boldsymbol{\epsilon}^{*(i)})$ in terms of polynomials of coordinates. Consequently $\bar{\epsilon}_{ij}^{d(1)}(\bar{\mathbf{x}})$ and $\epsilon_{ij}^{d(2)}(\mathbf{x})$ in Eqs. (9) can be expressed in terms of polynomials of degree m. After conversion of Eqs. (9) to new sets of expansions for $\bar{\boldsymbol{\epsilon}}^{d(1)}$ and $\boldsymbol{\epsilon}^{d(2)}$ in the manner explained above, the coefficients of like powers from these new expansions are equated to those from Eqs. (7) to yield $12m'$ additional equations, hence completing the necessary systems of equations.

3 Extension to Multi-Inhomogeneity Systems

The theory presented for double inhomogeneity in Section 2 is easily extended to the multi-inhomogeneity problem in which the interphase material surrounding the core inhomogeneity is a multilayer. The core and the layers are of ellipsoidal shape with arbitrary elastic properties, orientations, positions, and aspect ratios. The schematic representation of such a system, consisting of the core and n layers of coatings, is illustrated in Fig. 4. The ellipsoids are ordered as $\Sigma_1 \equiv \Psi_1 \subset \Sigma_2 \subset \Sigma_3 \subset \ldots \subset \Sigma_{n+1}$ such that for the layer i indicated by Ψ_i we have

$$\Psi_i = \Sigma_i - \Sigma_{i-1}, \quad i = 2,3,\ldots,n+1, \tag{13}$$

where the core Σ_1 and each region Ψ_i have arbitrary elastic constants C^1 and C^i, $i=2,3,\ldots,n+1$, respectively. The multi-inhomogeneity is surrounded by the matrix $\Phi - \Sigma_{n+1}$ with elastic moduli C. To formulate this problem, the multi-inhomogeneity system is replaced with the equivalent multi-inclusion (EMI) with proper choice of homogenizing eigenstrains $\epsilon_{kl}^{*(i)}$, i

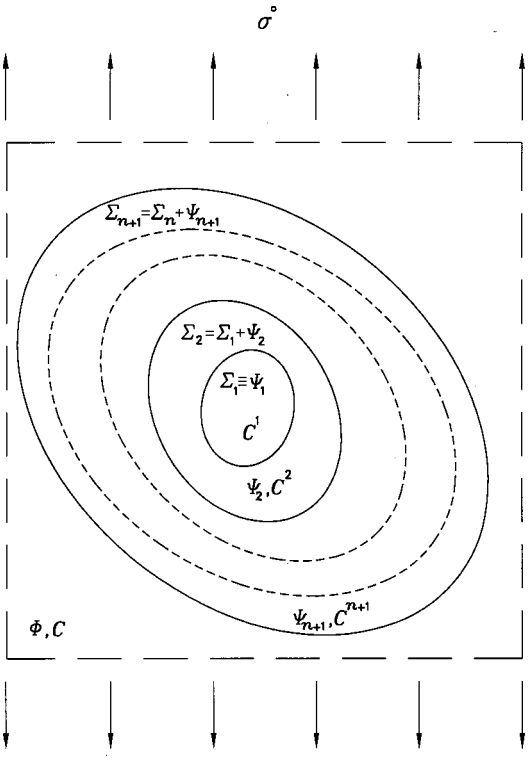


Fig. 4 A multi-inhomogeneity system consisting of *n*-layers of coatings

$=1,2,3,\ldots,n+1$. Next, the superposition described in the previous section is employed to the resulting EMI. Thus, referring to the $\mathbf{x}$-coordinate

$$C_{ijpq}^{k}(\epsilon_{pq}^{0} + \epsilon_{pq}^{d(k)}(\mathbf{x})) = C_{ijpq}(\epsilon_{pq}^{0} + \epsilon_{pq}^{d(k)}(\mathbf{x}) - \epsilon_{pq}^{*(k)}(\mathbf{x})), \quad \mathbf{x} \in \Psi_k, \quad k = 1,2,\ldots,n+1, \tag{14a}$$

$$\epsilon_{ij}^{d(k)}(\mathbf{x}) = \sum_{l=1}^{n+1} \epsilon_{ij}^{\Sigma_l}(\mathbf{x};\boldsymbol{\epsilon}^{*(l)}(\mathbf{x})) - \sum_{l=1}^{n} \epsilon_{ij}^{\Sigma_l}(\mathbf{x};\boldsymbol{\epsilon}^{*(l+1)}(\mathbf{x})), \quad \mathbf{x} \in \Psi_k, \quad k = 1,2,\ldots,n+1. \tag{14b}$$

Suppose that the eigenstrains and the disturbance strains are mth order polynomials of coordinates, each polynomial having m' terms. Following similar arguments as for the EDI case discussed in Section 2, the consistency Eqs. (14*a*) yield $6m'(n+1)$ equations with $12m'(n+1)$ unknowns. Moreover, depending on whether point $\mathbf{x}$ is an interior point of subdomain Σ_i or not, each quantity $\boldsymbol{\epsilon}^{\Sigma_i}(\mathbf{x};\boldsymbol{\epsilon}^{*(i)})$ can be expressed by an inhomogeneous polynomial or Taylors' series expansion of degree m, respectively. This leads to an additional set of $6m'(n+1)$ equations, which then completes the system of equations for the determination of $12m'(n+1)$ unknowns. Using the theory presented herein, the stress (strain) can be calculated pointwise with high accuracy within the inhomogeneities. In Appendix B, it is shown that by employing an averaging scheme to the present theory, the average consistency conditions obtained by Hori and Nemat-Nasser [13] for the evaluation of average strains and stresses are recovered.

4 Results and Discussion

In this section the numerical solutions for the stress field of three different double-inhomogeneity problems will be given. For the sake of comparison and in order to demonstrate the accuracy of the present theory, two problems considered by other investi-

gators are reexamined. In Subsection 4.1, the coated short fiber composites studied by Mikata and Taya [8] and in Subsection 4.2 the coated continuous fiber studied by Mikata and Taya [18], and by Benveniste et al. [12] is reconsidered. In Subsection 4.3 a double-inhomogeneity system is considered, consisting of a spherical inhomogeneity with a spherical cavity and centers that do not coincide. Hence, this system does not possess any of the symmetries of the problems of Subsections 4.1 and 4.2. Furthermore, the interphase layer has variable thickness due to the existing eccentricity.

4.1 Coated Short Fiber Composites Under Mechanical Loading. As mentioned earlier, Mikata and Taya [8] used the Boussinesq-Sadowsky stress function to calculate the stress field around a coated short fiber under axisymmetric loading. This problem is schematically shown in Fig. 5. It is solved under uniaxial far-field stress σ_z^0, fiber aspect ratio $b/a=20$, coating to matrix stiffness ratio $G_2/G_1=2$, fiber-to-matrix stiffness ratio $G_3/G_1=10$, and Poisson's ratio $\nu_1=\nu_2=\nu_3=0.3$. For coating thickness to fiber radius ratio of $c/a=0.1$, they found the normalized stress σ_z/σ_z^0 along the radial direction on the plane $z=0$ in fiber, coating and matrix to be 8.8, 1.8, and 0.9 respectively. Their results together with the ones obtained by the present theory are shown in Fig. 6. According to the present theory $\sigma_z/\sigma_z^0=9.25,1.85$ in fiber and coating, respectively. Inside the matrix, just outside the coating, $\sigma_z/\sigma_z^0=0.93$ which gradually approaches $\sigma_z/\sigma_z^0=1$ away from the coating. In this particular example, since the coating is thin, the stress fields in the fiber and coating are nearly constant and hence the zeroth-order term in the expansions of eigenstrains given by Eqs. (5) of the present analysis suffices. Mikata and Taya ([8]) pointed out that for a coated short fiber with a fiber aspect ratio of $b/a=20$, the value of the stress σ_z inside of each region (fiber, coating, and matrix) is roughly proportional to their corresponding shear modulus G. This proportionality becomes exact for a coated long cylindrical fiber. Specifically, in this example, when instead of a coated short fiber a coated long cylindrical fiber ($b/a\to\infty$) is considered, σ_z/σ_z^0 must equal 10, 2, 1 in the fiber, coating, and matrix, respectively. Applications of the present theory and the method used by Mikata and Taya to the coated short fiber problem considered in this section, show that the values of σ_z/σ_z^0 for the coated short fiber are nearly equal to the solution of the corresponding coated long fiber problem (i.e., nearly equal to 10, 2, 1 in the fiber, coating, and matrix, respectively). In the context of the present theory, this result is further supported by the fact that, the Eshelby's tensors for a cylinder and for an ellipsoid with aspect ratio of $b/a=20$ are approximately equal. It should be emphasized that the method used by Mikata and Taya is valid for axisymmetric loading only, however, such a limitation is not imposed on the present theory which is applicable for more general loading conditions.

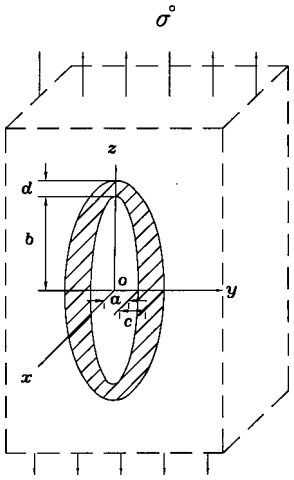


Fig. 5 A coated short fiber model considered by Mikata and Taya [8]

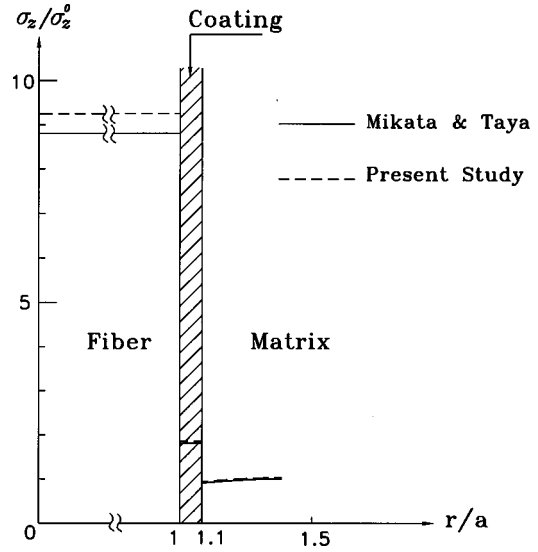


Fig. 6 Variation of σ_z along r in the plane of $z=0$ obtained by the method of the EIM presented herein for the problem shown in Fig. 5

4.2 Coated Continuous Fiber Composites Under Mechanical Loading. Mikata and Taya [18] and later Benveniste et al. [9] considered coated continuous fiber composites under thermomechanical loadings. Mikata and Taya [18] used a four-concentric circular cylindrical model for the stress analysis of coated fiber composites subjected to thermomechanical loadings. Benveniste et al. [9] based on an earlier work of Benveniste [12] calculated approximate stress fields for coated fiber composites. The model of coated continuous fiber composites under transverse normal loading which have been considered by Benveniste et al. [9] is shown in Fig. 7. In this subsection, this problem is re-examined using the method of EDI developed in the present work and the corresponding results are shown in Figs. 8 and 9. The material properties used for this example are given in Table 1. The numerical values of stress components at points A and B inside the coating which have been indicated in Fig. 7, are given in Table 2. The results obtained in the present work agrees very well with those of Benveniste et al. [9]. For this problem as in the previous example of Subsection 4.1, the coating is very thin and consequently the stress fields inside the fiber and coating are nearly uniform. Again, the zeroth-order term in the present analysis yields a reasonable results.

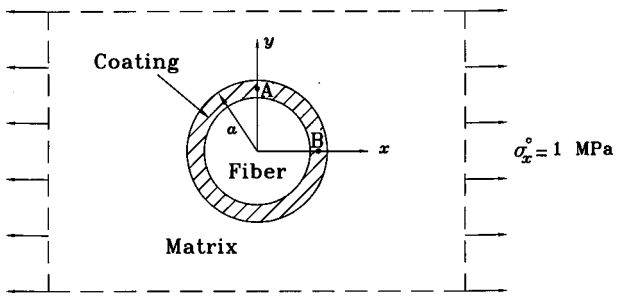


Fig. 7 A continuous coated fiber model under transverse loading considered by Benveniste et al. [9]

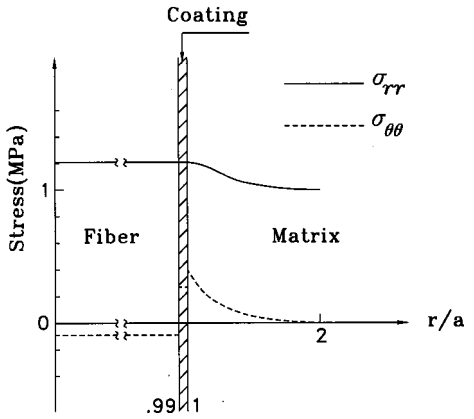


Fig. 8 Stress distributions along the *x*-axis obtained by the method of the EIM presented herein for the problem shown in Fig. 7

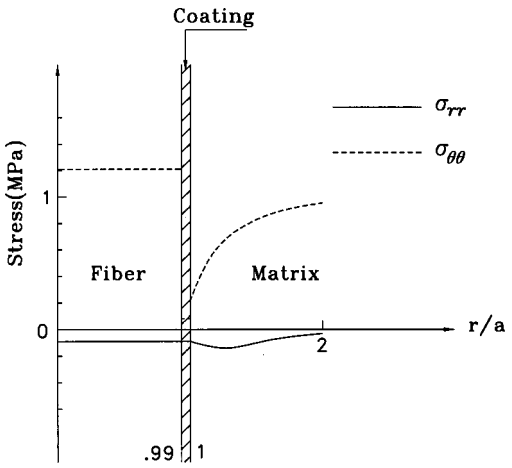


Fig. 9 Stress distributions along the *y*-axis obtained by the method of the EIM presented herein for the problem shown in Fig. 7

4.3 Double-Inhomogeneity System Consisting of Two Eccentric Spheres. Consider the double-inhomogeneity system under far-field uniaxial stress σ_x^0, as shown in Fig. 10. This system consists of a spherical cavity Ω embedded in another sphere Σ, which in turn is surrounded by an infinite matrix $\Phi-\Sigma$. The centers of spheres Ω and Σ whose radii are $r_1=1$ and $r_2=4$, respectively, are located on the *x*-axis and are set apart by the distance $\Delta=2$. Poisson's ratio for the matrix and coating are $\nu_1=\nu_2=0.3$, and coating to matrix stiffness ratio $G_2/G_1=10$. The plots of stress distributions, σ_x/σ_x^0 and $\sigma_y/\sigma_x^0=\sigma_z/\sigma_x^0$ along the *x*-axis are depicted in Fig. 11. This example demonstrates the nonuniformity and variation of stress (strain) fields within the coating, and it is observed that, as moving away from point B toward point A along the *x*-axis, only then the stresses become uniform. This is expected since the effect of cavity and its interaction with the coating becomes negligible, as if the cavity does not exist and consequently the stresses become uniform, in agreement with the Eshelby's single-ellipsoidal inclusion results. It should be noted that the values of σ_y and σ_z become negligible outside the sphere Σ. The values of stress components at the points A, B, C, and D inside Σ, for which the zeroth and first-degree terms of the expansions have been employed, are displayed in Table 3. The results obtained considering zeroth and first-order terms of the expansion are nearly the same as the results obtained using zeroth-order term only.

Table 1 Material properties of a coated fiber composite system, shown in Fig. 7

Material	Elastic Modulus E(Gpa)	Shear Modulus G(Gpa)	Volume Fraction
Sic Fiber	431	172	0.4
Carbon Coating	34.48	14.34	0.0107
Titanium-Al Matrix	96.5	37.1	0.5893

Table 2 Some stress components obtained for the system shown in Fig. 7 by the present theory

Point	$\sigma_{\theta\theta}$ (MPa)	σ_{rr} (MPa)	$\sigma_{r\theta}$ (MPa)
A	0.08	−0.09	0
B	0.27	1.21	0

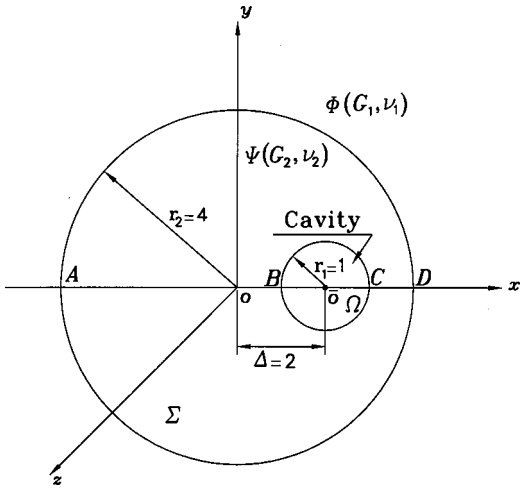


Fig. 10 A double-inhomogeneity system consisting of a cavity Ω surrounded by a spherical inhomogeneity Σ which in turn is surrounded by an infinite domain, under far-field stress σ_x^0

5 Conclusions

In this paper, Eshelby's equivalent inclusion method in conjunction with a superposition scheme described in Section 2 were employed to compute the stress fields in composites with coated inhomogeneities. In general, the elastic fields inside the inhomogeneities and the coatings are not uniform, and hence were expanded in terms of the Cartesian coordinates. As discussed in this

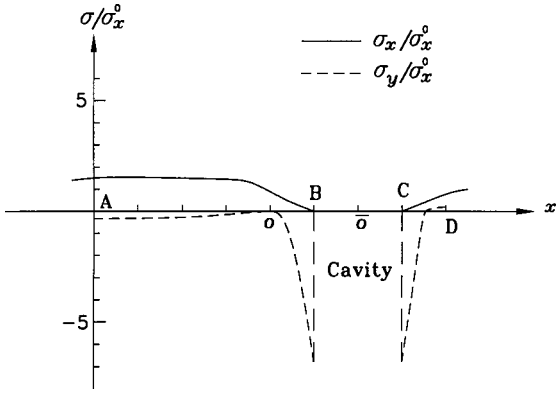


Fig. 11 Stress distributions along the *x*-axis for the problem depicted in Fig. 10

Table 3 Principal stresses for points A, B, C, and D along the *x*-axis, corresponding to Fig. 10

	Point A		Points B and C		Point D	
Terms in the Expansion	$\frac{\sigma_x}{\sigma_x^0}$	$\frac{\sigma_y}{\sigma_x^0}=\frac{\sigma_z}{\sigma_x^0}$	$\frac{\sigma_x}{\sigma_x^0}$	$\frac{\sigma_y}{\sigma_x^0}=\frac{\sigma_z}{\sigma_x^0}$	$\frac{\sigma_x}{\sigma_x^0}$	$\frac{\sigma_y}{\sigma_x^0}=\frac{\sigma_z}{\sigma_x^0}$
Zeroth-order term	1.55	−0.33	0.0	−6.82	0.85	−0.04
Zeroth and first-order terms	1.55	−0.33	0.0	−6.78	0.90	0.1

paper, the problems of confocal spheroids and concentric cylinders with thin coatings are within the range of applicability of the methods proposed in the literature. Re-examination of these problems using the present methodology shows that, even when only the zeroth-order term in the expansions is considered, the stress profiles obtained are in excellent agreement with the ones reported in the literature using other approaches. Moreover, in the present paper a more general problem of a double-inhomogeneity system consisting of two eccentric spheres has been examined. For this problem, the zeroth-order term in the expansions of eigenstrains $\varepsilon_{ij}^{*(1)}(\mathbf{x})$ and $\varepsilon_{ij}^{*(2)}(\mathbf{x})$ yields nearly the same results as when the first-order term of the expansions was also included.

It should be emphasized that the theory presented herein is valid not only for thin, but also for thick coatings, as well as for multilayers of variable thickness. Furthermore, the present theory can handle multi-inhomogeneity systems consisting of ellipsoidal inhomogeneities of arbitrary elastic constants, sizes, and orientations.

Acknowledgments

This research was supported by Sharif University of Technology.

Appendix A

Consider an ellipsoidal inclusion, α in an infinite isotropic elastic medium, whose principal axes coincide with coordinate axes, x_1, x_2, and x_3. Suppose that the eigenstrain in the inclusion is in the form of polynomial of coordinates as given by the expression (11). It follows that the disturbance in strain field can be presented by expression (12), for which the following relations involving $\gamma_{ijkl\ldots}$ hold:

$$8\pi(1-\nu)\gamma_{ijkl}=H_{,klij}-2\nu\delta_{kl}F_{,ij}-(1-\nu)\times[F_{,kj}\delta_{il}+F_{,ki}\delta_{jl}+F_{,lj}\delta_{ik}+F_{,li}\delta_{jk}],$$

$$8\pi(1-\nu)\gamma_{ijklq}=H_{q,klij}-2\nu\delta_{kl}F_{q,ij}-(1-\nu)\times[F_{q,kj}\delta_{il}+F_{q,ki}\delta_{jl}+F_{q,lj}\delta_{ik}+F_{q,li}\delta_{jk}],$$

$$\vdots,$$

such that

$$H(\mathbf{x})=\int_\alpha|\mathbf{x}-\mathbf{x}'|d\mathbf{x}',$$

$$H_{ij\ldots k}(\mathbf{x})=\int_\alpha x_i'x_j'\ldots x_k'|\mathbf{x}-\mathbf{x}'|d\mathbf{x}',$$

$$F(\mathbf{x})=\int_\alpha\frac{d\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|},$$

$$F_{ij\ldots k}(\mathbf{x})=\int_\alpha\frac{x_i'x_j'\ldots x_k'd\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|},$$

where, in general, $\gamma_{ijkl\ldots}(\mathbf{x})\neq\gamma_{klij\ldots}(\mathbf{x})$ unless $i\neq j$ and $k\neq l$. If $\mathbf{x}\in\alpha$ then ([2])

$$\gamma_{ijkl}(\mathbf{x})=\gamma_{ijkl}(0),\quad \mathbf{x}\in\alpha. \tag{A1}$$

Also, the higher order tensors, $\gamma_{ijkl\ldots}$ can be expanded into the following polynomial forms ([4]):

$$\gamma_{ijklq}(\mathbf{x})=\gamma_{ijklq,m}(0)x_m,\quad \mathbf{x}\in\alpha,$$
$$\gamma_{ijklqr}(\mathbf{x})=\gamma_{ijklqr}(0)+\tfrac{1}{2}\gamma_{ijklqr,mn}(0)x_mx_n,\quad \mathbf{x}\in\alpha \tag{A2}$$
$$\vdots,\quad \vdots.$$

On substituting (*A*1) and (*A*2) into the expression (12) an inhomogeneous polynomial of $\mathbf{x}$ for $\epsilon_{ij}^{\alpha}(\mathbf{x};\epsilon^{*(i)}),\mathbf{x}\in\alpha$ will be obtained. The constituent terms of this polynomial are of degree $m,(m-2),(m-4),\ldots$. But if point $\mathbf{x}_0$ is an exterior point with respect to the inclusion α, then Taylors' series expansions of $\gamma_{ijkl\ldots}(\mathbf{x})$ about the point $\mathbf{x}=\mathbf{x}_0$ are used,

$$\gamma_{ijkl}(\mathbf{x})=\gamma_{ijkl}(\mathbf{x}_0)+\gamma_{ijkl,m}(\mathbf{x}_0)x_m+\tfrac{1}{2}\gamma_{ijkl,mn}(\mathbf{x}_0)x_mx_n+\ldots,\quad \mathbf{x}_0\notin\alpha,$$

$$\gamma_{ijklq}(\mathbf{x})=\gamma_{ijklq}(\mathbf{x}_0)+\gamma_{ijklq,m}(\mathbf{x}_0)x_m+\tfrac{1}{2}\gamma_{ijklq,mn}(\mathbf{x}_0)x_mx_n+\ldots,\quad \mathbf{x}_0\notin\alpha, \tag{A3}$$
$$\vdots,\quad \vdots.$$

Substitution of (*A*3) into the expression (12) yields a polynomial for $\epsilon_{ij}^{\alpha}(\mathbf{x};\epsilon^{*(i)})$.

Appendix B

In this Appendix it is shown that by employing the expressions (9), which resulted from the superposition presented in Section 2, the average consistency conditions derived by Hori and Nemat-Nasser [13] are recovered. Denote the volume average of a field quantity (.) over v by $\langle(.)\rangle_v$, and apply volume averages to the consistency conditions (8) over Ω and Ψ, respectively, we obtain

$$C_{ijkl}^1(\epsilon_{kl}^0+\langle\epsilon_{kl}^{d(1)}\rangle_\Omega)=C_{ijkl}(\epsilon_{kl}^0+\langle\epsilon_{kl}^{d(1)}\rangle_\Omega-\langle\epsilon_{kl}^{*(1)}\rangle_\Omega), \tag{B1}$$

$$C_{ijkl}^2(\epsilon_{kl}^0+\langle\epsilon_{kl}^{d(2)}\rangle_\Psi)=C_{ijkl}(\epsilon_{kl}^0+\langle\epsilon_{kl}^{d(2)}\rangle_\Psi-\langle\epsilon_{kl}^{*(2)}\rangle_\Psi), \tag{B2}$$

Taking the volume average of expression (9*a*) over Ω

$$\langle\epsilon_{ij}^{d(1)}(\mathbf{x})\rangle_\Omega=\langle\epsilon_{ij}^{\Sigma}(\mathbf{x};\boldsymbol{\epsilon}^{*(2)})\rangle_\Omega+\langle\epsilon_{ij}^{\Omega}(\mathbf{x};\boldsymbol{\epsilon}^{*(1)})\rangle_\Omega-\langle\epsilon_{ij}^{\Omega}(\mathbf{x};\boldsymbol{\epsilon}^{*(2)})\rangle_\Omega, \tag{B3}$$

where without loss of generality the bars have been dropped. Using Tanaka-Mori result ([15]) we have

$$\langle\epsilon_{ij}^{\Omega}(\mathbf{x};\boldsymbol{\epsilon}^{*(1)})\rangle_\Omega=S_{ijkl}(\Omega)\langle\epsilon_{kl}^{*(1)}\rangle_\Omega, \tag{B4}$$

where $S_{ijkl}(\Omega)$ is given by Eq. (2*a*). The first term in (*B*3) is given by

$$\langle\epsilon_{ij}^{\Sigma}(\mathbf{x};\boldsymbol{\epsilon}^{*(2)})\rangle_\Omega=\frac{1}{\Omega}\int_\Omega\int_\Sigma\Gamma_{ijmn}(\mathbf{x}-\mathbf{x}')\epsilon_{mn}^{*(2)}(\mathbf{x}')d\mathbf{x}'d\mathbf{x}. \tag{B5}$$

If we let

$$\int_\Psi\Gamma_{ijmn}(\mathbf{x}-\mathbf{x}')(\epsilon_{mn}^{*(2)}(\mathbf{x}')-\langle\epsilon_{mn}^{*}\rangle_\Psi)d\mathbf{x}'\equiv0, \tag{B6}$$

as assumed by Hori and Nemat-Nasser [13] which can indeed be justified when the coating is thin enough, then (*B*5) becomes

$$\langle\epsilon_{ij}^{\Sigma}(\mathbf{x};\boldsymbol{\epsilon}^{*(2)})\rangle_\Omega=S_{ijmn}(\Omega)\langle\epsilon_{mn}^{*(2)}\rangle_\Omega+(S_{ijmn}(\Sigma)-S_{ijmn}(\Omega))\langle\epsilon_{mn}^{*(2)}\rangle_\Psi. \tag{B7}$$

The last term in (*B*3) becomes

$$\langle\epsilon_{ij}^{\Omega}(\mathbf{x};\boldsymbol{\epsilon}^{*(2)})\rangle_\Omega=S_{ijmn}(\Omega)\langle\epsilon_{mn}^{*(2)}\rangle_\Omega. \tag{B8}$$

Substitution of (*B*4), (*B*7), and (*B*8) into (*B*3) yields

$$\langle \epsilon_{ij}^{d(1)} \rangle_\Omega = S_{ijkl}(\Omega)\langle \epsilon_{kl}^{*(1)} \rangle_\Omega + (S_{ijkl}(\Sigma) - S_{ijkl}(\Omega))\langle \epsilon_{kl}^{*(2)} \rangle_\Psi . \tag{B9}$$

We have that

$$(\boldsymbol{\epsilon}^d(\mathbf{x}), \boldsymbol{\epsilon}^*(\mathbf{x})) = \begin{cases} (\boldsymbol{\epsilon}^{d(1)}(\mathbf{x}), \boldsymbol{\epsilon}^{*(1)}(\mathbf{x})), & \mathbf{x} \in \Omega, \\ (\boldsymbol{\epsilon}^{d(2)}(\mathbf{x}), \boldsymbol{\epsilon}^{*(2)}(\mathbf{x})), & \mathbf{x} \in \Psi. \end{cases}$$

Define the volume fraction of Ω as $f = \Omega/\Sigma$, then

$$\langle \epsilon_{ij}^d \rangle_\Sigma = S_{ijkl}(\Sigma)\langle \epsilon_{kl}^* \rangle_\Sigma = f\langle \epsilon_{ij}^{d(1)} \rangle_\Omega + (1-f)\langle \epsilon_{ij}^{d(2)} \rangle_\Psi , \tag{B10}$$

$$\langle \epsilon_{ij}^* \rangle_\Sigma = f\langle \epsilon_{ij}^{*(1)} \rangle_\Omega + (1-f)\langle \epsilon_{ij}^{*(2)} \rangle_\Psi . \tag{B11}$$

Combination of (*B*7), (*B*10), and (*B*11) leads to

$$\langle \epsilon_{ij}^{d(2)} \rangle_\Psi = S_{ijkl}(\Sigma)\langle \epsilon_{kl}^{*(2)} \rangle_\Psi + \frac{f}{1-f}(S_{ijkl}(\Sigma) - S_{ijkl}(\Omega)) \times (\langle \epsilon_{kl}^{*(1)} \rangle_\Omega - \langle \epsilon_{kl}^{*(2)} \rangle_\Psi). \tag{B12}$$

Upon substitution of (*B*9) and (*B*12) into (*B*1) and (*B*2) we obtain

$$(C_{ijkl}^1 - C_{ijkl})\{\epsilon_{kl}^0 + S_{klmn}(\Omega)\langle \epsilon_{mn}^{*(1)} \rangle_\Omega + (S_{klmn}(\Sigma) - S_{klmn}(\Omega)) \times \langle \epsilon_{mn}^{*(2)} \rangle_\Psi\} + C_{ijkl}\langle \epsilon_{kl}^{*(1)} \rangle_\Omega = 0, \tag{B13}$$

$$(C_{ijkl}^2 - C_{ijkl})\left\{ \epsilon_{kl}^0 + S_{klmn}(\Sigma)\langle \epsilon_{mn}^{*(2)} \rangle_\Psi + \frac{f}{1-f}(S_{klmn}(\Sigma) - S_{klmn}(\Omega))(\langle \epsilon_{mn}^{*(1)} \rangle_\Omega - \langle \epsilon_{mn}^{*(2)} \rangle_\Psi) \right\} + C_{ijkl}\langle \epsilon_{kl}^{*(2)} \rangle_\Psi = 0, \tag{B14}$$

which are the average consistency conditions obtained by Hori and Nemat-Nasser [13].

References

[1] Mura, T., Shodja, H. M., and Hirose, Y., 1996, ''Inclusion Problems,'' Appl. Mech. Rev., **49**, No. 10, Part 2, pp. S118–S127.

[2] Eshelby, J. D., 1957, ''The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems,'' Proc. R. Soc. London, Ser. A, **A241**, pp. 376–396.

[3] Eshelby, J. D., 1959, ''The Elastic Field Outside an Ellipsoidal Inclusion,'' Proc. R. Soc. London, Ser. A, **A252**, pp. 561–569.

[4] Eshelby, J. D., 1961, ''Elastic Inclusions and Inhomogeneities,'' *Progress in Solid Mechanics*, Vol. 2, I. N. Sneddon and R. Hill, eds., North-Holland, Amsterdam, pp. 89–140.

[5] Moschovidis, Z. A., and Mura, T., 1975, ''Two-Ellipsoidal Inhomogeneities by the Equivalent Inclusion Method,'' ASME J. Appl. Mech., **42**, pp. 847–852.

[6] Chen, T., Dvorak, G. J., and Benveniste, Y., 1990, ''Stress Fields in Composites Reinforced by Coated Cylindrically Orthotropic Fibers,'' Mech. Mater., **9**, pp. 17–32.

[7] Walpole, L. J., 1978, ''A Coated Inclusion in an Elastic Medium,'' Math. Proc. Cambridge Philos. Soc., **83**, pp. 495–506.

[8] Mikata, Y., and Taya, M., 1985, ''Stress Field in and Around a Coated Short Fiber in an Infinite Matrix Subjected to Uniaxial and Biaxial Loadings,'' ASME J. Appl. Mech., **52**, pp. 19–24.

[9] Benveniste, Y., Dvorak, G. J., and Chen, T., 1989, ''Stress Fields in Composites With Coated Inclusions,'' Mech. Mater., **7**, pp. 305–317.

[10] Christensen, R. M., and Lo, K. H., 1979, ''Solutions for Effective Shear Properties in Three Phase Sphere and Cylinder Models,'' J. Mech. Phys. Solids, **27**, pp. 315–330.

[11] Hill, R., 1972, ''An Invariant Treatment of Interfacial Discontinuities in Elastic Composites,'' *Continuum Mechanics and Related Problems of Analysis*, L. I. Sedov, ed., Academy of Sciences USSR, Moscow, pp. 597–604.

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

[13] Hori, M., and Nemat-Nasser, S., 1993, ''Double-Inclusion Model and Overall Moduli of Multi-Phase Composites,'' Mech. Mater., **14**, pp. 189–206.

[14] Nemat-Nasser, S., and Hori, M., 1993, *Micromechanics: Overall Properties of Heterogeneous Solids*, Elsevier, Amsterdam.

[15] Tanaka, K., and Mori, T., 1972, ''Note on Volume Integrals of the Elastic Field Around an Ellipsoidal Inclusion,'' J. Elast., **2**, pp. 199–200.

[16] Mura, T., 1987, *Micromechanics of Defects in Solids 2nd Edition*, Martinus Nijhoff Publishers, Dordrecht, The Netherlands.

[17] Asaro, R. J., and Barnett, D. M., 1975, ''The Non-uniform Transformation Strain Problem for an Anisotropic Ellipsoidal Inclusion,'' J. Mech. Phys. Solids, **23**, pp. 77–83.

[18] Mikata, Y., and Taya, M., 1985b, ''Stress Field in a Coated Continuous Fiber Composite Subjected to Thermomechanical Loadings,'' J. Compos. Mater., **19**, pp. 554–578.

M. R. Vaghar
Consulting Engineer,
Mem. ASME
Benedict Engineering Co., Inc.,
3660 Hartsfield Road,
Tallahassee, FL 32303

H. Garmestani
Associate Professor,
Mem. ASME
Department of Mechanical Engineering and
Center for Material Research and Technology,
FAMU-FSU College of Engineering,
Tallahassee, FL 32310

Three-Dimensional Axisymmetric Stress Analysis of Superconducting Magnets Using Green's Function Solution

A closed-form Green's function solution for the axisymmetric stresses in an elastic coil of superconducting magnets is presented, which provides the components of stress throughout the coil and includes the shear stress in addition to the normal stresses. The Green's function method permits the development of a solution irrespective of the type of magnetic body forces within the coil. Green's functions are derived by using finite Hankel transforms appropriate for a cylindrical coil. [DOI: 10.1115/1.1345700]

1 Introduction

The prediction of stress and strain is essential for both mechanical and electrical design of high-field solenoid magnets. These magnets are designed in a variety of configurations. A superconducting magnet is one example of such magnets, which can be treated as a combination of several solenoid coils, where each coil may be reinforced by a nonconducting layer. Depending upon the geometrical specifications of a coil, magnetic fields may behave differently. These fields result in magnetic body forces, and thus stresses. Traditionally only the tangential component of the stress at the plane perpendicular to the middle section of the longitudinal axis of a coil (midplane) has been considered for design and failure analysis. The value of shear stress has been determined to be small in the midplane but it becomes larger toward the ends of the coil. In the analytical solutions available in the literature, the stress analysis has been performed for just the midplane and shear stress is assumed to be negligible ([1–4]). As a result, a three-dimensional closed-form solution is desired to understand the distribution of stresses (including shear) throughout a solenoid coil. In the present work, a general closed-form solution, using the Green's function method, is derived for an elastic, isotropic coil of a high-field solenoid magnet. This solution is applied to the important case of a superconducting magnet. This analysis is not limited to the midplane and can be used for any type of solenoid magnet.

The use of a Green's function solution is not limited to magnetic body forces. It can be applied to other axisymmetric elasticity problems for finite bodies. The Green's function solution can be used for inclusion problem in composite materials where eigenstrains may be considered as body forces ([5,6]). The Green's function solution can also be applied to specific problems in fracture mechanics and composite materials ([7–9]). Fictitious body forces can be introduced in composite materials, where the difference between the thermal expansion coefficients of the fiber and the matrix results in residual stresses.

A limited number of publications on the approximation of the three-dimensional problem are available in the literature. Some solutions are obtained by neglecting shear throughout the coil ([10]). Other solutions are based upon numerical techniques ([11]) or power series expansion of the fields and displacements ([12]). The direct analytical solutions for infinite or semi-infinite domains are not appropriate for a finite domain such as a magnet ([13,14]).

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, October 1, 1999; final revision, May 8, 2000. Associate Editor: M.-J. Pindera. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Fundamental Equations for the Stress Functions

Consider an elastic isotropic coil with inside radius of a, outside radius of b, and length of $2L$ as shown in Fig. 1. Using the equilibrium equations, constitutive equations, and strain-displacement relationships, the governing equations for displacement vector $\mathbf{u}(r,z)$, for an axisymmetric distribution of body forces $\mathbf{X}(r,z)$, may be written as

$$(\lambda+\mu)\nabla(\nabla.\mathbf{u})+\mu\nabla^2\mathbf{u}+\mathbf{X}=0 \tag{1}$$

where ∇^2 is the three-dimensional Laplacian and λ and μ are (Lamé's) elastic coefficients ([15]).

From the Helmholtz theorem, any vector satisfying Eq. (1) may be resolved into a sum of a gradient and a curl

$$\mathbf{u}=\nabla\phi+\nabla\times\mathbf{A} \tag{2}$$

where $\phi(r,z)$ is a scalar potential and $\mathbf{A}(r,z)$ is a vector potential such that $\nabla.\mathbf{A}=0$. Incorporating the displacement vector from Eq. (2) into Eq. (1) yields an equation in terms of potential functions ϕ and $\mathbf{A}$.

$$(\lambda+2\mu)\nabla(\nabla^2\phi)+\mu\nabla\times(\nabla^2\mathbf{A})+\mathbf{X}=0. \tag{3}$$

The independent potential functions ϕ and $\mathbf{A}$ may be written as

$$\mathbf{A}=\alpha\nabla\times\boldsymbol{\Psi}\qquad \phi=\beta\nabla.\boldsymbol{\Psi} \tag{4}$$

where α and β are arbitrary constants, and components of the vector $\boldsymbol{\Psi}$ are the stress functions. Introducing Eq. (4) into Eq. (3) leads to a partial differential equation for vector $\boldsymbol{\Psi}$.

$$\beta(\lambda+2\mu)\nabla^4\boldsymbol{\Psi}+[\beta(\lambda+2\mu)+\mu\alpha]\nabla\times[\nabla\times(\nabla^2\boldsymbol{\Psi})]+\mathbf{X}=0 \tag{5}$$

In order to simplify Eq. (5), we may choose the arbitrary constants α and β as $-1/\mu$ and $1/(\lambda+2\mu)$, respectively. Thus, Eq. (5) reduces to the component form.

$$\left(\nabla^2-\frac{1}{r^2}\right)^2\Psi_r-\frac{4}{r^4}\frac{\partial^2\Psi_r}{\partial\theta^2}-\frac{4}{r^2}\left(\nabla^2-\frac{1}{r^2}\right)\frac{\partial\Psi_\theta}{\partial\theta}+X_r=0$$

$$\left(\nabla^2-\frac{1}{r^2}\right)^2\Psi_\theta-\frac{4}{r^4}\frac{\partial^2\Psi_\theta}{\partial\theta^2}+\frac{4}{r^2}\left(\nabla^2-\frac{1}{r^2}\right)\frac{\partial\Psi_r}{\partial\theta}+X_\theta=0$$

$$\nabla^4\Psi_z+X_z=0 \tag{6}$$

Because the geometry and loading are axisymmetric, the partial derivative with respect to tangential direction is zero, $\partial/\partial\theta=0$.

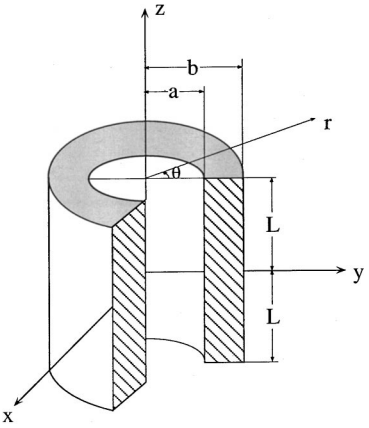


Fig. 1 Schematic diagram representing one coil of a magnet

Hence, the three partial differential equations in Eq. (6) reduce to three uncoupled partial differential equations for radial, tangential, and axial stress functions:

$$\left(\nabla^2-\frac{1}{r^2}\right)^2\Psi_r+X_r=0 \tag{7}$$

$$\left(\nabla^2-\frac{1}{r^2}\right)^2\Psi_\theta+X_\theta=0 \tag{8}$$

$$\nabla^4\Psi_z+X_z=0. \tag{9}$$

The body force in a magnet is the Lorentz force, a function of r and z related to the magnetic field, **B**, and current density, **J**, by

$$\mathbf{X}=\mathbf{J}\times\mathbf{B}. \tag{10}$$

For an axisymmetric distribution of Lorentz force, $\mathbf{J}=J_\theta\mathbf{e}_{\boldsymbol{\theta}}$ and $\mathbf{B}=B_r\mathbf{e}_\mathbf{r}+B_z\mathbf{e}_\mathbf{z}$. Thus, the vector product of the **J** and **B** leads to

$$X_r=J_\theta B_z \quad X_\theta=0 \quad X_z=-J_\theta B_r. \tag{11}$$

In the absence of a tangential magnetic body force in Eq. (8), the tangential stress function will be zero, resulting in a zero tangential displacement u_θ.

3 Finite Hankel Transform

The partial differential equations represented by Eqs. (7) and (9) may be solved by using finite Hankel transforms ([16]). The finite Hankel transform of order n of function $f(r)$ on a closed finite interval $[a, b]$ is defined by

$$\Re_n[f(r)]=\overline{f}(\zeta_i)=\int_a^b rf(r)K_n(\zeta_i,r)dr \tag{12}$$

where ζ_i is a root of the transcendental equation

$$J_n(\zeta_i a)Y_n(\zeta_i b)-J_n(\zeta_i b)Y_n(\zeta_i a)=0 \tag{13}$$

and $K_n(\zeta_i,r)$ is the Fourier Bessel kernel.

$$K_n(\zeta_i,r)=[J_n(\zeta_i r)Y_n(\zeta_i b)-J_n(\zeta_i b)Y_n(\zeta_i r)] \tag{14}$$

The inverse transform for the finite Hankel transform is

$$f(r)=\Re_n^{-1}[\overline{f}(\zeta_i)]=\frac{\pi^2}{2}\sum_i\frac{\zeta_i^2J_n^2(\zeta_i a)}{J_n^2(\zeta_i a)-J_n^2(\zeta_i b)}\overline{f}(\zeta_i)K_n(\zeta_i,r). \tag{15}$$

where the summation is extended over all positive roots ζ_i. The finite Hankel transform of a Laplacian of $f(r)$ (in cylindrical coordinate) is given by

$$\Re_n\left[\left(\nabla^2-\frac{n^2}{r^2}\right)f\right]=\Re_n\left[\left(\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{n^2}{r^2}\right)f\right]$$
$$=\frac{2}{\pi}\left(f(b)-\frac{J_n(\zeta_i b)}{J_n(\zeta_i a)}f(a)\right)-\zeta_i^2\Re_n[f]. \tag{16}$$

4 Radial Green's Function

The Radial Green's function for Ψ_r, radial stress function, is obtained by solving Eq. (7) with Dirichlet homogeneous boundary conditions. Applying the finite Hankel transform (of order one) to each term of Eq. (7) yields

$$\Re_1\left[\left(\nabla^2-\frac{1}{r^2}\right)^2\Psi_r\right]=-\Re_1[X_r]. \tag{17}$$

Use of Eq. (16) in Eq. (17) results in

$$\left(-\zeta_{1i}^2+\frac{d^2}{dz^2}\right)^2\overline{\Psi}_r(\zeta_{1i},z)=-\overline{X}_r(\zeta_{1i},z) \tag{18}$$

where ζ_{1i} is a root of the transcendental equation

$$J_1(\zeta_{1i}a)Y_1(\zeta_{1i}b)-J_1(\zeta_{1i}b)Y_1(\zeta_{1i}a)=0 \tag{19}$$

and

$$\overline{\Psi}_r(\zeta_{1i},z)=\int_a^b r\Psi_r(r,z)K_1(\zeta_{1i},r)dr \tag{20}$$

$$\overline{X}_r(\zeta_{1i},z)=\int_a^b rX_r(r,z)K_1(\zeta_{1i},r)dr, \tag{21}$$

where

$$K_1(\zeta_{1i},r)=[J_1(\zeta_{1i}r)Y_1(\zeta_{1i}b)-J_1(\zeta_{1i}b)Y_1(\zeta_{1i}r)] \tag{22}$$

is the Fourier Bessel kernel. An additional transform (in the axial direction) is needed for solving Eq. (18). Considering the radial body force is an even function of z and the interval is finite $[-L,L]$, an appropriate transform is the finite Fourier cosine transform. By introducing the finite Fourier cosine transform (in the axial direction) of Eq. (17),

$$\Im_C\left[\left(-\zeta_{1i}^2+\frac{\partial^2}{\partial z^2}\right)^2\overline{\Psi}_r(\zeta_{1i},z)\right]=-\Im_C[\overline{X}_r(\zeta_{1i},z)],$$

the differential equation is converted into the algebraic equation

$$\left(-\zeta_{1i}^2-\frac{n^2\pi^2}{L^2}\right)^2\bar{\bar{\Psi}}_r(\zeta_{1i},n)=-\bar{\bar{X}}_r(\zeta_{1i},n) \tag{23}$$

where n is an integer. The functions

$$\bar{\bar{\Psi}}_r(\zeta_{1i},n)=\int_{-L}^{L}\overline{\Psi}_r(\zeta_{1i},z)\cos\frac{n\pi z}{L}dz \tag{24}$$

$$\bar{\bar{X}}_r(\zeta_{1i},n)=\int_{-L}^{L}\overline{X}_r(\zeta_{1i},z)\cos\frac{n\pi z}{L}dz \tag{25}$$

are the finite Fourier cosine transforms of $\overline{\Psi}_r(\zeta_{1i},z)$ and $\overline{X}_r(\zeta_{1i},z)$. The inverse finite Fourier cosine transform of $\bar{\bar{\Psi}}_r(\zeta_{1i},n)$ and inverse finite Hankel transform of $\overline{\Psi}_r(\zeta_{1i},z)$ are defined by

$$\overline{\Psi}_r(\zeta_{1i},z)=\Im_C^{-1}[\bar{\bar{\Psi}}_r(\zeta_{1i},n)]$$
$$=\frac{1}{2L}\bar{\bar{\Psi}}_r(\zeta_{1i},0)+\frac{1}{L}\sum_{n=1}^{\infty}\bar{\bar{\Psi}}_r(\zeta_{1i},n)\cos\frac{n\pi z}{L} \tag{26}$$

$$\Psi_r(r,z)=\Re_1^{-1}[\bar{\Psi}_r(\zeta_{1i},z)]$$
$$=\frac{\pi^2}{2}\sum_{i=1}^{\infty}\frac{\zeta_{1i}^2J_1^2(\zeta_{1i}a)}{J_1^2(\zeta_{1i}a)-J_1^2(\zeta_{1i}b)}\bar{\Psi}_r(\zeta_{1i},z)K_1(\zeta_{1i},r). \quad (27)$$

Substitution of Eqs. (26) and (27) into Eq. (23) yields an equation for the radial stress function:

$$\Psi_r(r,z)=-\frac{\pi^2}{2}\sum_{i=1}^{\infty}\frac{\zeta_{1i}^2J_1^2(\zeta_{1i}a)}{J_1^2(\zeta_{1i}a)-J_1^2(\zeta_{1i}b)}\left[\frac{1}{2L\zeta_{1i}^4}\bar{\bar{X}}_r(\zeta_{1i},0)+L^3\sum_{n=1}^{\infty}\frac{1}{(L^2\zeta_{1i}^2+n^2\pi^2)^2}\bar{\bar{X}}_r(\zeta_{1i},n)\cos\frac{n\pi z}{L}\right]K_1(\zeta_{1i},r). \quad (28)$$

Introducing Eqs. (21) and (25) into Eq. (28) gives the solution of the radial stress function $\Psi_r(r,z)$ in terms of the radial body force $X_r(r,z)$.

$$\Psi_r(r,z)=-\frac{\pi^2}{2}\sum_{i=1}^{\infty}\left\{\frac{1}{2L\zeta_{1i}^4}\int_{-La}^{L\ b}r'X_r(r',z')K_1(\zeta_{1i},r')dr'dz'+L^3\sum_{n=1}^{\infty}\left[\frac{1}{(L^2\zeta_{1i}^2+n^2\pi^2)^2}\cos\frac{n\pi z}{L}\int_{-La}^{L\ b}r'X_r(r',z')K_1(\zeta_{1i},r')\cos\frac{n\pi z'}{L}dr'dz'\right]\frac{\zeta_{1i}^2J_1^2(\zeta_{1i}a)}{J_1^2(\zeta_{1i}a)-J_1^2(\zeta_{1i}b)}K_1(\zeta_{1i},r)\right\} \quad (29)$$

By interchanging integrals with summations, we may write Eq. (29) in the form

$$\Psi_r(r,z)=\int_{-L}^{L}\int_{a}^{b}X_r(r',z')G_r(r,r',z,z')dr'dz' \quad (30)$$

where

$$G_r(r,r',z,z')=-\sum_{i=1}^{\infty}\left\{\frac{\pi^2}{2}\frac{\zeta_{1i}^2J_1^2(\zeta_{1i}a)}{J_1^2(\zeta_{1i}a)-J_1^2(\zeta_{1i}b)}r'\times K_1(\zeta_{1i},r)K_1(\zeta_{1i},r')\times\left[\frac{1}{2L\zeta_{1i}^4}+L^3\sum_{n=1}^{\infty}\frac{1}{(L^2\zeta_{1i}^2+n^2\pi^2)^2}\times\cos\frac{n\pi z'}{L}\cos\frac{n\pi z}{L}\right]\right\} \quad (31)$$

is the radial Green's function.

5 Axial Green's Function

The axial Green's function is obtained by solving the partial differential equation for the axial stress function, Eq. (9), with Dirichlet homogeneous boundary conditions. Here, finite Hankel transform in the radial direction and finite Fourier sine transform in the axial direction (since the axial body force is an odd function of z) are used.

Applying the finite Hankel transform of order zero in r, and finite Fourier since transform in z, to Eq. (9) yields

$$\left(-\zeta_{0i}^2-\frac{n^2\pi^2}{L^2}\right)^2\bar{\bar{\Psi}}_z(\zeta_{0i},n)=-\bar{\bar{X}}_z(\zeta_{0i},n) \quad (32)$$

where n is an integer, ζ_{0i} satisfies

$$J_0(\zeta_{0i}a)Y_0(\zeta_{0i}b)-J_0(\zeta_{0i}b)Y_0(\zeta_{0i}a)=0 \quad (33)$$

and

$$\bar{\bar{\Psi}}_z(\zeta_{0i},n)=\Re_0[\Im_s[\Psi_z(r,z)]]=\int_{-La}^{L\ b}r\Psi_z(r,z)K_0(\zeta_{0i},r)\sin\frac{n\pi z}{L}drdz \quad (34)$$

$$\bar{\bar{X}}_z(\zeta_{0i},n)=\Re_0[\Im_s[X_z(r,z)]]=\int_{-La}^{L\ b}rX_z(r,z)K_0(\zeta_{0i},r)\sin\frac{n\pi z}{L}drdz \quad (35)$$

are transforms of the axial stress function and axial body force. Here,

$$K_0(\zeta_{0i},r)=[J_0(\zeta_{0i}r)Y_0(\zeta_{0i}b)-J_0(\zeta_{0i}b)Y_0(\zeta_{0i}r)] \quad (36)$$

is the Fourier Bessel kernel for the zero-order transformation. The inverse transform of $\bar{\bar{\Psi}}_z(\zeta_{0i},n)$ is

$$\Psi_z(r,z)=\Re_0^{-1}[\Im_s^{-1}[\bar{\bar{\Psi}}_z(\zeta_{0i},n)]]=\frac{\pi^2}{2L}\sum_{i=1}^{\infty}\sum_{n=1}^{\infty}\frac{\zeta_{0i}^2J_0^2(\zeta_{0i}a)}{J_0^2(\zeta_{0i}a)-J_0^2(\zeta_{0i}b)}\times\sin\frac{n\pi z}{L}\bar{\bar{\Psi}}_z(\zeta_{0i},n)K_0(\zeta_{0i},r). \quad (37)$$

Incorporating Eqs. (32) and (35) into Eq. (37) gives the solution to the axial stress function $\Psi_z(r,z)$ in terms of the axial body force $X_z(r,z)$.

$$\Psi_z(r,z)=\int_{-L}^{L}\int_{a}^{b}X_z(r',z')G_z(r,r',z,z')dr'dz' \quad (38)$$

where

$$G_z(r,r',z,z')=-\frac{\pi^2}{2}\sum_{i=1}^{\infty}\sum_{n=1}^{\infty}\left[\frac{L^3}{(L^2\zeta_{0i}^2+n^2\pi^2)^2}\frac{\zeta_{0i}^2J_0^2(\zeta_{0i}a)}{J_0^2(\zeta_{0i}a)-J_0^2(\zeta_{0i}b)}r'K_0(\zeta_{0i},r')K_0(\zeta_{0i},r)\sin\frac{n\pi z'}{L}\sin\frac{n\pi z}{L}\right] \quad (39)$$

is the axial Green's function.

6 Boundary Conditions

The displacement vector is related to the vector $\boldsymbol{\Psi}$ by Eq. (40).

$$\mathbf{u}=-\frac{1}{\mu}\nabla\times(\nabla\times\boldsymbol{\Psi})+\frac{1}{\lambda+2\mu}\nabla(\nabla.\boldsymbol{\Psi}) \tag{40}$$

The stress tensor in terms of the displacement vector is defined by

$$\boldsymbol{\sigma}=\lambda(\nabla.\mathbf{u})\mathbf{I}+\mu[\nabla\mathbf{u}+\nabla^T\mathbf{u}] \tag{41}$$

where $\mathbf{I}$ is the identity tensor. Substituting Eq. (40) into Eq. (41) results in stresses in terms of axial and radial stress functions

$$\sigma_r=2\frac{\partial}{\partial r}\left(\nabla^2-\frac{1}{r^2}\right)\Psi_r+\frac{1}{1-\nu}\left(\nu\nabla^2-\frac{\partial^2}{\partial r^2}\right)\varphi$$

$$\sigma_\theta=\frac{2}{r}\left(\nabla^2-\frac{1}{r^2}\right)\Psi_r+\frac{1}{1-\nu}\left(\nu\nabla^2-\frac{1}{r}\frac{\partial}{\partial r}\right)\varphi$$

$$\sigma_z=2\frac{\partial}{\partial z}(\nabla^2\Psi_z)+\frac{1}{1-\nu}\left(\nu\nabla^2-\frac{\partial^2}{\partial z^2}\right)\varphi$$

$$\sigma_{rz}=\frac{\partial}{\partial r}(\nabla^2\Psi_z)+\frac{\partial}{\partial z}\left(\nabla^2-\frac{1}{r^2}\right)\Psi_r-\frac{1}{1-\nu}\frac{\partial^2\varphi}{\partial r\,\partial z} \tag{42}$$

where ν is the Poisson's ratio and

$$\varphi=\frac{1}{r}\frac{\partial}{\partial r}(r\Psi_r)+\frac{\partial\Psi_z}{\partial z} \tag{43}$$

is the divergence of the vector $\boldsymbol{\Psi}$.

Traction-free boundary conditions are appropriate for a solenoid coil. Therefore, the radial and shear stresses should be zero at the inside and outside radii ($r=a$ and $r=b$), and the axial and shear stresses should be zero at the ends of the coil ($z=\pm L$).

Substituting the solutions for radial and axial stress functions from Eqs. (30), (31), (38), and (39) into Eq. (42) and computing radial and shear stresses at the inside and outside radii, and the axial and shear stresses at the ends of the coil, yields

$$\sigma_r(a,z)=\sum_{n=0}^{\infty}\wp_1(n)\cos\frac{n\pi z}{L} \tag{44}$$

$$\sigma_{rz}(a,z)=\sum_{n=1}^{\infty}\wp_2(n)\sin\frac{n\pi z}{L}$$

$$\sigma_r(b,z)=\sum_{n=0}^{\infty}\wp_3(n)\cos\frac{n\pi z}{L}$$

$$\sigma_{rz}(b,z)=\sum_{n=1}^{\infty}\wp_4(n)\sin\frac{n\pi z}{L}$$

$$\sigma_z(r,\pm L)=\sum_{i=1}^{\infty}[\wp_5(\zeta_{0i})K_0(\zeta_{0i},r)+\wp_6(\zeta_{1i})K_0^*(\zeta_{1i},r)]$$

$$\sigma_{rz}(r,\pm L)=0$$

where

$$K_0^*(\zeta_{1i},r)=[J_0(\zeta_{1i}r)Y_1(\zeta_{1i}b)-J_1(\zeta_{1i}b)Y_0(\zeta_{1i}r)]$$

$$=\frac{1}{r\zeta_{1i}}K_1(\zeta_{1i},r)+\frac{1}{\zeta_{1i}}\frac{\partial}{\partial r}K_1(\zeta_{1i},r), \tag{45}$$

and $\wp_1(n)$ through $\wp_4(n)$, $\wp_5(\zeta_{0i})$ and $\wp_6(\zeta_{1i})$ are given by

$$\wp_1(n)=\sum_{i=1}^{\infty}\frac{2}{\pi a^2}\frac{1}{1-\nu}\left\{-\Gamma_z(\zeta_{0i},n)\frac{n\pi}{L}\frac{J_0(\zeta_{0i}b)}{J_0(\zeta_{0i}a)}+a\Gamma_r(\zeta_{1i},n)\right.$$

$$\left.\times\left[\frac{n^2\pi^2}{L^2}(2-\nu)+\zeta_{1i}^2(1-\nu)\right]\frac{J_1(\zeta_{0i}b)}{J_1(\zeta_{0i}a)}\right\} \tag{46}$$

$$\wp_2(n)=\sum_{i=1}^{\infty}\frac{2}{\pi a}\frac{1}{1-\nu}\left\{-\Gamma_z(\zeta_{0i},n)\left[\nu\frac{n^2\pi^2}{L^2}\right.\right.$$

$$\left.\left.-\zeta_{0i}^2(1-\nu)\right]\frac{J_0(\zeta_{0i}b)}{J_0(\zeta_{0i}a)}\right\}$$

$$\wp_3(n)=\sum_{i=1}^{\infty}\frac{2}{\pi b^2}\frac{1}{1-\nu}\left\{-\Gamma_z(\zeta_{0i},n)\frac{n\pi}{L}\right.$$

$$\left.+b\Gamma_r(\zeta_{1i},n)\left[\frac{n^2\pi^2}{L^2}(2-\nu)+\zeta_{1i}^2(1-\nu)\right]\right\}$$

$$\wp_4(n)=\sum_{i=1}^{\infty}\frac{1}{1-\nu}\frac{2}{\pi b}\left\{-\Gamma_z(\zeta_{0i},n)\left[\nu\frac{n^2\pi^2}{L^2}-\zeta_{0i}^2(1-\nu)\right]\right\}$$

$$\wp_5(\zeta_{0i})=\sum_{n=0}^{\infty}\frac{(-1)^n}{1-\nu}\left\{-\Gamma_z(\zeta_{0i},n)\frac{n\pi}{L}\left[\frac{n^2\pi^2}{L^2}(1-\nu)\right.\right.$$

$$\left.\left.+\zeta_{0i}^2(2-\nu)\right]\right\}$$

$$\wp_6(\zeta_{1i})=\sum_{n=0}^{\infty}\frac{(-1)^n}{1-\nu}\left\{\Gamma_r(\zeta_{1i},n)\zeta_{1i}\left[\frac{n^2\pi^2}{L^2}(1-\nu)-\nu\zeta_{1i}^2\right]\right\}$$

with

$$\Gamma_z(\zeta_{0i},n)=-\frac{\pi^2}{2}\left\{\frac{L^3}{(L^2\zeta_{0i}^2+n^2\pi^2)^2}\frac{\zeta_{0i}^2J_0^2(\zeta_{0i}a)}{J_0^2(\zeta_{0i}a)-J_0^2(\zeta_{0i}b)}\right.$$

$$\left.\times\int_{-L}^{L}\int_a^b r'X_z(r',z')K_0(\zeta_{0i},r')\sin\frac{n\pi z'}{L}dr'dz'\right\} \tag{47}$$

$$\Gamma_r(\zeta_{1i},n)=-\left\{\frac{\pi^2}{2}\frac{\zeta_{1i}^2J_1^2(\zeta_{1i}a)}{J_1^2(\zeta_{1i}a)-J_1^2(\zeta_{1i}b)}\frac{L^3}{(L^2\zeta_{1i}^2+n^2\pi^2)^2}\right.$$

$$\left.\times\int_{-L}^{L}\int_a^b r'X_r(r',z')K_1(\zeta_{1i},r')\cos\frac{n\pi z'}{L}dr'dz'\right\}$$

$$\Gamma_r(\zeta_{1i},0)=-\frac{\pi^2}{2}\frac{J_1^2(\zeta_{1i}a)}{J_1^2(\zeta_{1i}a)-J_1^2(\zeta_{1i}b)}\frac{1}{2L\zeta_{1i}^2}$$

$$\times\int_{-L}^{L}\int_a^b r'X_r(r',z')K_1(\zeta_{1i},r')dr'dz'.$$

From Eq. (44), it can be observed that except for the shear stress at the ends of the coil, boundary conditions are not satisfied. The radial and shear stresses impose forcing functions of z at the radial boundaries and the axial stress asserts a forcing function of r at the axial boundaries. Thus, a complementary solution for either radial or axial stress function (since stresses are related to both) is needed to neutralize these forcing functions.

7 Complementary Solution for the Axial Stress Function

Let us consider function $\xi(r,z)$ (an odd function in z) as a complementary function for the axial stress function. From Eq. (9), $\xi(r,z)$ must satisfy the homogeneous part of the partial differential equation for the axial stress function.

$$\nabla^4\zeta(r,z)=0 \tag{48}$$

From Eq. (42) radial, axial, and shear stresses are expressed in terms of $\xi(r,z)$.

$$\sigma_r=\frac{1}{1-\nu}\frac{\partial}{\partial z}\left(\nu\nabla^2-\frac{\partial^2}{\partial r^2}\right)\xi(r,z) \tag{49}$$

$$\sigma_z = \frac{1}{1-\nu}\frac{\partial}{\partial z}\left[(2-\nu)\nabla^2 - \frac{\partial^2}{\partial z^2}\right]\xi(r,z)$$

$$\sigma_{rz} = \frac{1}{1-\nu}\frac{\partial}{\partial r}\left[(1-\nu)\nabla^2 - \frac{\partial^2}{\partial z^2}\right]\xi(r,z)$$

The complementary function must reverse the effect of the imposed forcing functions by stresses at the boundaries. As a result, from Eqs. (44) and (49) boundary conditions for $\xi(r,z)$ are obtained and given by

$$\frac{1}{1-\nu}\frac{\partial}{\partial z}\left(\nu\nabla^2 - \frac{\partial^2}{\partial r^2}\right)\xi(r,z)\bigg|_{r=a} = -\sum_{n=0}^{\infty}\wp_1(n)\cos\frac{n\pi z}{L} \tag{50}$$

$$\frac{1}{1-\nu}\frac{\partial}{\partial r}\left[(1-\nu)\nabla^2 - \frac{\partial^2}{\partial z^2}\right]\xi(r,z)\bigg|_{r=a} = -\sum_{n=1}^{\infty}\wp_2(n)\sin\frac{n\pi z}{L}$$

$$\frac{1}{1-\nu}\frac{\partial}{\partial z}\left(\nu\nabla^2 - \frac{\partial^2}{\partial r^2}\right)\xi(r,z)\bigg|_{r=b} = -\sum_{n=0}^{\infty}\wp_3(n)\cos\frac{n\pi z}{L}$$

$$\frac{1}{1-\nu}\frac{\partial}{\partial r}\left[(1-\nu)\nabla^2 - \frac{\partial^2}{\partial z^2}\right]\xi(r,z)\bigg|_{r=b} = -\sum_{n=1}^{\infty}\wp_4(n)\sin\frac{n\pi z}{L}$$

$$\frac{1}{1-\nu}\frac{\partial}{\partial r}\left[(2-\nu)\nabla^2 - \frac{\partial^2}{\partial z^2}\right]\xi(r,z)\bigg|_{z=L} = -\sum_{i=1}^{\infty}\left[\wp_5(\zeta_{0i})K_0(\zeta_{0i},r) + \wp_6(\zeta_{1i})K_0^*(\zeta_{1i},r)\right]$$

$$\frac{1}{1-\nu}\frac{\partial}{\partial r}\left((1-\nu)\nabla^2 - \frac{\partial^2}{\partial z^2}\right)\xi(r,z)\bigg|_{z=L} = 0.$$

The partial differential equation for $\xi(r,z)$ with the given boundary conditions may be solved by using the superposition principle. The substitution of $\xi(r,z) = \xi_1(r,z) + \xi_2(r,z)$ into Eq. (48) yields two partial differential equations for $\xi_1(r,z)$ and $\xi_2(r,z)$.

$$\nabla^4\xi_1(r,z) = 0 \tag{51}$$

$$\nabla^4\xi_2(r,z) = 0 \tag{52}$$

Solution to $\xi_1(r,z)$ is achieved by applying the finite Hankel transform of order zero to Eq. (51) and solving the resulting differential equation for z.

$$\xi_1(r,z) = \pi^2\sum_{i=1}^{\infty}\frac{\zeta_{0i}^2 I_0^2(\zeta_{0i}a)}{J_0^2(\zeta_{0i}a) - J_0^2(\zeta_{0i}b)}\left[A_i\sinh(\zeta_{0i}z) + B_i z\cosh(\zeta_{0i}z)\right]K_0(\zeta_{0i},r) \tag{53}$$

Here, A_i and B_i are arbitrary constants. Solution to $\xi_2(r,z)$ is obtained by employing the finite Fourier sine transform to Eq. (52) and solving the ensuing differential equation for r.

$$\xi_2(r,z) = \frac{1}{2\pi}\sum_{n=1}^{\infty}\left[\frac{1}{n}\hat{A}_n r I_1\left(\frac{n\pi}{L}r\right) + \frac{1}{n}\hat{B}_n r K_1\left(\frac{n\pi}{L}r\right) + \frac{2\pi}{L}\hat{C}_n I_0\left(\frac{n\pi}{L}r\right) + \frac{2\pi}{L}\hat{D}_n K_0\left(\frac{n\pi}{L}r\right)\right]\sin\frac{n\pi z}{L} \tag{54}$$

In Eq. (54), $\hat{A}_n$, $\hat{B}_n$, $\hat{C}_n$, and $\hat{D}_n$ are arbitrary constants, $I_0[(n\pi/L)r]$ and $I_1[(n\pi/L)r]$ are the modified Bessel functions of the first kind, and $K_0[(n\pi/L)r]$ and $K_1[(n\pi/L)r]$ are the modified Bessel functions of the second kind. The superposition of Eqs. (53) and (54) furnishes the solution to $\xi(r,z)$.

$$\xi(r,z) = \pi^2\sum_{i=1}^{\infty}\frac{\zeta_{0i}^2 J_0^2(\zeta_{0i}a)}{J_0^2(\zeta_{0i}a) - J_0^2(\zeta_{0i}b)}\left[A_i\sinh(\zeta_{0i}z) + B_i z\cosh(\zeta_{0i}z)\right]K_0(\zeta_{0i},r) + \frac{1}{2\pi}\sum_{n=1}^{\infty}\left[\frac{1}{n}\hat{A}_n r I_1\left(\frac{n\pi}{L}r\right) + \frac{1}{n}\hat{B}_n r K_1\left(\frac{n\pi}{L}r\right) + \frac{2\pi}{L}\hat{C}_n I_0\left(\frac{n\pi}{L}r\right) + \frac{2\pi}{L}\hat{D}_n K_0\left(\frac{n\pi}{L}r\right)\right]\sin\frac{n\pi z}{L} \tag{55}$$

The six arbitrary constants in Eq. (55) may be evaluated by using the six boundary conditions given by Eq. (50). Applying the shear stress boundary condition at $z = L$ yields

$$B_i = \omega(\zeta_{0i})A_i \tag{56}$$

where $\omega(\zeta_{0i})$ is expressed by Eq. (57).

$$\omega_{(\zeta_{0i})} = \frac{-\zeta_{0i}}{2\nu + \zeta_{0i}L\coth(\zeta_{0i}L)} \tag{57}$$

Employing the radial boundary conditions to $\xi(r,z)$ and using Eq. (57) results in

$$\chi_{11}(n)\hat{A}_n + \chi_{12}(n)\hat{B}_n + \chi_{13}(n)\hat{C}_n + \chi_{14}(n)\hat{D}_n + \sum_{i=1}^{\infty}\Lambda_1(\zeta_{0i},n)A_i = \wp_1(n) \tag{58}$$

$$\chi_{21}(n)\hat{A}_n + \chi_{22}(n)\hat{B}_n + \chi_{23}(n)\hat{C}_n + \chi_{24}(n)\hat{D}_n + \sum_{i=1}^{\infty}\Lambda_2(\zeta_{0i},n)A_i = \wp_2(n) \tag{59}$$

$$\chi_{31}(n)\hat{A}_n + \chi_{32}(n)\hat{B}_n + \chi_{33}(n)\hat{C}_n + \chi_{34}(n)\hat{D}_n + \sum_{i=1}^{\infty}\Lambda_3(\zeta_{0i},n)A_i = \wp_3(n) \tag{60}$$

$$\chi_{41}(n)\hat{A}_n + \chi_{42}(n)\hat{B}_n + \chi_{43}(n)\hat{C}_n + \chi_{44}(n)\hat{D}_n + \sum_{i=1}^{\infty}\Lambda_4(\zeta_{0i},n)A_i = \wp_4(n) \tag{61}$$

where $\chi_{11}(n)$ through $\chi_{44}(n)$ and $\Lambda_1(\zeta_{0i},n)$ through $\Lambda_4(\zeta_{0i},n)$ are given in Appendix A. The boundary condition for axial stress at $z = L$ provides

$$\Lambda_5(\zeta_{0i})A_i + \sum_{n=1}^{\infty}\left[\chi_{51}(\zeta_{0i},n)\hat{A}_n + \chi_{52}(\zeta_{0i},n)\hat{B}_n + \chi_{53}(\zeta_{0i},n)\hat{C}_n + \chi_{54}(\zeta_{0i},n)\hat{D}_n\right] = \Gamma_5(\zeta_{0i},\zeta_{1i}) \tag{62}$$

where $\chi_{51}(\zeta_{0i},n)$ through $\chi_{54}(\zeta_{0i},n)$, $\Lambda_5(\zeta_{0i})$ and $\Gamma_5(\zeta_{0i},\zeta_{1i})$ are given in Appendix B.

Equations (58)–(62) represent a system of equations where the unknowns are $\hat{A}_n$, $\hat{B}_n$, $\hat{C}_n$, $\hat{D}_n$ and A_i. To evaluate these unknowns, the infinite series in Eqs. (58)–(62) are replaced by finite summations with an acceptable truncation error. Hence, the infinite upper limits for i and n are changed to finite values of M and N, respectively. Expanding these finite summations would result in a system of equations with $4N + M$ unknowns and equations, where unknowns are $\hat{A}_1 - \hat{A}_N$, $\hat{B}_1 - \hat{B}_N$, $\hat{C}_1 - \hat{C}_N$, $\hat{D}_1 - \hat{D}_N$ and $A_1 - A_M$. Equation (62) gives M equations by letting i vary from 1 to M. Moreover, allowing n to advance from 1 to N in Eqs. (58)–(61), results in $4N$ equations. By solving this system of equations, the arbitrary constants for the complementary solution of the axial stress function are obtained. The combination of the

complementary and the Green's function solutions for the axial stress function yields a solution that satisfies both the boundary conditions and the axial body force. This solution together with the radial Green's function determines the distribution of stresses in a given coil.

8 Numerical Results

The Green's function solution is applied to a 23 Tesla superconducting coil. The parameters for this coil are given in Table 1. Figures 2 and 3 show the tangential and radial stresses through the coil along the radius at three different axial positions ($z=0$, $z=L/2$ and $z=L$). Figure 4 shows the characteristics of the axial stress through the coil along the radius at the midplane and $z=L/2$; and Fig. 5 shows the shear stress at $z=L/2$. Note that due to traction free boundary conditions, axial and shear stresses are zero at $z=L$ and shear stress is zero at the midplane ($z=0$) because of symmetry.

Table 1 Parameters for the 23 T superconducting coil

Name	Symbol	Value	Unit
Inner radius	a	100.00	mm
Outer radius	b	136.50	mm
Half length	L	28.00	mm
Elastic modulus	E	111.00	GPa
Poisson's ratio	ν	0.30	
Current density	J	530.10	A/mm^2

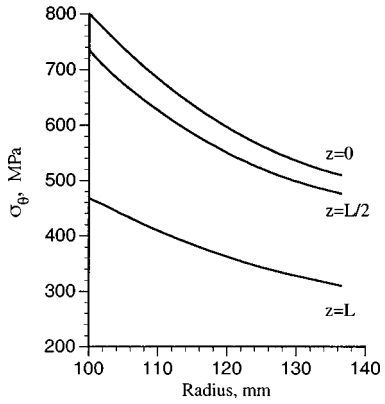


Fig. 2 Distribution of the tangential stress for a 23 Tesla superconducting coil

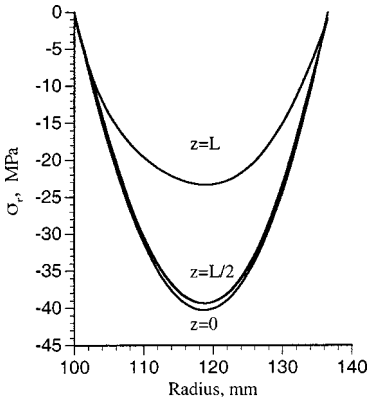


Fig. 3 Distribution of the radial stress for a 23 Tesla superconducting coil

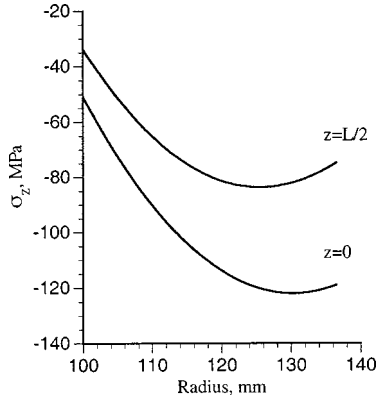


Fig. 4 Distribution of the axial stress for a 23 Tesla superconducting coil

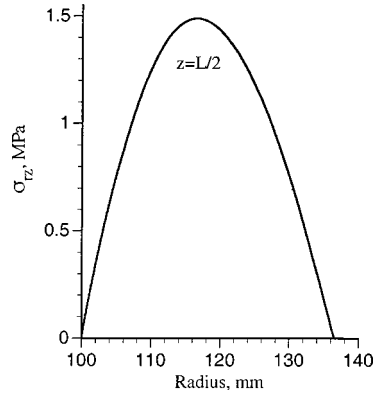


Fig. 5 Distribution of the shear stress for a 23 Tesla superconducting coil

9 Conclusions

Analytical closed-form solution for the distribution of stresses has been developed for a coil of high-field solenoid magnets, including superconducting magnets. This solution is presented in forms of the Green's functions, which permits the development of a solution irrespective of the type of the field or its distribution within a coil. The problem was formulated in terms of stress functions. Green's functions were derived by using finite Hankel and finite Fourier transforms. Boundary conditions were satisfied by introducing a complementary solution for the axial stress function. The radial Green's function with the superposition of the complementary and the axial Green's function provide a comprehensive analytical solution for the stresses.

The Green's function solution provides a complete analytical stress solution for an isotropic coil. This solution should be used as a foundation for the stress analysis of multilayer magnets. The future work should also extend this solution for an orthotropic coil.

Appendix A

$$\chi_{11}(n)=\frac{-1}{(1-\nu)}\left[\left(\nu-\frac{1}{2}\right)\frac{n\pi}{L^2}I_0\left(\frac{n\pi}{L}a\right)-\frac{n^2\pi^2}{2L^3}aI_1\left(\frac{n\pi}{L}a\right)\right]$$

$$\chi_{12}(n)=\frac{-1}{(1-\nu)}\left[\left(\nu-\frac{1}{2}\right)\frac{n\pi}{L^2}K_0\left(\frac{n\pi}{L}a\right)-\frac{n^2\pi^2}{2L^3}aK_1\left(\frac{n\pi}{L}a\right)\right]$$

$$\chi_{13}(n)=\frac{-1}{(1-\nu)}\left[-\frac{n^3\pi^3}{L^4}I_0\left(\frac{n\pi}{L}a\right)+\frac{n^2\pi^2}{L^3}\frac{1}{a}I_1\left(\frac{n\pi}{L}a\right)\right]$$

$$\chi_{14}(n)=\frac{-1}{(1-\nu)}\left[-\frac{n^3\pi^3}{L^4}K_0\left(\frac{n\pi}{L}a\right)+\frac{n^2\pi^2}{L^3}\frac{1}{a}K_1\left(\frac{n\pi}{L}a\right)\right]$$

$$\chi_{21}(n)=\frac{-1}{(1-\nu)}\left[\left(\nu-\frac{1}{2}\right)\frac{n\pi}{L^2}I_0\left(\frac{n\pi}{L}b\right)-\frac{n^2\pi^2}{2L^3}bI_1\left(\frac{n\pi}{L}b\right)\right]$$

$$\chi_{22}(n)=\frac{-1}{(1-\nu)}\left[\left(\nu-\frac{1}{2}\right)\frac{n\pi}{L^2}K_0\left(\frac{n\pi}{L}b\right)-\frac{n^2\pi^2}{2L^3}bK_1\left(\frac{n\pi}{L}b\right)\right]$$

$$\chi_{23}(n)=\frac{-1}{(1-\nu)}\left[-\frac{n^3\pi^3}{L^4}I_0\left(\frac{n\pi}{L}b\right)+\frac{n^2\pi^2}{L^3}\frac{1}{b}I_1\left(\frac{n\pi}{L}b\right)\right]$$

$$\chi_{24}(n)=\frac{-1}{(1-\nu)}\left[-\frac{n^3\pi^3}{L^4}K_0\left(\frac{n\pi}{L}b\right)+\frac{n^2\pi^2}{L^3}\frac{1}{b}K_1\left(\frac{n\pi}{L}b\right)\right]$$

$$\chi_{31}(n)=-\left[\frac{n\pi}{L^2}I_1\left(\frac{n\pi}{L}a\right)+\frac{1}{(1-\nu)}\frac{n^2\pi^2}{2L^3}aI_0\left(\frac{n\pi}{L}a\right)\right]$$

$$\chi_{32}(n)=-\left[\frac{n\pi}{L^2}K_1\left(\frac{n\pi}{L}a\right)+\frac{1}{(1-\nu)}\frac{n^2\pi^2}{2L^3}aK_0\left(\frac{n\pi}{L}a\right)\right]$$

$$\chi_{33}(n)=-\left[\frac{n^3\pi^3}{L^4}\frac{1}{(1-\nu)}I_1\left(\frac{n\pi}{L}a\right)\right]$$

$$\chi_{34}(n)=-\left[\frac{n^3\pi^3}{L^4}\frac{1}{(1-\nu)}K_1\left(\frac{n\pi}{L}a\right)\right]$$

$$\chi_{41}(n)=-\left[\frac{n\pi}{L^2}I_1\left(\frac{n\pi}{L}b\right)+\frac{1}{(1-\nu)}\frac{n^2\pi^2}{2L^3}bI_0\left(\frac{n\pi}{L}b\right)\right]$$

$$\chi_{42}(n)=-\left[\frac{n\pi}{L^2}K_1\left(\frac{n\pi}{L}b\right)+\frac{1}{(1-\nu)}\frac{n^2\pi^2}{2L^3}bK_0\left(\frac{n\pi}{L}b\right)\right]$$

$$\chi_{43}(n)=-\left[\frac{n^3\pi^3}{L^4}\frac{1}{(1-\nu)}I_1\left(\frac{n\pi}{L}b\right)\right]$$

$$\chi_{44}(n)=-\left[\frac{n^3\pi^3}{L^4}\frac{1}{(1-\nu)}K_1\left(\frac{n\pi}{L}b\right)\right]$$

$$\Lambda_1(\zeta_{0i},n)=\frac{1}{(1-\nu)}\frac{\pi}{a^2}\frac{2\zeta_{0i}^2J_0(\zeta_{0i}a)J_0(\zeta_{0i}b)}{J_0^2(\zeta_{0i}a)-J_0^2(\zeta_{0i}b)}\{\zeta_{0i}\Omega_2(\zeta_{0i},n)+\omega(\zeta_{0i})[\zeta_{0i}\Omega_1(\zeta_{0i},n)+\Omega_2(\zeta_{0i},n)]\}$$

$$\Lambda_2(\zeta_{0i},n)=\frac{1}{(1-\nu)}\frac{\pi}{b^2}\frac{2\zeta_{0i}^2J_0^2(\zeta_{0i}a)}{J_0^2(\zeta_{0i}a)-J_0^2(\zeta_{0i}b)}\{\zeta_{0i}\Omega_2(\zeta_{0i},n)+\omega(\zeta_{0i})[\zeta_{0i}\Omega_1(\zeta_{0i},n)+\Omega_2(\zeta_{0i},n)]\}$$

$$\Lambda_3(\zeta_{0i},n)=\frac{1}{(1-\nu)}\frac{\pi}{a}\frac{2\zeta_{0i}^3J_0(\zeta_{0i}a)J_0(\zeta_{0i}b)}{J_0^2(\zeta_{0i}a)-J_0^2(\zeta_{0i}b)}\{-\zeta_{0i}\Omega_4(\zeta_{0i},n)+\omega(\zeta_{0i})[2\nu\Omega_4(\zeta_{0i},n)-\zeta_{0i}\Omega_3(\zeta_{0i},n)]\}$$

$$\Lambda_4(\zeta_{0i},n)=\frac{1}{(1-\nu)}\frac{\pi}{b}\frac{2\zeta_{0i}^3J_0^2(\zeta_{0i}a)}{J_0^2(\zeta_{0i}a)-J_0^2(\zeta_{0i}b)}\{-\zeta_{0i}\Omega_4(\zeta_{0i},n)+\omega(\zeta_{0i})[2\nu\Omega_4(\zeta_{0i},n)-\zeta_{0i}\Omega_3(\zeta_{0i},n)]\}$$

where

$$\Omega_1(\zeta_{0i},n)=\int_{-L}^{L}z\cos\frac{n\pi z}{L}\sinh(\zeta_{0i}z)dz$$
$$=\frac{2L^2(-1)^n}{(\zeta_{0i}^2L^2+n^2\pi^2)^2}[-(\zeta_{0i}^2L^2-n^2\pi^2)\sinh(\zeta_{0i}L)+(\zeta_{0i}^2L^2+n^2\pi^2)\zeta_{0i}L\cosh(\zeta_{0i}L)]$$

$$\Omega_2(\zeta_{0i},n)=\int_{-L}^{L}\cos\frac{n\pi z}{L}\cosh(\zeta_{0i}z)dz$$
$$=\frac{2\zeta_{0i}L^2(-1)^n}{(\zeta_{0i}^2L^2+n^2\pi^2)}\sinh(\zeta_{0i}L)$$

$$\Omega_3(\zeta_{0i},n)=\int_{-L}^{L}z\sin\frac{n\pi z}{L}\cosh(\zeta_{0i}z)dz$$
$$=\frac{2L^2(-1)^n}{(\zeta_{0i}^2L^2+n^2\pi^2)^2}[2n\pi L\zeta_{0i}\sinh(\zeta_{0i}L)-(\zeta_{0i}^2L^2+n^2\pi^2)n\pi\cosh(\zeta_{0i},L)]$$

$$\Omega_4(\zeta_{0i},n)=\int_{-L}^{L}\sin\frac{n\pi z}{L}\sinh(\zeta_{0i}z)dz$$
$$=\frac{-2n\pi L(-1)^n}{(\zeta_{0i}^2L^2+n^2\pi^2)}\sinh(\zeta_{0i}L).$$

Appendix B

$$\chi_{51}(\zeta_{0i},n)=(-1)^n\frac{n\pi}{L^2}\left[(2-\nu)\int_a^b rK_0(\zeta_{0i},r)I_0\left(\frac{n\pi}{L}r\right)dr+\frac{n\pi}{2L^2}\int_a^b r^2K_0(\zeta_{0i},r)I_1\left(\frac{n\pi}{L}r\right)dr\right]$$

$$\chi_{52}(\zeta_{0i},n)=(-1)^n\frac{n\pi}{L^2}\left[(2-\nu)\int_a^b rK_0(\zeta_{0i},r)K_0\left(\frac{n\pi}{L}r\right)dr+\frac{n\pi}{2L^2}\int_a^b r^2K_0(\zeta_{0i},r)K_1\left(\frac{n\pi}{L}r\right)dr\right]$$

$$\chi_{53}(\zeta_{0i},n)=(-1)^n\frac{n^3\pi^3}{L^4}\left[\int_a^b rK_0(\zeta_{0i},r)I_0\left(\frac{n\pi}{L}r\right)dr\right]$$

$$\chi_{54}(\zeta_{0i},n)=(-1)^n\frac{n^3\pi^3}{L^4}\left[\int_a^b rK_0(\zeta_{0i},r)K_0\left(\frac{n\pi}{L}r\right)dr\right]$$

$$\Lambda_5(\zeta_{0i})=2\zeta_{0i}^2[(1-2\nu)\omega(\zeta_{0i})\cosh(\zeta_{0i}L)-\zeta_{0i}\cosh(\zeta_{0i}L)-\zeta_{0i}\omega(\zeta_{0i})L\sinh(\zeta_{0i}L)]$$

$$\Gamma_5(\zeta_{0i},\zeta_{1i})=-(1-\nu)\left\{\frac{2}{\pi^2\zeta_{0i}^2}\left[1-\left(\frac{J_0(\zeta_{0i}b)}{J_0(\zeta_{0i}a)}\right)^2\right]\wp_5(\zeta_{0i})+\wp_6(\zeta_{1i})\int_a^b rK_0(\zeta_{0i},r)K_0^*(\zeta_{1i},r)dr\right\}$$

References

[1] Lontai, L. M., and Marston, P. G., 1995, ''A 100 Kilogauss Quasi-Continuous Cryogenic Solenoid—Part I,'' *Proceedings of the International Symposium on Magnet Technology*, Stanford University, CA, pp. 723–732.

[2] Arp, V., 1977, ''Stresses in Superconducting Solenoids,'' J. Appl. Phys., **48**, No. 5, pp. 2026–2036.

[3] Gray, W. H., and Ballou, J. K., 1977, ''Electromechanical Stress Analysis of Transversely Isotropic Solenoids,'' J. Appl. Phys., **48**, No. 7, pp. 3100–3109.

[4] Mitchell, N., and Mszanowski, U., 1992, ''Stress Analysis of Structurally Graded Long Solenoid Coils,'' IEEE Trans. Magn., **28**, No. 1, pp. 226–229.
[5] Mori, T., and Tanaka, K., 1973, ''Average Stress in Matrix and Average Elastic Energy of Materials With Misfitting Inclusions,'' Acta Metall., **21**, No. 5, pp. 571–574.
[6] Hasegawa, H., Lee, V., and Mura, T., 1991, ''Stress Fields Caused by a Circular Cylindrical Inclusion,'' ASME Winter Annual Meeting Atlanta, GA, Paper No. 91-WA/APM-32.
[7] Hu, K. X., and Chandra, A., 1993, ''Interactions Among General Systems of Cracks and Anticracks: An Integral Equation Approach,'' ASME J. Appl. Mech., **60**, No. 4, pp. 920–927.
[8] Kuo, C. H., and Keer, L. M., 1995, ''Three-Dimensional Analysis of Cracking in a Multilayered Composite,'' ASME J. Appl. Mech., **62**, No. 2, pp. 273–281.
[9] Noda, N., and Matsuo, T., 1995, ''Singular Integral Equation Method in Optimization of Stress-Relieving Hole: A New Approach Based on the Body Force Method,'' Int. J. Fract., **70**, No. 2, pp. 147–165.
[10] Markiewicz, W. D., Vaghar, M. R., Dixon, I. R., and Garmestani, H., 1994, ''Generalized Plane Strain Analysis of Solenoid Magnets,'' IEEE Trans. Magn., **30**, No. 4, pp. 2233–2236.
[11] Bobrov, E. S., 1984, ''Electrically Conducting Orthotropic Cylinder Shell in Axial and Radial Magnetic Field,'' *The Mechanical Behavior of Electromagnetic Solid Continua*, Elsevier, Amsterdam, pp. 407–413.
[12] Cox, A., Garmestani, H., Markiewicz, W. D., and Dixon, I. R., 1996, ''Power Series Stress Analysis of Solenoid Magnets,'' IEEE Trans. Magn., **32**, No. 4, pp. 3012–3015.
[13] Hasegawa, H., 1976, ''Axisymmetric Body Force Problems of an Elastic Half Space,'' Japan Soc. Mech. Eng., **19**, No. 137, pp. 1262–1269.
[14] Hasegawa, H., Lee, V., and Mura, T., 1992, ''Green's Functions for Axisymmetric Problems of Dissimilar Elastic Solids,'' ASME, Summer Mechanics and Materials Meeting, Tempe, AZ, pp. 1–9.
[15] Boresi, P., and Chong, K., 1987, *Elasticity in Engineering Mechanics*, Elsevier, New York.
[16] Sneddon, I. N., 1951, *Fourier Transforms*, McGraw-Hill, New York, pp. 82–91.

Static and Dynamic Characterization of Some Tensegrity Modules

H. Murakami[1]
Professor
Mem. ASME
e-mail: murakami@mae.uscd.edu

Y. Nishimura
Graduate Student

Department of Mechanical and
Aerospace Engineering,
University of California,
9500 Gilman Drive,
La Jolla, CA 92093-0411

A set of procedures was presented for characterizing static and dynamic response of tensegrity modules. The procedures were applied to two tensegrity modules: a six-bar spherical module and a two-stage cylindrical module with three bars at each stage. The singular value decomposition of the initial equilibrium matrix revealed prestress and infinitesimal mechanism modes. The prestress stiffening effect of infinitesimal mechanism modes was found to be isotropic at each node. In the initial quasi-static loading, infinitesimal mechanisms exhibited soft response. As the deformation advanced, the stiffness of tensegirty modules increased almost quadratically with infinitesimal mechanism amplitudes. Modal analyses revealed that the lowest modes were those of infinitesimal mechanism modes and their natural frequencies were an order of magnitude smaller than those of higher deformation modes. [DOI: 10.1115/1.1331058]

1 Introduction

Kenneth Snelson invented a cylindrical tensegrity tower in 1948 ([1]). By extending the concept of geodesics, Fuller developed spherical tensegrity modules ([2]). According to Marks and Fuller [2] and Pugh [3], tensegrity is a class of truss structures consisting of a continuous set of cables and discrete bars. Figures 1(*a*) and 1(*b*) illustrate, respectively, a six-bar spherical tensegrity module ([4]) and a two-stage cylindrical tensegrity module with three bars at each stage ([5]).

Aerospace engineers have adopted lightweight tensegrity structures as a new deployable structural concept. Motro [6], Furuya [7], and Hanaor [8,9] proposed tensegrity structures as deployable space structures. Skelton and Sultan [5] presented a smart structural system integrating tensegrity structures with modern control theory.

For a truss structure with n_E elements or members and n_N nodes or joints with n_C linearly independent displacement constraints, there are $n_V \equiv 3\ n_N - n_C$ unknown displacement components. In this paper it is assumed that each structure is constrained against rigid-body motion, $n_C \geqslant 6$. Let the element internal-force vector be denoted by $\mathbf{s}$, an $n_E \times 1$ column matrix, and the external nodal force vector by $\mathbf{f}$, an $n_V \times 1$ column matrix. The equilibrium equation at time t for quasi-static loading $\mathbf{f}(t)$ is expressed as

$$\mathbf{A}(t)\mathbf{s}(t) = \mathbf{f}(t), \tag{1}$$

where $\mathbf{A}(t)$ is an $n_V \times n_E$ matrix consisting of direction cosines of truss elements. (The equilibrium matrix $\mathbf{A}(t)$ will be defined in (7*a–c*).

Clark Maxwell [10] classified the stiffness of truss structures by using the difference between the element number n_E and the number of unknown displacement components n_V:

$$Mx \equiv n_E - n_V. \tag{2}$$

which is referred to as the Maxwell number in this paper. Maxwell observed that if $M_x > 0$, truss structures were redundant or statically indeterminate. If $M_x = 0$, trusses were statically determinate. If $M_x < 0$, trusses became (kinematically indeterminate) mechanisms. As exceptions to the above for $M_x \leqslant 0$, Maxwell noted structures which exhibited ''inferior order of stiffness,'' *i.e.*, stiffness is on the order of prestress instead of on the order of Young's modulus.

Calladine [11] observed that most of tensegrity structures introduced by Mark and Fuller [2] possess mechanisms that could be stiffened by prestressing to achieve ''infinitesimal mechanisms,'' or Maxwell's inferior order stiffness. By investigating the vector spaces associated with the initial equilibrium matrix $\mathbf{A}(0)$ in (1), Calladine [11] obtained the relationship between the number of prestress modes n_S and the number of mechanisms n_M as

$$n_S - n_M = Mx. \tag{3}$$

Tensegrity structures exist under prestressed configurations. These prestressable configurations must be found by solving initial equilibrium Eqs. (1) at $t = 0$ without external forces, $\mathbf{f}(0) = \mathbf{0}$. For a prestressable initial configuration, there is a unique prestress mode with tension in cables and compression in bars. Initial element forces must also be computed by solving initial equilibrium problems.

Cable networks are another class of truss structures and are similar to tensegrity structures in their need for finding initial geometry and prestress modes. Therefore, previous works on the mechanics of tensegrity structures have benefited from the large deformation analysis of prestressed cable networks by Argyris and Scharpf [12]. Based upon their theory, a pre-stressed initial configuration of a cable network has been computed by using a nonlinear finite element code based upon a guess of the initial geometry. In order to solve nonlinear equilibrium equations either the Newton method ([13]) or dynamic relaxation methods ([14,15]) have been employed. However, these numerical procedures for finding initial configurations are not useful for control applications. Analytical expressions can describe initial configurations by using an order of magnitude less number of parameters than what is used for numerical procedures.

The objective of the paper is to present a set of procedures for characterizing static and dynamic response of a class of tensegrity modules with $Mx \leqslant 0$. The procedures are illustrated by using two simple tensegrity modules shown in Figs. 1(*a*) and (*b*). The first step in characterizing tensegrity modules is to compute the Maxwell number (2) and observe Calladine's relation (3). For a class of tensegrity modules with ''discrete'' bars, the total number of nodes n_N is two times the number of bars. Both the spherical tensegrity module in Fig. 1(*a*) and the two-stage tensegrity module in Fig. 1(*b*) consist of six bars and 24 cables, i.e., $n_N = 12$ and $n_E = 30$. The Maxwell number $Mx \equiv n_E - (3n_N - n_C)$ with $n_C = 6$

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, July 2, 1999; final revision, May 6, 2000. Associate Editor: R. C. Benson. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

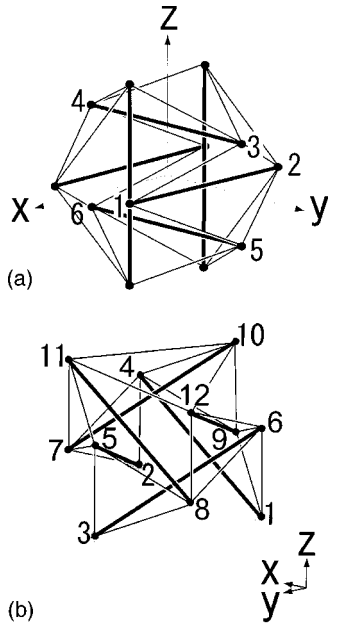


Fig. 1 (*a*) A six-bar spherical tensegrity module; (*b*) a two-stage tensegrity module with three bars at each stage

of the modules becomes $Mx=0$. Therefore, from Calladine's relation (3) they have the same number of prestress modes and mechanism modes, *i.e.*, $n_S=n_M$.

2 Summary of Nonlinear Equations of Motion

In this section, nonlinear equations of motion for elastic truss structures under large deformation are summarized ([16]). A motion of a truss structure with n_N nodes in Euclidean space $\mathbf{R}^3$ is described by the nodal coordinate $\mathbf{x}(t)$ with respect to an inertial Cartesian coordinate system $\{x_1,x_2,x_3\}$ and time t. By employing a finite element kinematical representation, nodes of the structure are identified by using both the global node numbers and local (or elemental) node numbers ([13,17]). The global node numbers, $1,2,\ldots,n_N$, identify the nodes of the entire truss structure, while local node numbers $\hat{1}$ and $\hat{2}$ identify the end nodes of each truss element, as illustrated in Figs. 1 and 2. Truss elements are identified by element numbers $1,2,\ldots,n_E$ in parentheses.

In a finite element description, the variation of the velocity field $\mathbf{v}$ in the axial θ_1-direction in each element is linearly interpolated by element nodal velocities, as illustrated in Fig. 2. Let the nodal velocity be denoted by $\mathbf{w}$, an $n_V\times 1$ column matrix, and the nodal velocity of element (e) by $\mathbf{w}^{(e)}$, a 6×1 column matrix. When a truss structure is subjected to the nodal force $\mathbf{f}(t)$ the principle of virtual velocity yields ([18])

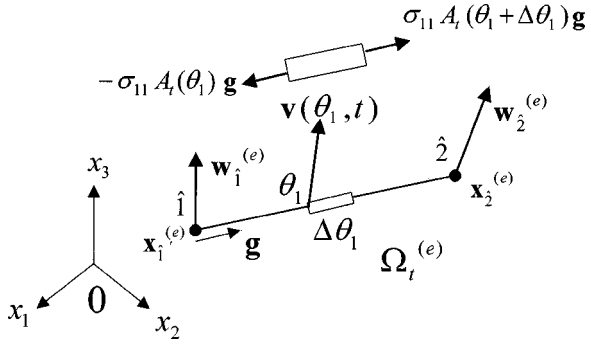


Fig. 2 Kinematics and kinetics of a deformed truss element

$$\sum_{e=1}^{n_E}\delta\mathbf{w}^{(e)}\cdot[\mathbf{M}^{(e)}\dot{\mathbf{w}}^{(e)}+\mathbf{p}^{(e)}]+\delta\mathbf{w}\cdot\mathbf{f}=0. \tag{4a}$$

In (4*a*) $\delta\mathbf{w}$ denotes virtual nodal velocity, $\dot{\mathbf{w}}$ is the nodal acceleration, $\mathbf{M}^{(e)}$ and $\mathbf{p}^{(e)}$ denote, respectively, the element mass matrix and the element internal force vector due to the Cauchy axial normal-stress σ_{11}:

$$\mathbf{M}^{(e)}\equiv\frac{1}{6}(m_t l_t)^{(e)}\begin{bmatrix}2\mathbf{I}_3 & \mathbf{I}_3\\ \mathbf{I}_3 & 2\mathbf{I}_3\end{bmatrix}, \tag{4b}$$

$$\mathbf{p}^{(e)}\equiv\begin{Bmatrix}-\mathbf{g}\\ \mathbf{g}\end{Bmatrix}^{(e)}(\sigma_{11}A_t)^{(e)}, \tag{4c}$$

where m_t, l_t, and A_t are, respectively, the mass per unit axial length, the element length, and the area of cross section at time t, $\mathbf{I}_3$ denotes the 3×3 identity matrix, and $\mathbf{g}^{(e)}(t)$ is the current direction cosine of element (e).

A relationship between the global and local (elemental) velocity components for element (e) is established by a Boolean map $\mathbf{L}g^{(e)}$, an $6\times n_V$ matrix, between the global nodal velocity $\mathbf{w}$, an $n_V\times 1$ column matrix, and the nodal velocity $\mathbf{w}^{(e)}$ of element (e):

$$\mathbf{w}^{(e)}=\mathbf{L}g^{(e)}\mathbf{w}, \tag{5a}$$

$$\delta\mathbf{w}^{(e)}=\mathbf{L}g^{(e)}\delta\mathbf{w}. \tag{5b}$$

By assembling (4*a*) for global degrees-of-freedom by using (5*a*, 5*b*), the equations of motion for the current configuration are obtained (see for example [17]):

$$\mathbf{M}\dot{\mathbf{w}}+\mathbf{p}(t)-\mathbf{f}(t)=\mathbf{0}, \tag{6a}$$

where $\mathbf{M}$ and $\mathbf{p}$ are the global mass matrix and the global internal force vector

$$\mathbf{M}\equiv\sum_{e=1}^{n_E}\mathbf{L}g^{(e)^T}\mathbf{M}^{(e)}\mathbf{L}g^{(e)}, \tag{6b}$$

$$\mathbf{p}\equiv\sum_{e=1}^{n_E}\mathbf{L}g^{(e)^T}\mathbf{p}^{(e)}. \tag{6c}$$

Let $\mathbf{s}(t)$ denote the $n_E\times 1$ column matrix of the element force vector whose eth element is the axial force $(\sigma_{11}A_t)^{(e)}$. From (4*c*) and (6*c*), the internal force vector $\mathbf{p}$ is expressed as

$$\mathbf{p}(t)=\mathbf{A}(t)\mathbf{s}(t), \tag{7a}$$

where

$$\mathbf{A}\equiv[\mathbf{a}_1\ \ \mathbf{a}_2\cdots\mathbf{a}_{n_E}], \tag{7b}$$

$$\mathbf{a}_e\equiv\mathbf{L}g^{(e)^T}\begin{Bmatrix}-\mathbf{g}\\ \mathbf{g}\end{Bmatrix}^{(e)}. \tag{7c}$$

For initial equilibrium analyses, Eqs. (6*a*) and (7*a*) at a natural configuration, denoted by time $t=0$, are utilized:

$$\mathbf{A}(0)\mathbf{s}(0)=\mathbf{f}(0), \tag{8}$$

in which the initial direction cosine $\mathbf{g}^{(e)}(0)$ is used in (7*c*). By applying the singular value decomposition to $\mathbf{A}(0)$, prestress and infinitesimal mechanism modes are computed by Murakami [19].

In order to investigate the stiffening effect by a prestress mode, natural frequencies, and corresponding mode shapes, the equations of motion (6*a*) are linearized at a prestressed configuration, denoted also by $t=0$. Let Young's modulus, the area of cross section, the element length, and the second Piola-Kirchhoff stress at the prestressed state be denoted, respectively, by Y_0, A_0, l_0, and $S_{11}(0)(=\sigma_{11}(0))$. In terms of the nodal displacement vector $\mathbf{d}$ and the nodal acceleration vector $\ddot{\mathbf{d}}$, the linearized equations of motion become

$$\mathbf{M}_L\ddot{\mathbf{d}}+\mathbf{K}_T\mathbf{d}=\mathbf{f}(t), \tag{9a}$$

where the symmetric tangent stiffness matrix $\mathbf{K}_T$ is decomposed into the initial stiffness $\mathbf{K}_0$, employed for small deformation truss analyses with $\mathbf{G}^{(e)} \equiv \mathbf{g}^{(e)}(0)$, and the prestress stiffness $\mathbf{K}_s$ as

$$\mathbf{K}_T = \mathbf{K}_0 + \mathbf{K}_s, \tag{9b}$$

in which

$$\mathbf{K}_0 \equiv \sum_{e=1}^{n_E} \mathbf{L}g^{(e)T} \mathbf{K}_0^{(e)} \mathbf{L}g^{(e)}, \tag{9c}$$

$$\mathbf{K}_s \equiv \sum_{e=1}^{n_E} \mathbf{L}g^{(e)T} \mathbf{K}_s^{(e)} \mathbf{L}g^{(e)}, \tag{9d}$$

$$\mathbf{K}_0^{(e)} \equiv \left(\frac{Y_0 A_0}{l_0} \right)^{(e)} \begin{bmatrix} \mathbf{G}\mathbf{G}^T & -\mathbf{G}\mathbf{G}^T \\ -\mathbf{G}\mathbf{G}^T & \mathbf{G}\mathbf{G}^T \end{bmatrix}^{(e)}, \tag{9e}$$

$$\mathbf{K}_s^{(e)} \equiv \left(\frac{S_{11}(0) A_0}{l_0} \right)^{(e)} \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix}. \tag{9f}$$

It is observed that the prestress stiffening in (9*f*) is ''isotropic'' at each node. (In large deformation analyses of prestressed networks, Argyris and Scharpf [12] predicted isotropic stiffening due to prestresses.) In the sequel, the above equations will be utilized to characterize static and dynamic response of tensegrity modules.

3 Initial Shape Finding

The initial geometry of tensegrity modules was originally found by the ingenuity of pioneers, such as Snelson and Fuller. Element lengths and nodal coordinates were later justified analytically by considering initial equilibrium Eq. (8) with $\mathbf{f}(0) = \mathbf{0}$. An obvious condition for tensegrity modules is the existence of nontrivial prestress modes $\mathbf{s}(0)$ with tension in cables and compression in bars. This condition is referred to as the ''tensegrity condition.'' Tarni [20] analytically constructed equilibrium equations and obtained the tensegrity conditions by using local coordinate systems for cyclic cylindrical truss structures. By extremizing the length of a single family of cable elements, for example vertical cables, and prescribing remaining geometrical parameters, Tobie [21] found the twist angle for regular cylindrical tensegrity modules. Tobie's work was introduced by Kenner [4] in his book. The equivalence between Tobie's and Tarnai's conditions was proven by Murakami [19].

At this moment, there are two methods available for finding element lengths and nodal coordinates of existing regular tensegrity modules: (i) by extremizing the length of a single family of cable elements for a prescribed set of geometrical parameters and (ii) by analytically solving reduced equilibrium equations. In this section, the former method is applied to a six-bar spherical tensegrity module in Fig. 1(*a*). The latter method is applied to a two-stage cylindrical tensegrity module in Fig. 1(*b*).

A Cartesian coordinate system $\{x, y, z\}$ is selected with the origin at the center of the sphere which circumscribes all the nodes 1–12. Due to the spherical symmetry, it suffices to determine the coordinates of nodes 1, 2, and 3, illustrated in Fig. 3. The nodal coordinates are completely determined in terms of the separation h between a pair of parallel bars with length b as: $1(b/2,h/2,0), 2(-b/2,h/2,0), 3(0,b/2,h/2)$. Let the length of cable elements be denoted by l. For a prescribed bar length b, the cable length of the element connecting nodes 1 and 3 becomes a function of the parameter h:

$$l(h) = \frac{1}{2} \sqrt{b^2 + (b-h)^2 + h^2}. \tag{10}$$

Spherical tensegrity modules built with arbitrary h collapse indicating that in $\mathbf{A}(0)$ all columns are linearly independent. Spherical tensegrity modules can exist without external supports if the columns of $\mathbf{A}(0)$ become linearly dependent and Eq. (8) with $\mathbf{f}(0) = \mathbf{0}$ has a nontrivial prestress mode $\mathbf{s}(0)$.

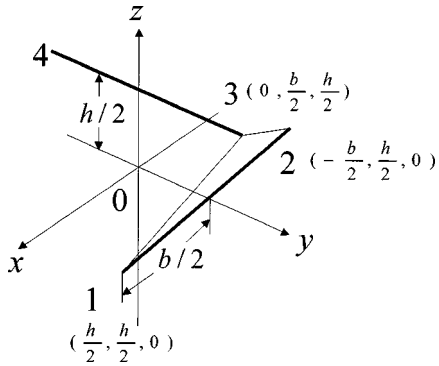


Fig. 3 Nodal coordinates of the six-bar spherical tensegrity module

By extremizing the cable length with respect to h, *i.e.*, $dl/dh = 0$, one finds $h = b/2$. Since the elements of the initial equilibrium matrix $\mathbf{A}(0) = [a_{ij}]$ are functions of $l(h)$, $dl/dh = 0$ is equivalent to satisfying

$$\hat{Q}(h) \equiv \det(\mathbf{A}^T(0)\mathbf{A}(0)) = 0. \tag{11a}$$

It can be easily shown that $\mathbf{A}^T(0)\mathbf{A}(0)$ is positive semi-definite, $\hat{Q}(h) \geqslant 0$. Further, the rank of $\mathbf{A}(0)$ is n_E, except for the cases when the tensegrity condition is satisfied, *i.e.*, $n_S = 1$. Therefore, $\hat{Q}(h)$ takes the minimum value, $\hat{Q}(h) = 0$, if and only if $d\hat{Q}/dh = 0$:

$$\frac{d\hat{Q}}{dh} = \left(\sum_{i,j=1}^{n_E} \mathrm{Cof}(c_{ij}) \frac{dc_{ij}}{dl} \right) \frac{dl}{dh} = 0, \tag{11b}$$

where $\mathrm{Cof}(c_{ij})$ denotes the cofactor of $c_{ij} \equiv a_{ki} a_{kj}$ with summation over $k = 1, \ldots, n_V$. For the tensegrity module, $r_A \equiv \mathrm{rank}\, \mathbf{A}(0) = n_E - 1$ and $n_S = \dim(\mathrm{null}\mathbf{A}^T(0)\mathbf{A}(0)) = 1$, there is at least one nonzero cofactor of $[c_{ij}]$. Therefore, $d\hat{Q}/dh = 0$ if and only if $dl(h)/dh = 0$.

By using $h = b/2$ and computing both nodal coordinates and the direction cosines of the bar and cables connected to the node, one can write equilibrium Eqs. (8) at node 3. The prestress mode $\mathbf{s}(0)$ relates element forces in the bar s_b and cable s_c as follows:

$$s_c = -\frac{1}{\sqrt{6}} s_b. \tag{12}$$

A two-stage cylindrical tensegrity module with three bars at each stage is shown in Fig. 1(*b*). Skelton and Sultan [5] obtained the equations of motion and tensegrity conditions for modules consisting of ''rigid'' bars and cables with negligible mass. They demonstrated unprecedented possibilities offered by controlled smart tensegrity structures. The structure in Fig. 1(*b*) is built by stacking two three-bar cylindrical modules ([4]). Figure 4(*a*) illustrates the connections of bars and cables by using a developed connectivity diagram of the two-stage tensegrity ([3,22]). In the figure, bold lines indicate bars and thin lines indicate top, base, saddle, and vertical cables. Dashed lines represent diagonal cables. In order to define nodal coordinates, a Cartesian coordinate system $\{x, y, z\}$ is selected with the origin at the center of the base equilateral triangle defined by nodes 1, 2, and 3, also denoted by 1–3. The x, y-plane coincides with the plane of the base triangle 1–3, and cable (1, 2) is parallel to the x-axis. The z-axis connects the centers of the base and top equilateral triangles. The top view of a two-stage tensegrity is illustrated in Fig. 4(*b*) by using equilateral triangles formed by nodes at equal elevations. Let the radii of the circles circumscribing the base triangle 1–3 and the top triangle 10–12 be r_0. The bars of the base module consisting of nodes 1–6 connect nodes (1, 4), (2, 5), (3, 6) and are twisted in the counterclockwise direction by $2\pi/3 + \alpha$. The pa-

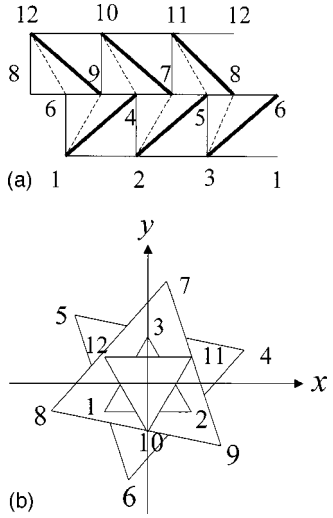


Fig. 4 (*a*) A developed connection diagram of the two-stage tensegrity module; (*b*) a top view of equilateral triangles of the two-stage tensegrity module

rameter α is referred to as the additional twist angle ([4]). The top module 7–12 is an upside-down version of the base module 1–6. The bars of the top module connecting nodes (7, 10), (8, 11), (9, 12) are twisted in the clockwise direction by $2\pi/3+\alpha$. The top module is stacked on the saddle cables of the base module in such a way that nodes 7, 8, and 9, respectively, rest in the middle of saddle cables (4, 5), (5, 6), and (6, 4). The diagonal cables connect nodes (1, 9), (2, 7), (3, 8), (12, 6), (10, 4), and (11, 5).

All nodal coordinates can be analytically expressed for the prescribed bar length b, the radius r_0 of the top and base circles, the radius r_m of the circumscribing circles of triangles 4–6 and 7–9, the additional twist angle α, and the overlap ratio γ between the triangles 4–6 and 7–9 in elevation. If the z-coordinate of the triangle 4–6 is denoted by $h\,(b,\alpha,r_0,r_m)$, the z-coordinate of the triangle 7–9 becomes $(1-\gamma)h$. (The nodal coordinates and element lengths are presented in the Appendix.) The previous extremization method does not work here since there are two unknown cable lengths. A remaining method is to construct the initial equilibrium matrix of (8) by analytically calculating the initial direction cosine $\mathbf{G}$ of each element. By virtue of the cyclic symmetry, it suffices to examine the equilibrium equations at two representative nodes 1 and 9 by incorporating the constraint that the elements in the same family possess the same prestresses. Let the element forces due to prestresses of bars, top and base cables, vertical cables, and diagonal cables be denoted by s_b, s_0, s_v, and s_d, respectively. Saddle cables are divided into two groups: group I consisting of cables (6, 9), (4, 7), (5, 8) and group II consisting of (9, 4), (7, 5) and (8, 6). Their element forces are denoted respectively by s_1 and s_2. With the above simplifications, the equilibrium matrix of (8) reduces to a 6×6 matrix. By setting the determinant of $\mathbf{A}(0)$ to zero, the following characteristic equation in terms of r_m/r_0, α, and γ is obtained:

$$\frac{r_m}{r_0}\sin\left(\frac{\pi}{6}-\alpha\right)=\frac{\gamma(\gamma+1)}{2(\gamma^2-\gamma+1)}. \tag{13}$$

A reduced set of initial equilibrium equations is presented in the Appendix. The above analytical result agrees with the numerical examples presented by Skelton and Sultan [5] and Sultan [23]. In order to find a set of nodal coordinates, one has to prescribe two parameters out of $\{r_m/r_0,\alpha,\gamma\}$ and solve (13) for the remaining parameter. For a prescribed γ, Eq. (13) describes a line in the x, y-plane. Let the lengths of top and base, vertical, saddle, and diagonal cables be denoted by l_0, l_v, l_s, and l_d, respectively. The cable lengths were computed and shown in the Appendix.

After finding the initial configuration, a nontrivial prestress mode $\mathbf{s}(0)$ can be computed from the reduced equilibrium equations as follows:

$$\frac{s_0}{l_0}=-\frac{s_b}{bD'}\left\{-\frac{\gamma}{\sqrt{3}}\sin\left(\frac{\pi}{6}-\alpha\right)+\frac{1+\gamma}{2\sqrt{3}}\frac{r_m}{r_0}\right\}, \tag{14a}$$

$$\frac{s_v}{l_v}=-\frac{s_b}{bD'}\left\{\gamma\sin\left(\frac{\pi}{3}-\alpha\right)+\sin\alpha\right\}, \tag{14b}$$

$$\frac{s_1}{l_s}=\frac{s_2}{l_s}=-\frac{s_b}{bD'}\frac{\sqrt{3}(1-\gamma)}{\gamma}\sin\left(\frac{\pi}{6}-\alpha\right), \tag{14c}$$

$$\frac{s_d}{l_d}=-\frac{s_b}{bD'}\sqrt{3}\sin\left(\frac{\pi}{6}-\alpha\right), \tag{14d}$$

where

$$D'\equiv\sin\left(\frac{\pi}{3}-\alpha\right)+\gamma\sin\alpha. \tag{14e}$$

The parameters that satisfy (13) yield an admissible prestress mode with tension in cables and compression in bars.

A key step in obtaining a reduced set of equilibrium equations is to incorporate the cyclic symmetry conditions and to impose the constraint of the same element forces due to prestresses on the same family of elements.

In the initial shape finding, Eq. (11*a*) is imposed on the initial geometry for the existence of a prestress mode. It is instructive to examine the quadratic form associated with $\mathbf{A}^T(0)\mathbf{A}(0)$ in the vector space $\mathbf{R}^{n_E}$ of element internal forces $\mathbf{s}(0)$. By using (8), the quadratic form $\mathbf{s}^T\mathbf{A}^T(0)\mathbf{A}(0)\mathbf{s}$ is identified to be $\mathbf{f}\cdot\mathbf{f}$ the square length of the nodal force vector in $\mathbf{R}^{n_V}$ and is positive semidefinite. The condition (11*a*) only assures the existence of a prestress mode. Therefore, after computing a prestress mode, one has to check the admissibility of the prestress mode with tension in cables and compression in bars.

4 Infinitesimal Mechanism and Prestress Modes

Once an initial geometry is found, infinitesimal mechanism and prestress modes can be numerically obtained by performing the singular value decomposition on the ''full'' initial equilibrium matrix $\mathbf{A}(0)$ in (8) ([19]). The matrix $\mathbf{A}(0)$ defines a linear transformation from the vector space $\mathbf{R}^{n_E}$ of the element internal forces $\mathbf{s}$ and elongations $\mathbf{e}$ to the vector space $\mathbf{R}^{n_V}$ of the nodal forces $\mathbf{f}$ and displacements $\mathbf{d}$.

For small deformations, Clapeyron's theorem (for example, Sokolnikoff [24]) states that the work done by surface traction and body forces acting through the displacements from the natural configuration to the deformed configuration is equal to twice the strain energy of the body if it obeys Hooke's law. The theorem yields at the natural configuration, denoted by $t=0$:

$$\langle\mathbf{f},\mathbf{d}\rangle_{n_V}=\langle\mathbf{s},\mathbf{e}\rangle_{n_E}. \tag{15}$$

By substituting the linear transformation (8) into (15), one obtains the adjoint transformation of $\mathbf{A}(0)$ from $\mathbf{R}^{n_V}$ to $\mathbf{R}^{n_E}$ as (see, for example, Naylor and Sell [25]):

$$\langle\mathbf{A}(0)\mathbf{s},\mathbf{d}\rangle_{n_V}\equiv\langle\mathbf{s},\mathbf{A}^T(0)\mathbf{d}\rangle_{n_E}. \tag{16}$$

This definition of the element elongation $\mathbf{e}$ agrees, as it should, with that obtained from the small deformation finite element analyses:

$$\mathbf{e}=\mathbf{A}^T(0)\mathbf{d}. \tag{17}$$

By investigating the dimensions of $\mathbf{R}^{n_V}$ and $\mathbf{R}^{n_E}$, Calladine [11] observed

$$n_E=n_S+r_A, \tag{18a}$$

$$n_V = n_M + r_A, \tag{18b}$$

where $r_A \equiv \text{rank}\mathbf{A}(0), n_S \equiv \dim(\text{null}\mathbf{A}(0))$, the number of prestress modes, and $n_M \equiv \dim(\text{null}\mathbf{A}^T(0))$, the number of infinitesimal mechanism modes. From (2) and (18*a*, 18*b*), Calladine's relation (3) was obtained. A physical interpretation of the null and range spaces of $\mathbf{A}(0)$ and $\mathbf{A}^T(0)$ was presented by Pellegrino and Calladine [26]. Prestress modes span the null space of $\mathbf{A}(0)$, while mechanism modes span the null space of $\mathbf{A}^T(0)$.

In order to obtain base vectors of $\mathbf{R}^{n_V}$ and $\mathbf{R}^{n_E}$, the singular value decomposition theorem (for example, Noble and Daniel [27]) is utilized. It is known that $n \times n$ symmetric matrices have real eigenvalues and orthonormal sets of n eigenvectors. Therefore, $\mathbf{A}^T(0)\mathbf{A}(0)$ and $\mathbf{A}(0)\mathbf{A}^T(0)$ have, respectively, n_E and n_V eigenpairs. Further, $\det(\mathbf{A}^T(0)\mathbf{A}(0)) \geqslant 0$ and $\det(\mathbf{A}(0)\mathbf{A}^T(0)) \geqslant 0$ indicate that the eigenvalues are positive semi-definite. There are $r_A (\equiv \text{rank}\mathbf{A}(0))$ positive eigenvalues: $\sigma_1^2 \geqslant \sigma_2^2 \geqslant \ldots \geqslant \sigma_{r_A}^2 > 0$ where positive σ's are called the singular values of $\mathbf{A}(0)$. The ordered eigenpairs in the decreasing singular values satisfy

$$\mathbf{A}^T(0)\mathbf{A}(0)\mathbf{s}_i = \sigma_i^2 \mathbf{s}_i, \quad i = 1,2,\cdots,n_E, \tag{19a}$$

$$\mathbf{A}(0)\mathbf{A}^T(0)\mathbf{d}_j = \sigma_j^2 \mathbf{s}_j, \quad j = 1,2,\cdots,n_V. \tag{19b}$$

In $\mathbf{R}^{n_E}, \tilde{\mathbf{S}} \equiv [\mathbf{s}_1 \mathbf{s}_2 \ldots \mathbf{s}_{n_E}]$ is an orthonormal basis, while $\tilde{\mathbf{D}} \equiv [\mathbf{d}_1 \mathbf{d}_2 \ldots \mathbf{d}_{n_V}]$ is an orthonormal basis in $\mathbf{R}^{n_V}$. By using the $n_V \times n_V$ orthonormal matrix $\tilde{\mathbf{D}}$ and the $n_E \times n_E$ orthonormal matrix $\tilde{\mathbf{S}}$, $\mathbf{A}(0)$ is decomposed into

$$\mathbf{A}(0) = \tilde{\mathbf{D}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{S}}, \tag{20a}$$

where $\tilde{\mathbf{\Sigma}}$ is an $n_V \times n_E$ matrix with singular values on the diagonals of the first r_A rows:

$$\tilde{\mathbf{\Sigma}} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \sigma_{r_A} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{20b}$$

The eigenvectors $\mathbf{s}_j$ corresponding to zero eigenvalues of (19*a*) are prestress modes and span the null space of $\mathbf{A}(0)$, while eigenvectors $\mathbf{d}_j$ corresponding to zero eigenvalues of (19*b*) are infinitesimal mechanism modes and span the null space of $\mathbf{A}^T(0)$. The eigenproblems of (19*a*, 19*b*) can easily be solved by using either Jacobi's method ([13]) or Lanczos' method ([17]). For static characterization of tensegrity modules, only eigenvectors corresponding to zero singular values are required. Since $Mx = 0$ and $n_S = 1$ for the tensegrity modules in Figs. 1(*a*) and (*b*), Calladine's relation (3) yields $n_M = 1$. Therefore, $\mathbf{s}_{n_E}$ is a prestress mode and $\mathbf{d}_{n_V}$ becomes an infinitesimal mechanism mode. One can validate the numerical values in $\mathbf{s}_{n_E}$ by comparing them with the analytical prestress mode in (14*a*)–(14*d*).

Figures 5(*a*) and (*b*) illustrate the infinitesimal mechanism modes of the spherical tensegrity module and the two-stage tensegrity module, respectively. In the figures, dashed lines show the undeformed configurations. The spherical module in Fig. 5(*a*) exhibits a spherically symmetric, radial deformation mode. The two-stage module in Fig. 5(*b*) shows an axially deforming mode with the equilateral triangles 4–6 and 7–9, shown in Fig. 4(*b*), rotating in the same direction. This implies that top bars and base bars are rotating in opposite directions with respect to the z-axis. (A snap shot of Fig. 5(*b*) indicate an axially compressed state.) In the next section, it will be shown that these infinitesimal mechanism modes with zero element elongation exhibit a two-orders-of-magnitude softer response than the deformation modes with nonzero elongation in (17).

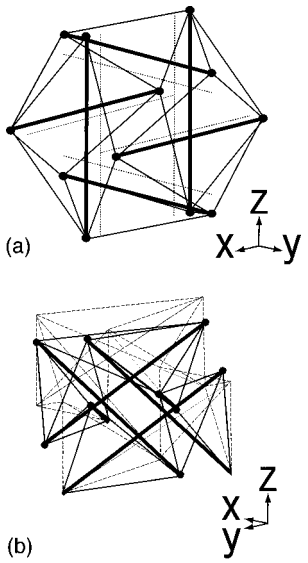


Fig. 5 (*a*) **An infinitesimal mechanism mode of the spherical tensegrity module;** (*b*) **an infinitesimal mechanism mode of the two-stage tensegrity module**

5 Stiffness of Prestressed Tensegrity Modules

For a prestressed tensegrity module, static equilibrium equations for small deformations are obtained from (9*a*) with $\ddot{\mathbf{d}} = \mathbf{f} = \mathbf{0}$. The tangent stiffness $\mathbf{K}_T$ was decomposed into the initial stiffness $\mathbf{K}_0$, employed for small deformation truss analyses, and the stiffness $\mathbf{K}_s$ induced by prestresses, as shown in (9*b*). For the two tensegrity modules in Figs. 1(*a*) and (*b*), it can be numerically shown that $\det(\mathbf{K}_T) > 0$. Since $\mathbf{e} = \mathbf{A}^T(0)\mathbf{d}_{n_V} = \mathbf{0}$ and $\mathbf{K}_0 \mathbf{d}_{n_V} = \mathbf{0}$, the stiffness of an infinitesimal mechanism mode, $\mathbf{d}_{n_V}$ in $\tilde{\mathbf{D}}$ of (20*a*), is induced only by $\mathbf{K}_s$. Further, Eq. (9*f*) reveals that the prestress stiffness at each node is ''isotropic'' since the nodal stiffness is expressed by identity matrices ([12,16]). It is noted that the assumption of ''moderate rotation'' is economical and popularly employed for truss structures. However, due to an inconsistent linearization, the assumption of moderate rotation incorrectly predicts ''anisotropic'' prestress stiffening instead of ''isotropic'' stiffening for tensegrity structures with infinitesimal mechanism modes.

6 Modal Analyses

Dynamic characterization of tensegrity modules involves modal analysis at a ''prestressed configuration.'' Consider a small harmonic motion of the form $\mathbf{d} = \tilde{\mathbf{d}} \exp(i\omega t)$ in (9*a*) where $\tilde{\mathbf{d}}$ is the amplitude and ω is the angular frequency. The standard finite element eigenproblem is obtained from (9*a*) with $\mathbf{f} = \mathbf{0}$ as

$$\mathbf{K}_T \tilde{\mathbf{d}} = \omega^2 \mathbf{M} \tilde{\mathbf{d}}. \tag{21}$$

The above eigenproblem can be solved by either using the Lanczos method ([17]) or the subspace iteration method ([13]).

For the numerical examples, steel bars and cables with Young's modulus $Y_0 = 200\,\text{GPa}$ and mass density $\rho = 7860\,\text{kg/m}^3$ are considered. The diameters of bars and cables are, respectively, $10^{-2}\,\text{m}$ and $10^{-3}\,\text{m}$. The first three natural frequencies of a spherical tensegrity module with $b = 2m$ are shown in Table 1 for increasing prestresses. The natural frequencies of a two-stage tensegrity module with the geometry $r_0 = r_m = h = 1m$ and $\gamma = 1/2$ are shown in Table 2. The first natural frequencies are an

Table 1 Natural frequencies of the spherical tensegrity module for various prestress levels

Prestress of bars [MPa]	1st mode [Hz]	2nd mode [Hz]	3rd mode [Hz]
0.01	0.2056	10.62	16.22
0.1	0.6502	10.62	16.22
1.0	2.056	10.63	16.23
2.0	2.908	10.63	16.23
3.0	3.561	10.63	16.23
4.0	4.112	10.64	16.24

Table 2 Natural frequencies of the two-stage tensegrity module for various prestress levels

Prestress of bars [MPa]	1st mode [Hz]	2nd mode [Hz]	3rd mode [Hz]
0.01	0.1819	20.93	20.93
0.10	0.5753	20.94	20.94
1.00	1.819	20.96	20.96
2.00	2.571	20.99	20.99
3.00	3.149	21.02	21.02
4.00	3.635	21.05	21.05

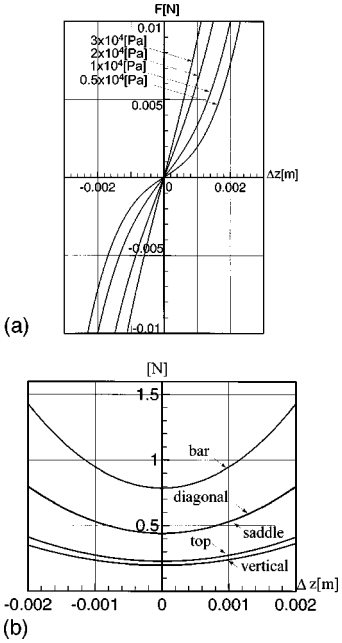


Fig. 6 (*a*) Load-displacement relation of the two-stage tensegrity module; (*b*) element forces—load-point displacement relation of the two-stage tensegrity module

order of magnitude smaller than those of higher modes. In the two modules, the first modes are indistinguishable from the infinitesimal mechanism modes, illustrated in Figs. 5(*a*) and (*b*). As shown by (9*e*) and (9*f*), the stiffness of the first modes is on the order of prestress, while that of the higher order modes is on the order of Young's modulus.

7 Nonlinear Stiffening Effect and Critical Loads of Bars

In this section, a quasi-static load-displacement relationship is investigated by using the updated Lagrangian finite element code which solves (6*a*). Newton's method was used with the linearized equation at each load increment. (Equation (9*a*) was linearized at $t=0$.) Figure 6(*a*) illustrates the vertical load and load-point displacement relation of the two-stage tensegrity module. The same vertical load is applied in the *z*-direction at each top node. The solid lines are the predictions of the finite element code at various initial prestress amplitudes. The initial tangent stiffness near the origin of Fig. 6(*a*) increases linearly with increasing prestress, as predicted by (9*f*). Further, the figure illustrates that the linear range is extremely small. If bars do not buckle, the load-displacement relation exhibits hardening response. As the load increases, the asymptotic stiffness converges to that obtained from (9e) with the direction cosines computed for the current deformed configuration. For the loading shown in Fig. 6(*a*) the element forces are plotted in Fig. 6(*b*) as a function of the load-point displacement. The cable tensile forces quadratically increase without slacking with the vertical load in both tension and compression. In the figure, the absolute value of bar compressive force is plotted. Skelton and Adhikari [28] first reported a hardening-type axial load-displacement relation for a two-stage tensegrity module. In addition, they reported cable slacking for the bending deformation of a two-stage tensegrity.

For tensegrity modules with infinitesimal mechanisms, the load-displacement relation is characterized by a nonlinear hard spring. Due to the hardening response of tensegrity modules, the necessary design consideration is to prevent both tensile failure of cables and buckling of bars. The buckling of bars is a bifurcation type, and a post-buckling behavior is known to be ''imperfection sensitive'' ([29,30]). Therefore, a critical load of bars should be determined by either conducting experiments or performing nonlinear finite element analyses of column buckling by imposing axial compressive forces predicted by the updated-Lagrangian truss finite element code. When the sum of the element force increment and the initial internal force due to prestresses reaches a critical load P_{cr} of the bar, buckling could take place. It is noted that the response of tensegrity modules with $Mx>0$ do not exhibit the hardening response shown in Figs. 6(*a*) and 6(*b*). For example, double-layer tensegrity grids with $Mx>0$ exhibit ''linear'' response until some bars buckle or some cables slack ([8,31]).

8 Tensegrity Configuration Spaces

In constructing tensegrity structures, it is not possible to build them without initial geometrical imperfections. Bars and cables may not have the exact intended lengths computed from the initial equilibrium analyses. Nodes may not be placed precisely at specified positions. Further, designers may wish to move nodes to neighboring locations to improve structural functions or architectural appearances. Tensegrity structures with a Maxwell number $Mx\leqslant 0$ collapse with minute disturbances if the tensegrity conditions are violated. The initial geometry in $\mathbf{R}^{n_V}$ must satisfy the characteristic Eq. (11*a*) that assures the existence of a prestress mode $\mathbf{s}_{n_E}$ in the element force vector space $\mathbf{R}^{n_E}$. In the configuration space $\mathbf{R}^{n_V}$, Eq. (11*a*) defines a hypersurface and imposes a holonomic constraint with respect to adjustable nodal coordinates. For the spherical tensegrity module, the hypersurface at the neighborhood of node 5 is drawn in Fig. 7 by fixing remaining nodal coordinates. (The dimension of the module is the same as that presented in Section 6.) The ideal position of node 5 is indicated by the base point of the grad $\hat{Q}$ vector. The shaded hypersurface indicates admissible nodal positions of node 5. Figure 8 illustrates the hypersurfaces for node 9 and node 10 of the two-stage cylindrical tensegrity module by fixing the remaining nodal coordinates. For example, the hypersurface of node 9 was computed by fixing all remaining nodal coordinates, including those of node 10, at the ideal initial configuration. The ideal positions of node 9 and 10 are indicated in Fig. 8 by the base positions of grad $\hat{Q}$ vectors. Figures 7 and 8 illustrate the existence of connected admissible nodal positions in the neighborhood of ideal nodal positions. The figures also indicate the prohibited directions indicated by grad $\hat{Q}$ normal to the hypersurface (11*a*). Moving a node to a position connected to an ideal or current nodal position on the admissible

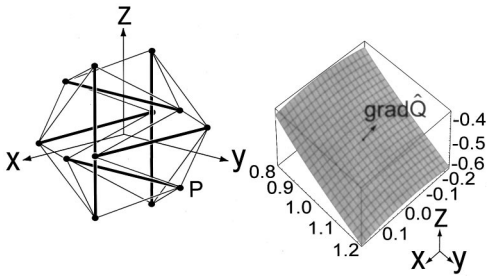


Fig. 7 The configuration space near node 5 of the spherical tensegrity module

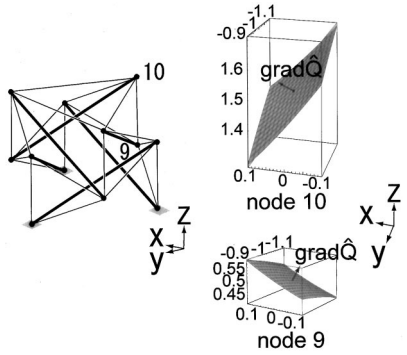


Fig. 8 The configuration spaces near node 9 and node 10 of the two-stage tensegrity module

hypersurface only satisfies the condition for the existence of a prestress mode $\mathbf{s}(0)$ that satisfies $\mathbf{A}(0)\mathbf{s}(0)=\mathbf{0}$. In addition, the prestress mode must be admissible with tension in cables and compression in bars.

In order to show that there exists such an admissible prestress mode for small change of nodal coordinates $\Delta\xi_0$, Murakami [19] performed a perturbation analysis in which the initial equilibrium matrix was perturbed from the ideal $\mathbf{A}_0 \equiv \mathbf{A}(0)$ to $\mathbf{A}=\mathbf{A}_0+\varepsilon\mathbf{A}_1$ where ε is a small real number indicating a norm of $\Delta\xi_0$. It was shown that the resulting changes in prestress and infinitesimal mechanism modes in (19*a*, 19*b*) are also on the order of ϵ. The above existence of admissible prestress modes is shown only for small nodal changes based upon local analyses.

For redundant tensegrity structures with $r_A=n_V$ and $Mx>0$, infinitesimal mechanism modes do not exist due to Calladine's relation (3) (see Section 4). The quadratic form associated with $\mathbf{A}^T(0)\mathbf{A}(0)$ becomes positive definite instead of positive semidefinite. The redundant tensegrity structures can take any configuration as long as connections between bars and cables do not change. (A continuous change in nodal coordinates of tensegrity structures without altering connections involves ''algebraic topology'' dealing with simplicial complexes (for example, Frankel [32]), which is beyond the scope of the present paper.)

9 Concluding Remarks

A set of characterization procedures was presented and illustrated for a six-bar spherical tensegrity module and a two-stage cylindrical tensegrity module. The procedures include: (i) computation of the Maxwell number (2), (ii) analytically finding initial shape and an admissible prestress mode, (iii) numerical computation of prestress and infinitesimal mechanism modes, (iii) modal analyses at a prestressed configuration, (iv) computation of load-displacement curves for the determination of a critical load of bars, and (v) the computation of hypersurfaces in the configuration space.

It was found that all infinitesimal mechanism modes were isotropically stiffened at each node by a single prestress mode, as shown in $(9f)$. Further, if bars are properly designed against buckling, tensegrity modules with infinitesimal mechanisms exhibit stable hardening load-displacement relations.

Acknowledgments

The authors would like to thank Professor Robert E. Skelton and Professor Emeritus Theodore Frankel at University of California, San Diego for stimulating discussions on tensegrity structures.

Appendix

Reduced Equilibrium Equations for a Two-Stage Tensegrity Module. In order to obtain both prestress modes and the characteristic Eq. (13), the initial equilibrium equations are derived for a two-stage cylindrical tensegrity module with three bars at each stage. The following parameters are used to analytically describe nodal coordinates:

b: the bar length;

r_0: the radius of the circumscribing circles of the equilateral triangles 1–3 and 10–12;

r_m: the radius of the circumscribing circles of the equilateral triangles 4–6 and 7–9;

α: the additional twist angle;

γ: the overlap ratio in elevation.

Further, for simplicity of notation, the x, y-coordinates of each node are expressed by using (complex variable) phasor notation, while the z-coordinate is expressed by using the standard Cartesian component. For example, the x, y, and z-coordinates of node 2 are expressed as $(r_0 e^{-i\pi/6},0)$ instead of $(r_0\cos\pi/6, -r_0\sin\pi/6,0)$. With this notation, the nodal coordinates are expressed as follows:

$$1(r_0e^{i7\pi/6},0),\quad 2(r_0e^{-i\pi/6},0),\quad 3(r_0e^{i\pi/2},0),$$

$$4(r_me^{i(-\pi/6+\alpha)},h),\quad 5(r_me^{i(\pi/2+\alpha)},h),\quad 6(r_me^{i(7\pi/6+\alpha)},h),$$

$$7(r_me^{i(\pi/6+\alpha)},(1-\gamma)h),\quad 8(r_me^{i(5\pi/6+\alpha)},(1-\gamma)h),$$

$$9(r_me^{t(-\pi/2+\alpha)},(1-\gamma)h),\quad 10(r_0e^{-i\pi/2},(2-\gamma)h),$$

$$11(r_0e^{i\pi/6},(2-\gamma)h),\quad 12(r_0e^{i5\pi/6},(2-\gamma)h).$$

Having all nodal coordinates, it is a routine calculation to find element lengths:

$$\frac{l_0}{r_0}=\sqrt{3}, \tag{A1a}$$

$$\frac{l_v}{r_0}=\sqrt{\left(\frac{b}{r_0}\right)^2-2\sqrt{3}\,\frac{r_m}{r_0}\sin\left(\frac{\pi}{3}+\alpha\right)}, \tag{A1b}$$

$$\frac{l_s}{r_0}=\sqrt{(1-\gamma^2)\left(\frac{r_m}{r_0}\right)^2+\gamma^2\left\{\left(\frac{b}{r_0}\right)^2-1-2\,\frac{r_m}{r_0}\sin\left(\frac{\pi}{6}+\alpha\right)\right\}}, \tag{A1c}$$

$$\frac{l_d}{r_0}=\sqrt{(1-\gamma)^2\left(\frac{b}{r_0}\right)^2+\gamma(2-\gamma)\left\{1+\left(\frac{r_m}{r_0}\right)^2\right\}-2\,\frac{r_m}{r_0}\left\{\cos\alpha-\gamma(2-\gamma)\sin\left(\frac{\pi}{6}+\alpha\right)\right\}}, \tag{A1d}$$

$$\frac{h}{r_0}=\sqrt{\left(\frac{b}{r_0}\right)^2-1-\left(\frac{r_m}{r_0}\right)^2-2\frac{r_m}{r_0}\sin\left(\frac{\pi}{6}+\alpha\right)}. \qquad (A1e)$$

In what follows, the initial equilibrium Eq. (8) at nodes 1 and 9 with $\mathbf{f}(0)=\mathbf{0}$ are presented in vector form by expressing direction cosines $\mathbf{G}\equiv\mathbf{g}(0)$ in terms of nodal position vectors and element lengths. By utilizing the connection diagram in Fig. 4(*a*), the initial equilibrium equation at node 1 is expressed as

$$(\mathbf{x}_2-\mathbf{x}_1+\mathbf{x}_3-\mathbf{x}_1)\frac{s_0}{l_0}+(\mathbf{x}_6-\mathbf{x}_1)\frac{s_v}{l_v}+(\mathbf{x}_9-\mathbf{x}_1)\frac{s_d}{l_d}$$
$$+(\mathbf{x}_4-\mathbf{x}_1)\frac{s_b}{b}=\mathbf{0}, \qquad (A2)$$

where $\mathbf{x}_i$ denotes the position vector of node i.

The equilibrium equation at node 9 becomes

$$(\mathbf{x}_{10}-\mathbf{x}_9)\frac{s_v}{l_v}+(\mathbf{x}_1-\mathbf{x}_9)\frac{s_d}{l_d}+(\mathbf{x}_6-\mathbf{x}_9)\frac{s_1}{l_s}+(\mathbf{x}_4-\mathbf{x}_9)\frac{s_2}{l_s}$$
$$+(\mathbf{x}_{12}-\mathbf{x}_9)\frac{s_b}{b}=\mathbf{0}. \qquad (A3)$$

Equations (*A*2) and (*A*3) could be written in matrix form to yield the reduced initial equilibrium equation: $\mathbf{A}(0)\mathbf{s}(0)=\mathbf{0}$. However, in order to expedite subsequent analyses, element internal forces are combined with element lengths as

$$\mathbf{A}'(0)\mathbf{s}'(0)=\mathbf{0}, \qquad (A4a)$$

where

$$\mathbf{s}'(0)=\left[\frac{s_0}{l_0}\frac{s_v}{l_v}\frac{s_d}{l_d}\frac{s_1}{l_s}\frac{s_2}{l_s}\frac{s_b}{b}\right]^T, \qquad (A4b)$$

and the elements of $\mathbf{A}'(0)=[a'_{ij}]$ are defined as follows:

$$a'_{11}=\frac{3\sqrt{3}}{2}r_0,\quad a'_{12}=\frac{\sqrt{3}}{2}r_0-r_m\cos\left(\frac{\pi}{6}+\alpha\right),$$

$$a'_{13}=\frac{\sqrt{3}}{2}r_0+r_m\sin\alpha,$$

$$a'_{14}=a'_{15}=0,\quad a'_{16}=\frac{\sqrt{3}}{2}r_0+r_m\cos\left(-\frac{\pi}{6}+\alpha\right),$$

$$a'_{21}=\frac{3}{2}r_0,\quad a'_{22}=\frac{r_0}{2}-r_m\sin\left(\frac{\pi}{6}+\alpha\right),\quad a'_{23}=\frac{r_0}{2}-r_m\cos\alpha,$$

$$a'_{24}=a'_{25}=0,\quad a'_{26}=\frac{r_0}{2}+r_m\sin\left(-\frac{\pi}{6}+\alpha\right),$$

$$a'_{31}=0,\quad a'_{32}=h,\quad a'_{33}=(1-\gamma)h,\quad a'_{34}=a'_{35}=0,\quad a'_{36}=h,$$

$$a'_{41}=0,\quad a'_{42}=-r_m\sin\alpha,\quad a'_{43}=-\frac{\sqrt{3}}{2}r_0-r_m\sin\alpha,$$

$$a'_{44}=-r_m\sin\left(\frac{\pi}{3}+\alpha\right),$$

$$a'_{45}=r_m\sin\left(\frac{\pi}{3}-\alpha\right),\quad a'_{46}=-\frac{\sqrt{3}}{2}r_0-r_m\sin\alpha,$$

$$a'_{51}=0,\quad a'_{52}=-r_0+r_m\cos\alpha,$$

$$a'_{53}=-\frac{r_0}{2}+r_m\cos\alpha,\quad a'_{54}=r_m\sin\left(\frac{\pi}{6}-\alpha\right),$$

$$a'_{55}=r_m\sin\left(\frac{\pi}{6}+\alpha\right),\quad a'_{56}=\frac{r_0}{2}+r_m\cos\alpha,$$

$$a'_{61}=0,\quad a'_{62}=h,\quad a'_{63}=-(1-\gamma)h,\quad a'_{64}=a'_{65}=\gamma h,\quad a'_{66}=h. \qquad (A4c)$$

When a nontrivial $\mathbf{s}'(0)$ exists, the determinant of $\mathbf{A}'(0)$ becomes zero. (The resulting characteristic Eq. (13) may be obtained by using a symbolic manipulator.) In what follows, an analytical method is used to obtain both prestress modes and the characteristic equation by knowing that a nontrivial solution $\mathbf{s}'(0)$ exists only when rank of $\mathbf{A}'(0)$ reduces from six to five. Therefore, Eqs. (*A*4*a*) can be solved uniquely for s_0/l_0, s_v/l_v, s_d/l_d, s_1/l_s, and s_2/l_s when s_b/b is prescribed by using the first five rows of (*A*4*a*). The last row yields the consistency condition, which becomes the characteristic Eq. (13).

In order to eliminate s_0/l_0 from the first and second rows, one subtracts $\sqrt{3}$ times the second row from the first row. The resulting equation and the third row can be used to solve for s_v/l_v and s_d/l_d, as shown in (14*b*) and (14*d*), for a prescribed s_b/b. By using the second row of (*A*4*a*) with (14*b*) and (14*d*), one solves for s_0/l_0 to find (14*a*). The fourth and fifth rows of (*A*4*a*) are used to find s_1/l_s and s_2/l_s in (14*c*). Since there are only five linearly independent rows in (*A*4*a*), the sixth row is linearly dependent of the first five rows. Therefore, the sixth row furnishes a consistency condition that yields the characteristic Eq. (13). The above analytical procedure can be easily applied to two-stage cylindrical tensegrity modules with m-bars at each stage and n-stage cylindrical tensegrity modules with m-bars at each stage.

References

[1] Schultz, D. G., 1981, *Kenneth Snelson, an Exhibition, Albright-Knox Art Gallery, Buffalo, NY*, Thorney-Sidney Press Inc., Buffalo, NY.

[2] Marks, R., and Fuller, R. B., 1973, *The Dymaxion World of Buckminster Fuller*, Anchor Books, Garden City, NY.

[3] Pugh, A., 1976, *An Introduction to Tensegrity*, University of California Press, Berkeley, CA.

[4] Kenner, H., 1976, *Geodesic Math and How to Use It*, University of California Press, Berkeley, CA.

[5] Skelton, R. E., and Sultan, C., 1997, ''Controllable Tensegrity, a New Class of Smart Structures,'' *Proceedings of the International Society for Optical Engineers, 4th Symposium on Smart Structures and Materials*, Vol. 3039, V. V. Varadan and J. Chandra, eds., San Diego, CA, Mar. 3–7, SPIE, Bellingham, WA, pp. 166–177.

[6] Motro, R., 1990, ''Tensegrity Systems and Geodesic Domes,'' Int. J. Space Struct., **5**, pp. 341–351.

[7] Furuya, H., 1992, ''Concept of Deployable Tensegrity Structures in Space Application,'' Int. J. Space Struct., **7**, pp. 143–151.

[8] Hanaor, A., 1993, ''Double-Layer Tensegrity Grids as Deployable Structures,'' Int. J. Space Struct., **8**, pp. 135–143.

[9] Hanaor, A., 1997, ''Tensegrity: Theory and Application,'' *Beyond the Cube: The Architecture of Space Frames and Polyhedra*, J. F. Gabriel, ed., John Wiley and Sons, New York, pp. 385–408.

[10] Maxwell, J. C., 1864, ''On the Calculation of the Equilibrium and Stiffness of Frames,'' Philos. Mag., **27**, p. 250; also 1890 in *The Scientific Papers of James Clerk Maxwell*, W. D. Niven, ed., Dover New York, pp. 598–604.

[11] Calladine, C. R., 1978, ''Buckminster Fuller's ''Tensegrity'' Structures and Clerk Maxwell's Rules for the Construction of Stiff Frames,'' Int. J. Solids Struct., **14**, pp. 161–172.

[12] Argyris, J. H., and Scharpf, D. W., 1972, ''Large Deflection Analysis of Prestressed Networks,'' J. Struct. Div. ASCE, **98**, No. ST3, pp. 633–654.

[13] Bathe, K.-J., 1982, *Finite Element Procedure in Engineering Analysis*, Prentice-Hall, Inc., Englewood Cliffs, NJ, Chapter 11.

[14] Underwood, P., 1983, ''Dynamic Relaxation,'' *Computational Methods for Transient Analysis*, T. Belytschko and T. J. R. Hughes, eds., Elsevier, New York, Chapter 5.

[15] Barnes, M., 1994, ''Form and Stress Engineering of Tension Structures,'' Structural Engineering Review, **6**, No. 3–4, pp. 175–202.

[16] Murakami, H., 2000, ''Static and Dynamic Analysis of Tensegrity Structures, Part 1. Nonlinear Equations of Motion,'' Int. J. Solids Struct., in press.

[17] Hughes, T. J. R., 1987, *The Finite Element Method, Linear Static and Dynamic Finite Element Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

[18] Malvern, L., 1969, *Introduction to the Mechanics of a Continuous Medium*, Prentice-Hall, Englewood Cliffs, NJ.

[19] Murakami, H., 2000, ''Static and Dynamic Analysis of Tensegrity Structures, Part 2. Quasi-Static Analysis,'' Int. J. Solids Struct., in press.

[20] Tarnai, T., 1980, ''Simultaneous Static and Kinematic Indeterminancy of Space Trusses With Cyclic Symmetry,'' Int. J. Solids Struct., **16**, pp. 347–359.

[21] Tobie, R. S., 1967, ''A Report on an Inquiry into the Existence, Formation and Representation of Tensile Structures,'' Masters' thesis, Pratt Institute, Brooklyn, NY.

[22] Williamson, D., and Skelton, R. E., 1998, ''A General Class of Tensegrity

Systems: Geometric Definition,'' *Engineering Mechanics: A Force for the 21st Century, the Proceedings of the 12th ASCE Engineering Mechanics Conference*, H. Murakami and J. E. Luco, eds., La Jolla, CA, May 17–20, ASCE, Reston, VA, pp. 732–735.
[23] Sultan, C., 1999, ''Modeling, Design and Control of Tensegrity Structures with Applications,'' Ph.D. dissertation, Purdue University, West Lafayette, IN.
[24] Sokolnikoff, I. S., 1956, *Mathematical Theory of Elasticity*, 2nd Ed., McGraw-Hill, New York, Section 26.
[25] Naylor, A. W., and Sell, G. R., 1971, *Linear Operator Theory in Engineering and Science*, Holt, Rinehart and Winston, New York, Chapter 5.
[26] Pellegrino, S., and Calladine, C. R., 1986, ''Matrix Analysis of Statically and Kinematically Indeterminate Frameworks,'' Int. J. Solids Struct., **22**, pp. 409–428.
[27] Noble, B., and Daniel, J. W., 1977, *Applied Linear Algebra*, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ.
[28] Skelton, R. E., and Akhikari, R., 1998, ''An Introduction to Smart Tensegrity Structures,'' *Engineering Mechanics: A Force for the 21st Century, the Proceedings of the 12th ASCE Engineering Mechanics Conference*, H. Murakami and J. E. Luco, eds., La Jolla, CA, May 17–20, ASCE, Reston, VA, pp. 24–27.
[29] Koiter, W. T., 1945, ''On the Stability of Elastic Equilibrium,'' thesis, Delft University, Amsterdam, The Netherlands (in Dutch); also 1970, AFFDL-TR-70-25, (English Translation).
[30] Budiansky, B., 1974, ''Theory of Buckling and Post-Buckling Behavior of Elastic Structures,'' *Advances in Applied Mechanics*, Vol. 14, C.-H. Yih, ed., Academic Press, San Diego, pp. 1–65.
[31] Hanaor, A., 1988, ''Prestressed Pin-Jointed Structures—Flexibility Analysis and Prestress Design,'' Comput. Struct., **28**, pp. 757–769.
[32] Frankel, T., 1997, *The Geometry of Physics, an Introduction*, Cambridge University Press, Cambridge, MA, Chapter 6.

Nonlinear Stability, Thermoelastic Contact, and the Barber Condition

J. A. Pelesko
School of Mathematics,
Georgia Institute of Technology,
Atlanta, GA 30332-0160

The behavior of a one-dimensional thermoelastic rod is modeled and analyzed. The rod is held fixed and at constant temperature at one end, while at the other end it is free to separate from or make contact with a rigid wall. At this free end we impose a pressure and gap-dependent thermal boundary condition. This condition, known as the Barber condition, couples the thermal and elastic problems. Such systems have previously been shown to undergo a bifurcation from a unique linearly stable steady-state solution to multiple steady-state solutions with alternating stability. Here, the system is studied using the asymptotic matching techniques of boundary layer theory to derive short-time, long-time, and uniform expansions. In this manner, the analysis is extended into the nonlinear regime and dynamic information about the history dependence and temporal evolution of the solution is obtained. [DOI: 10.1115/1.1345699]

1 Introduction

The analysis of thermal contact problems has revealed a wealth of interesting phenomena. Beginning with J. R. Barber in 1978 ([1]), who pointed out that the solution of such problems poses certain difficulties, and continuing to this day, numerous researchers have turned their attention to these problems. Barber observed that the classical assumption of perfect insulation during a separated phase and perfect thermal contact during contact led to models with solutions which were unacceptable on physical grounds. Introducing a pressure and temperature-dependent boundary condition, which would subsequently become known as the Barber condition, he allowed for a smooth transition between the insulated and perfect thermal contact states. Studying a linearized version of a thermal contact problem which included the Barber condition, he showed that the paradoxes inherent in simpler models could be avoided and physically relevant solutions recovered.

In 1980, Barber, Dundurs, and Comninou [2] investigated a thermal contact problem using the Barber condition in a one-dimensional model of a thermoelastic rod. Imposing a temperature gradient across the rod, they demonstrated that the system underwent a bifurcation from a unique linearly stable steady-state solution to multiple solutions with alternating stability as the magnitude of the thermal gradient was varied.

Since that time, various authors have explored the Barber condition and its implications for thermal contact problems ([3,4]). While such analyses have been extended to multiple materials ([3,5,6]), various geometries ([7,8]), and to numerical simulations ([4]), most theoretical work to date has relied upon linear stability theory. In a recent article ([9]) we developed a nonlinear theory which described the history dependence and dynamics of solutions near the bifurcation point for a simplified model of a one-dimensional thermoelastic rod. Our model did not, however, include the Barber condition. Since the Barber condition is much more physically realistic than the boundary condition used in ([9]), it is desirable to have a nonlinear theory for a model which incorporates the Barber condition. We carry out such an analysis here.

While the model studied here differs from the model studied in ([9]), only in the use of the Barber condition, the method of analysis differs significantly. In particular, here we use the asymptotic matching techniques of boundary layer theory to derive short-time, long-time, and uniform asymptotic expansions of the solution. In our prior analysis we used the method of multiple scales, or two-timing, to accomplish similar goals. The switch in techniques is not merely a matter of taste. Rather, any attempt to apply multiple scale techniques to the model considered herein will soon encounter algebraic difficulties. That is, such an attempt becomes analytically intractable. However, as is shown, boundary layer theory may be applied with little difficulty. This not only allows us to carry out the analysis for the one-dimensional rod model with the Barber condition, but gives us hope that similar techniques will yield a nonlinear stability theory for more complicated multidimensional problems.

We begin in Section 2 by formulating the governing equations for our model. We make the standard assumption that quasi-static uncoupled thermoplasticity is valid and use the Signorini contact condition to capture periods of separation and contact. We impose the Barber condition on the thermal part of the problem, leaving the contact resistance function unspecified. A solution is constructed for the elastic problem and the system of governing equations is reduced to a nonlocal and nonlinear heat conduction problem. In Section 3, we impose physically realistic constraints on the contact resistance and develop a linear theory. We review the analysis due to Barber [2], and show that the system studied undergoes a bifurcation from a single linearly stable steady-state solution to multiple steady-state solutions. Finally, in Section 4, we study the behavior of our system near the bifurcation point. That is, we inquire as to what happens when the system is started nearby the now linearly unstable steady-state solution. Using asymptotic matching techniques, we incorporate the effect of stabilizing nonlinear terms into our theory and obtain information about the dynamics and history dependence of the solution. We show that as conjectured, the solution does indeed approach one of the stable solutions obtained in the linear theory.

2 Formulation of the Model

We consider a one-dimensional thermoelastic rod of length L suspended between two rigid walls as pictured in Fig. 1. We assume that the rod possesses constant thermal and elastic material properties, is homogeneous and isotropic, and that uncoupled quasi-static thermoelastic theory is valid. With these assumptions

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Sept. 24, 1999; final revision, June 26, 2000. Associate Editor: J. R. Barber. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

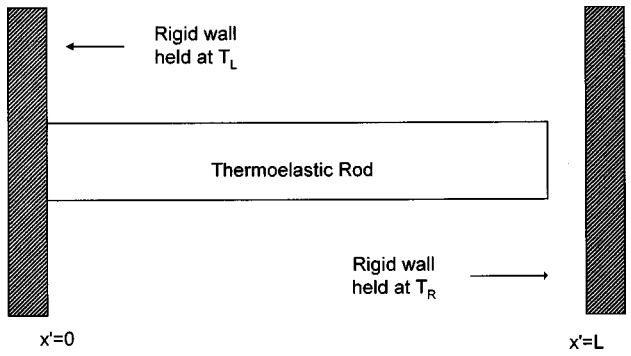


Fig. 1 Sketch of the model geometry

in mind, we formulate the equations governing the temperature distribution, T, elastic displacement, u', and stress, σ', within the rod. In the dimensionless variables

$$\theta=\frac{T-T_L}{T_R-T_L},\quad t=\frac{\kappa}{\rho c_p L^2}t',\quad x=\frac{x'}{L},\quad u=\frac{u'}{L},\quad \sigma=\frac{\sigma'}{E}, \tag{2.1}$$

these equations take the form

$$\frac{\partial\theta}{\partial t}=\frac{\partial^2\theta}{\partial x^2}\quad 0<x<1 \tag{2.2}$$

$$\frac{\partial^2 u}{\partial x^2}=\mu\frac{\partial\theta}{\partial x}\quad 0<x<1 \tag{2.3}$$

$$\sigma=\frac{\partial u}{\partial x}-\mu\theta\quad 0<x<1 \tag{2.4}$$

$$\theta(0,t)=0 \tag{2.5}$$

$$u(0,t)=0 \tag{2.6}$$

$$\left\{\begin{array}{c} u\leqslant 0\\ \sigma\leqslant 0\\ u\sigma=0\end{array}\right\}\ \text{at}\quad x=1 \tag{2.7}$$

$$R(\eta)\frac{\partial\theta}{\partial x}(1,t)=1-\theta(1,t) \tag{2.8}$$

where here

$$\mu=\alpha(T_R-T_L),\quad R(\eta)=\frac{\kappa\hat{R}(\eta)}{L} \tag{2.9}$$

and

$$\eta=\sigma(1,t)-u(1,t). \tag{2.10}$$

Note that μ may be interpreted as a nondimensional coefficient of thermal expansion or as a dimensionless measure of a thermal gradient in the problem, while R is a dimensionless form of the contact resistance function. In fact, R may be thought of as a variable Biot number, measuring the relative strengths of heat conduction within the rod and ''convection'' through the rod's right end. The variable η is equal to the contact pressure during contact ($\eta<0$) and the gap size during periods of separation ($\eta>0$). For a full derivation of the model above, the reader is referred to ([9]). As mentioned in the Introduction, the model above differs from that in ([9]) in that the boundary condition at the right end of the rod in ([9]) is replaced here with the Barber condition, Eq. (2.8). The reader will also notice that in the above we have assumed a reference gap width of zero in Eq. (2.7).

Now, we note that the problems for u and σ are linear and may be solved exactly. That is, we may integrate Eq. (2.3) twice and use Eqs. (2.4), (2.6) and (2.7) to solve for u and σ. We find

$$u(x,t)=\mu\int_0^x\theta(\zeta,t)d\zeta-x\max\left\{\delta\int_0^1\theta(\zeta,t)d\zeta,0\right\} \tag{2.11}$$

and

$$\sigma(x,t)=-\max\left\{\mu\int_0^1\theta(\zeta,t)d\zeta,0\right\}. \tag{2.12}$$

Using these solutions, we may evaluate η, i.e.,

$$\eta=\sigma(1,t)-u(1,t)=-\mu\int_0^1\theta(\zeta,t)d\zeta. \tag{2.13}$$

Hence, we have reduced the problem to one for θ only. We are left with

$$\frac{\partial\theta}{\partial t}=\frac{\partial^2\theta}{\partial x^2}\quad 0<x<1 \tag{2.14}$$

$$\theta(0,t)=0 \tag{2.15}$$

$$R(\eta)\frac{\partial\theta}{\partial x}(1,t)=1-\theta(1,t) \tag{2.16}$$

$$\eta=-\mu\int_0^1\theta(\zeta,t)d\zeta. \tag{2.17}$$

3 Linear Theory

In order to proceed with the analysis, we must further characterize the contact resistance function, $R(\eta)$. We recall from the definition of η, Eq. (2.10), that $\eta>0$ corresponds to separation from the wall, and that in this case η measures the size of this gap. Physically, we expect the contact resistance to increase monotonically with gap size. On the other hand, $\eta<0$ corresponds to contact with the wall, and in this case η measures the contact pressure. Here, we expect contact resistance to decrease monotonically with increasing pressure. Further, contact resistance must be a positive quantity and on physical grounds we are led to expect that $R(\eta)$ appears as pictured in Fig. 2.

With these assumptions about R in mind, we may investigate steady-state solutions of the system (2.14), (2.15), and (2.16). We begin by setting the time derivative to zero in Eq. (2.14), integrating the resulting ode and using the boundary conditions, Eqs. (2.15) and (2.16), to determine that steady solutions must have the form

$$\theta^*(x)=ax \tag{3.1}$$

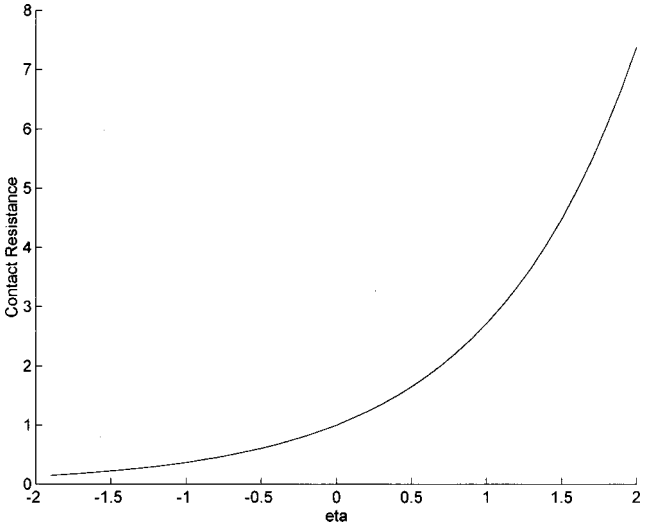


Fig. 2 A typical contact resistance function

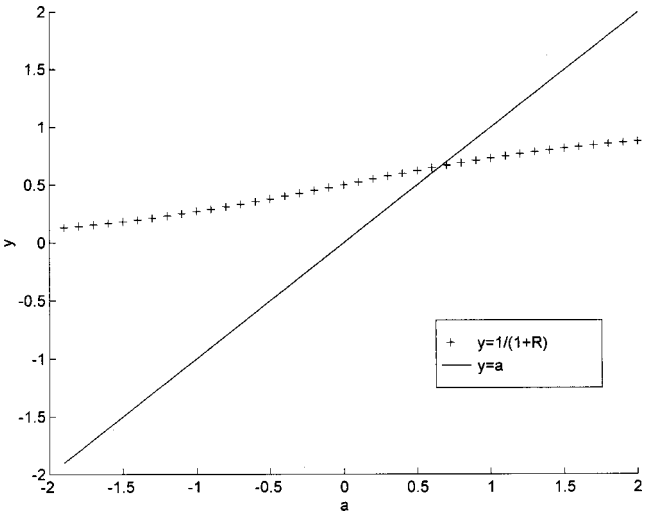


Fig. 3 Geometric solution of the steady-state problem

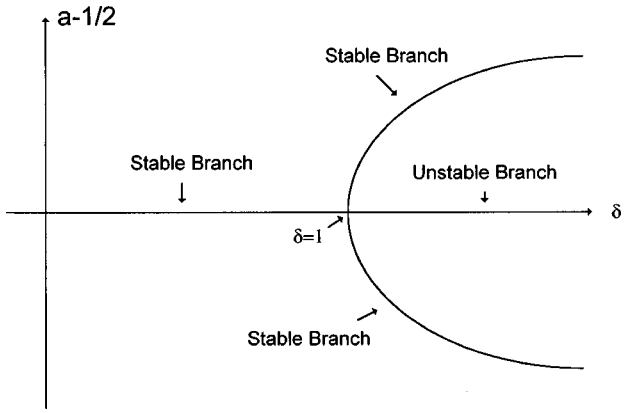


Fig. 4 Bifurcation diagram showing the constant in the steady solution as a function of the bifurcation parameter, δ

where a satisfies

$$a=\frac{1}{1+R(-\mu a/2)}. \tag{3.2}$$

Our observations about the nature of R allow us to plot the left and right sides of Eq. (3.2) on the same plot as functions of a. This is done in Fig. 3. With physically realistic assumptions on R, it is clear that we will always have at least one point of intersection, and hence at least one steady solution. We also note that depending upon the exact nature of R, we may have more than one steady-state solution. To clarify this situation further, we need more detailed information about the contact resistance. For simplicity, we specify the value of R at a convenient point. In particular, we shall assume that $R(-\mu/4)=1$. This implies that $a=1/2$ is a solution of Eq. (3.2) and hence $\theta^*(x)=x/2$ is a steady-state solution of the system (2.14), (2.15), and (2.16). Next, we define

$$F(\eta)=\frac{1}{1+R(\eta)} \tag{3.3}$$

and note that $F(-\mu/4)=1/2$. In order to have a bifurcation of the type investigated by Barber [2], it is easy to see that we must have that $\eta=-\mu/4$ be an inflection point for F. That is, we assume $F''(-\mu/4)=0$ and $F'''(-\mu/4)>0$. This implies that we may expand F in a Taylor series about $\eta=-\mu/4$ as follows:

$$F(\eta)=\frac{1}{2}+(\eta+\mu/4)F'(-\mu/4)+\frac{(\eta+\mu/4)^3}{6}F'''(-\mu/4)+\cdots. \tag{3.4}$$

Throughout the remainder of this paper we shall localize the analysis about the steady-state solution $\theta^*(x)=x/2$. That is, in addition to assuming that $\eta=-\mu/4$ is an inflection point for F, we assume that nearby this inflection point F is well approximated by the first three nonzero terms in the Taylor series (3.4).

Next, we investigate the linear stability of the steady solution $\theta^*(x)=x/2$. Accordingly we seek a solution to (2.14), (2.15), and (2.16) in the form

$$\theta(x,t)=\frac{x}{2}+\phi(x)e^{-\lambda^2 t} \tag{3.5}$$

where $|\phi(x)|\ll 1$. Inserting this ansatz into our governing equations, expanding the nonlinear terms in Taylor series, and omitting quadratic and higher order terms in ϕ, we obtain the eigenvalue problem

$$\frac{d^2\phi}{dx^2}+\lambda^2\phi=0 \tag{3.6}$$

$$\phi(0)=0 \tag{3.7}$$

$$\phi'(1)+\phi(1)=4\delta\int_0^1\phi(\zeta)d\zeta. \tag{3.8}$$

Here $\delta=\mu R'(-\mu/4)/8$. This linear eigenvalue problem has a solution ϕ when λ satisfies

$$\lambda^2\cos(\lambda)+4\delta(\cos(\lambda)-1)+\lambda\sin(\lambda)=0. \tag{3.9}$$

The solutions of this equation in conjunction with Eq. (3.5) determine the stability of the perturbation ϕ; if $\mathrm{Re}(\lambda^2)>0$ (<0), then the steady state is linearly stable (unstable). The solutions of Eq. (3.9) were studied by Barber [2]; we do not repeat his analysis here. Rather, we simply note that in our notation, Barber's result is that $\delta<1$ corresponds to linear stability, $\delta>1$ corresponds to linear instability, while $\delta=1$ is the marginally stable case.

With the assumptions mentioned above concerning F, another alternative characterization of the bifurcation as δ passes through one is possible. Retaining only up to cubic terms, and using (3.4) in the equation defining the steady states, (3.2), we obtain a cubic polynomial for a. By construction, one solution is of course, $a=1/2$. The other two solutions are given by

$$a=\frac{1}{2}\pm\left(\frac{2}{\mu}\right)^{3/2}\sqrt{\frac{6(\delta-1)}{F'''(-\mu/4)}}. \tag{3.10}$$

We see that these solutions are unphysical (imaginary) for $\delta<1$, and that we pick up two new physical solutions as δ passes through one. At least locally, the bifurcation is of the standard pitchfork type. The linear theory is summarized in Fig. 4.

4 Nonlinear Theory

In the previous section we found and investigated the linear stability of steady-state solutions to our model, i.e., Eqs. (2.14)–(2.16). We made physically realistic assumptions about the contact resistance function R, and determined that $\theta^*(x)=x/2$ was a solution for all positive values of the parameter δ. We showed that the linear stability of this solution changed as δ passed through one. In particular, for $\delta<1$, this solution was found to be linearly stable, while for $\delta>1$, linear theory predicts that any infinitesimal perturbation will grow exponentially. Clearly, in this parameter range, the linear theory is only valid for a limited time. In this section, we use the asymptotic matching techniques of boundary layer theory to extend our analysis into the nonlinear regime. That is, we investigate the nature of the solution to our governing equations when δ is nearby, but greater than one, and the initial con-

ditions are such that the system starts near the now unstable solution, θ^*. Our goal is to develop an approximate solution which is valid for all time, thereby allowing us to understand the dynamics and history dependence of solutions near this bifurcation point.

We begin, by imposing the initial condition

$$\theta(x,0)=\frac{x}{2}+\epsilon h(x). \tag{4.1}$$

Here, $\epsilon \ll 1$ and $h(x)$ is an arbitrary $O(1)$ function. Note that this defines ϵ and starts our system near $\theta^*(x)$. It is now convenient to rescale by setting $\theta=\epsilon v+\theta^*(x)$. Introducing this rescaling into Eqs. (2.14), (2.15), (2.16), and (4.1), expanding the nonlinear terms in a Taylor series and retaining terms up to $O(\epsilon^2)$ we obtain

$$\frac{\partial v}{\partial t}=\frac{\partial^2 v}{\partial x^2} \tag{4.2}$$

$$v(0,t)=0 \tag{4.3}$$

$$\frac{\partial v}{\partial x}(1,t)+v(1,t)=\left(4\delta\int_0^1 v(\zeta,t)d\zeta\right)\left[1+\epsilon\frac{\partial v}{\partial x}(1,t)-\epsilon v(1,t)\right]-\epsilon^2 c_0^2\left(\int_0^1 v(\zeta,t)d\zeta\right)^3 \tag{4.4}$$

$$v(x,0)=h(x) \tag{4.5}$$

where here $c_0^2=\mu F'''/3$ and is, by assumption, a positive number. We assume $c_0^2=O(1)$.

Next, we let $\delta=1+\gamma\epsilon^2$ where $\gamma=O(1)$ and we seek a solution in the form

$$v(x,t)\sim v_0(x,t)+\epsilon v_1(x,t)+\epsilon^2 v_2(x,t)+\cdots. \tag{4.6}$$

Inserting this expansion into Eqs. (4.2)–(4.5), and equating to zero coefficients of powers of ϵ, we find that $v_0(x,t)$ satisfies

$$\frac{\partial v_0}{\partial t}=\frac{\partial^2 v_0}{\partial x^2} \tag{4.7}$$

$$v_0(0,t)=0 \tag{4.8}$$

$$\frac{\partial v_0}{\partial x}(1,t)+v_0(1,t)=4\int_0^1 v_0(\zeta,t)d\zeta \tag{4.9}$$

$$v_0(x,0)=h(x). \tag{4.10}$$

We construct a solution using eigenfunction expansion. Accordingly, we seek solutions in the form $A(t)\phi(x)$, separate variables and obtain the eigenvalue problem, Eqs. (3.6)–(3.8), with $\delta=1$ for the spatial eigenfunctions, $\phi(x)$. Hence the eigenvalues are given by Eq. (3.9) with $\delta=1$. Further, from the linear theory in the previous section and from Barber's analysis, we note that zero is an eigenvalue and that all other eigenvalues are purely real. Next, we must take our analysis one step further and explicitly construct the eigenfunctions and derive an expansion theorem. Towards this end it is useful to remove the integral from the boundary condition, Eq. (3.8). We integrate Eq. (3.6) from zero to one, solve for the integral, and use this result to eliminate the integral in Eq. (3.6). This yields the equivalent system

$$\frac{d^2\phi}{dx^2}+\lambda^2\phi=0 \tag{4.11}$$

$$\phi(0)=0 \tag{4.12}$$

$$4\left(\frac{d\phi}{dx}(0)-\frac{d\phi}{dx}(1)\right)=\lambda^2\left(\frac{d\phi}{dx}(1)+\phi(1)\right). \tag{4.13}$$

We note that $\lambda=0$ remains an eigenvalue of this system. However, the new formulation, (4.11)–(4.13), makes clear the fact that we are faced with a nonstandard eigenvalue problem. That is, the eigenvalue parameter, λ, appears in the boundary conditions. Consequently, we cannot simply consider the operator $L=-d^2/dx^2$ and rely upon the theory of eigenfunction expansion for $L\phi=-\lambda^2\phi$. Rather, we must exercise care in defining an operator, constructing an adjoint, and in deriving an expansion theorem. We follow a typical approach as outlined, for example, in Friedman [10].

We begin by considering the space of two component vectors U, whose first component is a real-valued C^2 function, $u(x)$, and whose second component is a real number, u_1. We define the inner product of two vectors in this space by

$$\langle U,V\rangle=\int_0^1 u(x)v(x)dx+u_1v_1. \tag{4.14}$$

Next, we restrict our attention to the subspace, D, of vectors U such that $u(0)=0$ and $u_1=u(1)+u'(1)$. Then, we define an operator L acting on elements U of D by

$$LU=\begin{pmatrix} -\dfrac{d^2u}{dx^2} \\ 4\left(\dfrac{du}{dx}(0)-\dfrac{du}{dx}(1)\right)\end{pmatrix}. \tag{4.15}$$

Note that our eigenvalue problem, Eqs. (4.11)–(4.13), is now simply stated as find a vector U in D such that $LU=\lambda^2U$. Further, we may define an adjoint operator, L^*, where

$$L^*V=\begin{pmatrix} -\dfrac{d^2u}{dx^2} \\ \dfrac{dv}{dx}(1)\end{pmatrix} \tag{4.16}$$

and acts on elements, V, of the subspace D^* defined as two component vectors satisfying $v'(1)=v(0)-v(1)$ and $v_1=-v(0)/4$. The reader may easily verify that with the inner product, (4.14), we have $\langle LU,V\rangle=\langle U,L^*V\rangle$.

Next, we may attempt to derive an expansion theorem and solve our leading order problem. First, we note, of course, that the discrete spectrum of L is given by Eq. (3.9). The fact that our operator is not self-adjoint raises the possibility that L also possesses a continuous spectrum which would effect the nature of an eigenfunction expansion. By using a Green's function approach, we may rule out this possibility. The details of obtaining this null result are lengthy, the interested reader is referred to Appendix B of ([9]) for an example of this calculation. This having been said, we now construct eigenvectors. We find

$$U_n=\begin{pmatrix} \dfrac{a_n\sin(\lambda_n x)}{\lambda_n} \\ \dfrac{a_n\lambda_n\cos(\lambda_n)+a_n\sin(\lambda_n)}{\lambda_n}\end{pmatrix} \tag{4.17}$$

where $\lambda_0=0$ and the remaining λ_n's are the real nonzero solutions of Eq. (3.9) for $\delta=1$. Similarly, we can construct the following adjoint eigenfunctions from our adjoint eigenvalue problem:

$$V_n=\begin{pmatrix} b_n\left(\cos(\lambda_n x)+\left(\dfrac{4\sin(\lambda_n)-\lambda_n}{4\cos(\lambda_n)}\right)\sin(\lambda_n x)\right) \\ -\dfrac{b_n}{4}\end{pmatrix} \tag{4.18}$$

where, of course, the λ_n's are the same as above. We note that $\langle U_n,V_m\rangle=0$ for all $n\neq m$ and that $\langle U_n,V_n\rangle\neq 0$ for all $n\neq 0$ while $\langle U_0,V_0\rangle=0$. Further, we choose the a_n's, for $n\neq 0$ so that $\langle U_n,V_n\rangle=1$ and we choose a_0 so that $\langle U_0,U_0\rangle=1$.

Now, we can construct a solution to our leading order problem by using the eigenvectors and the adjoint eigenvectors just defined. We seek a solution lying in the subspace D in the following form:

$$\begin{pmatrix} v_0(x,t) \\ \dfrac{\partial v_0}{\partial x}(1,t)+v_0(1,t) \end{pmatrix} = \sum_{n=0}^{\infty} A_n e^{-\lambda_n^2 t} U_n . \tag{4.19}$$

The governing equation and the boundary conditions are of course satisfied, while the A_n's are still unknown. They will be determined by our initial conditions. We require that

$$\begin{pmatrix} v_0(x,0) \\ \dfrac{\partial v_0}{\partial x}(1,0)+v_0(1,0) \end{pmatrix} = \begin{pmatrix} h(x) \\ h_1 \end{pmatrix} = \sum_{n=0}^{\infty} A_n U_n \tag{4.20}$$

where h_1 and the A_n are yet to be determined. If we take an inner product with V_m, where $m \neq 0$, we find $A_m = \langle H, V_m \rangle$ where

$$H = \begin{pmatrix} h(x) \\ h_1 \end{pmatrix}. \tag{4.21}$$

Next, we take an inner product with V_0 and find $\langle H, V_0 \rangle = 0$ which implies

$$4\int_0^1 h(x)dx = h_1 \tag{4.22}$$

and hence uniquely determines h_1. Now, A_0 is still undetermined. To remedy this situation, we take an inner product with U_0 across Eq. (4.20) and solve for A_0 to find

$$A_0 = \langle H, U_0 \rangle - \sum_{n=1}^{\infty} A_n \langle U_n, U_0 \rangle. \tag{4.23}$$

We now have a complete solution for $v_0(x,t)$.

Noting that $\lambda_0 = 0$ and that all other λ_n's are real, we see from Eq. (4.19) that all modes except for the λ_0 mode decay in the large time limit. This implies that as $t \to \infty$ we have $v_0(x,t) \to A_0\sqrt{3/13}x$. If we now attempted to compute a solution for $v_1(x,t)$, which is forced by the v_0 solution, we would find that $v_1 \to \infty$ as $t \to \infty$! This implies that our expansion is nonuniform in time, and hence only serves as a *short-time* solution. To obtain the long-time behavior of our system, we turn to boundary layer theory.

We begin by changing to the slow or long time scale $\tau = \epsilon^2 t$. Our problem for v, Eqs. (4.2)–(4.5), becomes

$$\epsilon^2 \frac{\partial v}{\partial \tau} = \frac{\partial^2 v}{\partial x^2} \tag{4.24}$$

$$v(0,\tau) = 0 \tag{4.25}$$

$$\frac{\partial v}{\partial x}(1,\tau) + v(1,\tau) = \left(4\delta \int_0^1 v(\zeta,t)d\zeta\right)\left[1 + \epsilon \frac{\partial v}{\partial x}(1,\tau) - \epsilon v(1,\tau)\right] - \epsilon^2 c_0^2 \left(\int_0^1 v(\zeta,\tau)d\zeta\right)^3. \tag{4.26}$$

Here, we seek a solution in the form

$$v(x,\tau) \sim v_0(x,\tau) + \epsilon v_1(x,\tau) + \epsilon^2 v_2(x,\tau) + \cdots . \tag{4.27}$$

Introducing this expansion into our long-time Eqs. (4.24)–(4.26), and equating to zero coefficients of powers of ϵ we again obtain an infinite set of equations which sequentially determine the v_n. In order to determine the leading order solution, we shall need the equations up to order ϵ^2. Our order one equations are

$$\frac{\partial^2 v_0}{\partial x^2} = 0 \tag{4.28}$$

$$v_0(0,\tau) = 0 \tag{4.29}$$

$$\frac{\partial v_0}{\partial x}(1,\tau) + v_0(1,\tau) = 4\int_0^1 v_0(\zeta,\tau)d\zeta. \tag{4.30}$$

This system may be solved and we find $v_0(x,\tau) = A(\tau)x$ where $A(\tau)$ is an undetermined function of the slow time variable, τ. Using this we may simplify the $O(\epsilon)$ system and we find

$$\frac{\partial^2 v_1}{\partial x^2} = 0 \tag{4.31}$$

$$v_1(0,\tau) = 0 \tag{4.32}$$

$$\frac{\partial v_0}{\partial x}(1,\tau) + v_0(1,\tau) = 4\int_0^1 v_1(\zeta,\tau)d\zeta. \tag{4.33}$$

This system may also be solved and we find $v_1(x,\tau) = B(\tau)x$ where B is an unknown function. Using our $O(1)$ and $O(\epsilon)$ solutions we can simplify the $O(\epsilon^2)$ problem. We find that $v_2(x,\tau)$ satisfies

$$\frac{\partial v_0}{\partial \tau} = \frac{\partial^2 v_2}{\partial x^2} \tag{4.34}$$

$$v_2(0,\tau) = 0 \tag{4.35}$$

$$\frac{\partial v_2}{\partial x}(1,\tau) + v_2(1,\tau) = 4\int_0^1 v_2(\zeta,\tau)d\zeta + 4\gamma\int_0^1 v_0(\zeta,\tau)d\zeta - c_0^2\left(\int_0^1 v_0(\zeta,\tau)d\zeta\right)^3. \tag{4.36}$$

Using our solution for v_0, integrating Eq. (4.34) with respect to x and applying the boundary conditions, we find that this system only possesses a solution if $A(\tau)$ satisfies

$$\frac{dA}{d\tau} = 4\gamma A - \frac{c_0^2}{4}A^3. \tag{4.37}$$

This ordinary differential equation determines $A(\tau)$ up to an arbitrary constant which is obtained by matching back to the short-time solution. That is the initial condition for $A(\tau)$ is given by

$$A(0) = \lim_{t\to\infty} \frac{v_0(x,t)}{x} = A_0\sqrt{\frac{3}{13}}. \tag{4.38}$$

Our short-time solution, Eq. (4.19), and our long-time solution, $A(\tau)x$, may be assembled into a uniformly valid solution. That is

$$v(x,t) \sim A(\epsilon^2 t)x - A_0\sqrt{\frac{3}{13}}x + \sum_{n=0}^{\infty} A_n e^{-\lambda_n^2 t}\frac{a_n \sin(\lambda_n x)}{\lambda_n} + O(\epsilon) \tag{4.39}$$

gives the leading order behavior of solutions for all time.

5 Discussion

We began by formulating a model of a one-dimensional thermoelastic rod subjected to conditions which allowed for thermoelastic contact and the possibility of a thermoelastic instability. In contrast to our earlier nonlinear stability theory, ([9]), in this model we included a general form of the Barber condition. Physically based assumptions about the nature of the contact resistance function, $R(\eta)$, were made. With these assumptions we set out to verify Barber's linear theory ([2]), and to extend his analysis into the nonlinear regime.

The linear theory showed that for a certain class of contact resistance functions, or more precisely for suitable assumptions on the reciprocal contact resistance function, F, the system underwent a bifurcation as δ passed through one. That is, just as Barber discovered, there is a transition from one stable stationary solution

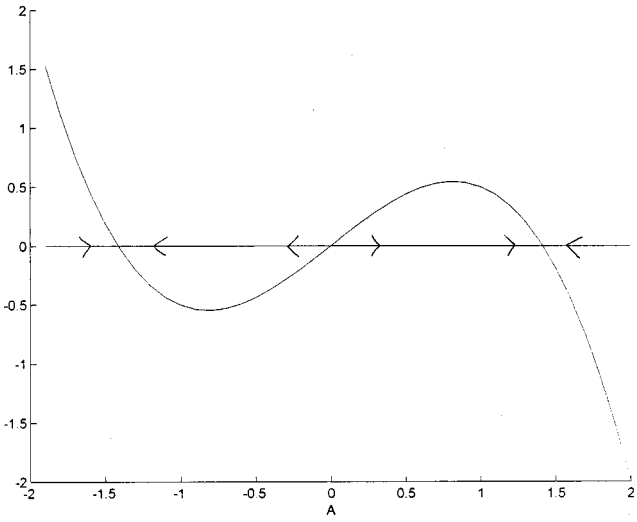


Fig. 5 Behavior of solutions to the amplitude equation, which governs $A(\tau)$

to three solutions with alternating stability. The solution that was stable undergoes an exchange of stabilities and becomes unstable for $\delta > 1$.

Next, we attempted to extend the linear analysis into the nonlinear regime in the neighborhood of $\delta = 1$. In doing so, we hoped to verify the conjecture of Barber that solutions which start near the now unstable steady-state approach one of the two linearly stable solutions uncovered in the linear analysis. Further, we would like the nonlinear analysis to clarify what initial conditions go to which solution and how they get there. That is, we want to understand history dependence and dynamics in the neighborhood of the bifurcation. To accomplish this goal, we developed a uniformly valid asymptotic approximation to the solution using the asymptotic matching techniques of boundary layer theory. These techniques yielded the following asymptotic approximation to the solution:

$$\theta(x,t) \sim \frac{x}{2} + \epsilon A(\epsilon^2 t)x - \epsilon A_0 \sqrt{\frac{3}{13}}x + \epsilon \sum_{n=0}^{\infty} A_n e^{-\lambda_n^2 t} \frac{a_n \sin(\lambda_n x)}{\lambda_n} + O(\epsilon^2) \tag{5.1}$$

where $A(\tau)$ satisfied the amplitude equation

$$\frac{dA}{d\tau} = 4\gamma A - \frac{c_0^2}{4} A^3 \tag{5.2}$$

with initial condition

$$A(0) = A_0 \sqrt{\frac{3}{13}}. \tag{5.3}$$

Now, as time tends to infinity all contributions from the sum in Eq. (5.1) decay to zero or cancel with other terms. This means that the limiting behavior is given by

$$\lim_{t\to\infty} \theta(x,t) \sim \frac{x}{2} + \epsilon x \lim_{t\to\infty} A(\epsilon^2 t) + O(\epsilon^2) \tag{5.4}$$

and our questions concerning nonlinear stability may be answered by examining the governing equation for A, i.e., Eq. (5.2). We note that Eq. (5.2) includes a cubic nonlinearity. Recall that this term arose due to the nonlinear nature of the contact resistance function, $R(\eta)$. Further note that this nonlinearity exerts a stabilizing influence on the solution. In Fig. 5, we sketch the phase plane for this amplitude equation. We see that A approaches $\pm 4\sqrt{\gamma}/c_0$ according as the initial condition is positive or negative. This implies that the solution tends to ax where

$$a = \frac{1}{2} \pm \left(\frac{2}{\mu}\right)^{3/2} \sqrt{\frac{6(\delta - 1)}{F'''(-\mu/4)}}. \tag{5.5}$$

These are precisely the solutions uncovered by the linear theory and hence Barber's conjecture is verified. Further, questions of history dependence may now be answered by simply examining the initial condition on A. The sign of this condition dictates whether we tend to the positive or negative solution. This sign in turn simply depends on the direction of the perturbation to the unstable steady solution. Similarly, questions concerning dynamics of solutions are answered by the time behavior of Eqs. (5.1) and (5.2).

Finally, a comment about the method of analysis is in order. As stated in the Introduction, the switch from the method of multiple scales to boundary layer theory was necessary in order to be able to carry out the analysis. As can be seen from the section on the nonlinear theory, this technique allows one to explicitly solve the reduced equations at each order. Such solutions are algebraically intractable with the multiscale approach. This phenomena has been observed in nonlinear stability theory for other types of problems, examples may be found in ([11]) or ([12]). This simplicity does, however, come at a price. In particular, we only discover the slow time behavior of the *dominant* mode. As all other modes decay, this price is not too steep, but yet it should be acknowledged. The gift of simplicity, however, gives one hope that multidimensional nonlinear stability theories are within reach. As a final note of inspiration to the reader, we observe that Barber's linear theory has now been extended to a nonlinear theory in a neighborhood of the bifurcation. Such a nonlinear theory is often referred to as a weakly nonlinear stability theory. A global nonlinear stability theory would be of interest and remains a challenge for the curious researcher.

References

[1] Barber, J. R., 1978, ''Contact Problems Involving a Cooled Punch,'' J. Elast., **8**, pp. 409–423.

[2] Barber, J. R., Dundurs, J., and Comninou, M., 1980, ''Stability Considerations in Thermoelastic Contact,'' ASME J. Appl. Mech., **47**, pp. 871–874.

[3] Panek, C., 1980, ''A Thermomechanical Example of Auto-Oscillation,'' ASME J. Appl. Mech., **47**, pp. 875–878.

[4] Olesiak, Z. S., and Pyryev, Y. A., 1996, ''Transient Response in a One-Dimensional Model of Thermoelastic Contact,'' ASME J. Appl. Mech., **63**, pp. 575–581.

[5] Barber, J. R., 1981, ''Stability of Thermoelastic Contact for the Aldo Model,'' ASME J. Appl. Mech., **48**, pp. 555–558.

[6] Zhang, R., and Barber, J. R., 1990, ''Effect of Material Properties on the Stability of Static Thermoelastic Contact,'' ASME J. Appl. Mech., **57**, pp. 365–369.

[7] Yeo, T., and Barber, J. R., 1995, ''Stability of a Semi-Infinite Strip in Thermoelastic Contact With a Rigid Wall,'' Int. J. Solids Struct., **32**, pp. 553–567.

[8] Li, C., and Barber, J. R., 1997, ''Stability of Thermoelastic Contact of Two Layers of Dissimilar Materials,'' J. Therm. Stresses, **20**, pp. 169–184.

[9] Pelesko, J. A., 1999, ''Nonlinear Stability Considerations in Thermoelastic Contact,'' ASME J. Appl. Mech., **66**, pp. 109–116.

[10] Friedman, B., 1990, *Principles and Techniques of Applied Mathematics*, Dover, New York.

[11] Segel, L. A., 1966, ''Nonlinear Hydrodynamic Stability Theory and Its Application to Thermal Convection and Curved Flows,'' *Non Equilibrium Thermodynamics: Variational Techniques and Stability*, University of Chicago Press, Chicago, IL.

[12] Kriegsmann, G. A., and Wagner, B. A., 1995, ''Microwave Heating of Carbon-Coated Ceramic Fibers: A Mathematical Model,'' IMA J. Appl. Math., **55**, pp. 243–255.

O. Gendelman
e-mail: ovgend@center.chph.ras.ru

L. I. Manevitch
e-mail: Lmanev@center.chph.ras.ru

Institute of Chemical Physics,
Russian Academy of Sciences,
Kosygin Str. 4,
117977 Moscow, Russia

A. F. Vakakis
Department of Mechanical and Industrial Engineering,
University of Illinois,
1206 W. Green Street,
Urbana, IL 61801
e-mail: avakakis@uiuc.edu

R. M'Closkey
Department of Mechanical and Aerospace Engineering,
University of California,
38-137N Engineering IV,
405 Hilgard Avenue,
Los Angeles, CA 90024-1597
e-mail: obsidian.seas.ucla.edu

Energy Pumping in Nonlinear Mechanical Oscillators: Part I—Dynamics of the Underlying Hamiltonian Systems

The systems considered in this work are composed of weakly coupled, linear and essentially nonlinear (nonlinearizable) components. In Part I of this work we present numerical evidence of energy pumping in coupled nonlinear mechanical oscillators, i.e., of one-way (irreversible) ''channeling'' of externally imparted energy from the linear to the nonlinear part of the system, provided that the energy is above a critical level. Clearly, no such phenomenon is possible in the linear system. To obtain a better understanding of the energy pumping phenomenon we first analyze the dynamics of the underlying Hamiltonian system (corresponding to zero damping). First we reduce the equations of motion on an isoenergetic manifold of the dynamical flow, and then compute subharmonic orbits by employing nonsmooth transformation of coordinates which lead to nonlinear boundary value problems. It is conjectured that a 1:1 stable subharmonic orbit of the underlying Hamiltonian system is mainly responsible for the energy pumping phenomenon. This orbit cannot be excited at sufficiently low energies. In Part II of this work the energy pumping phenomenon is further analyzed, and it is shown that it is caused by transient resonance capture on a 1:1 resonance manifold of the system. [DOI: 10.1115/1.1345524]

1 Introduction

In this and a companion paper we study nonlinear energy pumping in coupled mechanical oscillators. By this terminology, we denote the controlled spatial transfer of vibrational energy from the point of its initial generation to a different (predetermined) point where it eventually localizes. In essence, the energy pumping phenomenon corresponds to the controlled one-way channeling of the vibrational energy to a passive nonlinear ''sink'' where it localizes and diminishes in time due to damping dissipation. There exist numerous studies in the literature on ''static'' mode localization, spatial motion confinement, and on energy transfer due to internal resonances in coupled mechanical oscillators. The nonlinear energy pumping phenomenon discussed herein is a distinct nonlinear mechanism of energy transfer since it is realized through resonance capture ([1]).

Linear and nonlinear passive ''static'' mode localization and spatial motion confinement in periodic and nonperiodic coupled oscillators have been studied extensively in the literature ([2–9]). In these studies, linear and nonlinear standing wave motions were analyzed in ordered and disordered periodic coupled oscillators, and the existence of spatially localized free and forced standing waves was rigorously proven by means of theoretical, numerical, and experimental techniques. The standing wave localization considered in these previous works can be classified as ''static'' since it does not involve any controlled spatial transfer (transition) of energy through the system; indeed linear or nonlinear mode localization can be realized through appropriate selection of the initial conditions of the system, and does not involve any spatial ''flow'' of energy through the system.

Nonlinear transfer of energy between nonlinear modes in internal resonance has also been studied extensively (cf. [10]). In addition, as recently shown by Nayfeh and co-workers, under certain conditions energy transfer from high to low-frequency modes of a weakly nonlinear structure can also occur ([11]). However, these nonlinear energy exchanges are solely due to modal interactions and do not necessarily involve controlled, one-way spatial transfer of energy through the system.

To the authors' best knowledge the only previous study of the nonlinear energy pumping phenomenon is the one by Gendelman [12]. In that work a system of two weakly coupled oscillators, a linear and an essentially (nonlinearizable) nonlinear one, was considered. Pumping of energy was demonstrated numerically by showing that, under certain conditions, energy initially imparted in the linear oscillator transfers to the essentially nonlinear one, even though this later oscillator is not directly excited. However, no rigorous analysis and explanation of this phenomenon is given in that work.

In Part I of this work we present numerical evidence of energy pumping in two and three-degrees-of-freedom coupled oscillators with essential nonlinearities and weak viscous damping. We then focus (for simplicity) in the two-degrees-of-freedom case, and analyze systematically the bifurcation structure of the free nonlinear periodic orbits of the underlying Hamiltonian system with no damping. We show that the occurrence (or lack of) energy pumping can be explained by considering the 1-1 and higher order resonant orbits of the Hamiltonian system. A direct analysis of the energy pumping phenomenon is carried in Part II by transforming the damped equations of motion using the action-angle variables of the underlying Hamiltonian system. We show that energy

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Sept. 29, 1999; final revision, May 2, 2000. Associate Editor: N. C. Perkins. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

pumping is a resonance capture phenomenon on a 1-1 resonant manifold, and construct analytical approximations of energy pumping.

2 Nonlinear Energy Pumping: Numerical Evidence

Consider the following two-degrees-of-freedom system composed of two weakly coupled and weakly damped oscillators:

$$\begin{aligned} \ddot{y}_1+\varepsilon\lambda\dot{y}_1+Cy_1^3+\varepsilon(y_1-y_2)=0 \\ \ddot{y}_2+\varepsilon\lambda\dot{y}_2+\omega_2^2y_2+\varepsilon(y_2-y_1)=0 \end{aligned} \tag{1}$$

Weak coupling is assured by requiring that $\varepsilon \ll 1$, and all other variables are assumed to be $O(1)$ quantities; dots denote differentiation with respect to the independent variable t (time). For $\varepsilon=0$ the system decomposes into two uncoupled nonlinear and linear oscillators, labeled ''Oscillators 1 and 2,'' respectively. We note that oscillator 1 is essentially nonlinear (nonlinearizable).

In Fig. 1 we depict the transient responses of the two oscillators for $\lambda=0.5$, $\omega_2^2=0.9$, $C=5.0$, $\varepsilon=0.1$, and initial conditions $y_1(0)=y_2(0)=0$, $\dot{y}_1(0)=0$, $\dot{y}_2(0)=\sqrt{2h}$, where h (the energy of the system at $t=0+$) varies; these initial conditions correspond to impulsive excitation of oscillator 2 at $t=0$. For $h=0.5$ (cf. Fig. 1(*a*)) both oscillators perform damped free oscillations and no energy pumping occurs, since most energy is stored in the directly excited oscillator 2. By increasing the initial energy level to $h=0.8$ and 1.1.25 (cf. Figs. 1(*b*,*c*)), it is observed that energy transfer from the directly excited oscillator 2 to the unexcited oscillator 1 takes place; indeed, after an initial transient state most of the vibrational energy is irreversibly transferred (''pumped'') to oscillator 1. By further increasing the initial energy level the energy pumping phenomenon becomes less pronounced. This numerical simulation indicates that, for fixed system parameter values, *energy pumping in the weakly coupled system takes place, above a specific value of the initial energy level (strength of the excitation).*

Similar results are obtained for the three-degrees-of-freedom system governed by

$$\begin{aligned} \ddot{y}_1+\varepsilon\lambda\dot{y}_1+Cy_1^3+\varepsilon(y_1-y_2)=0 \\ \ddot{y}_2+\varepsilon\lambda\dot{y}_2+\omega_2^2y_2+\varepsilon(y_2-y_1)+d(y_2-y_3)=0 \\ \ddot{y}_3+\varepsilon\lambda\dot{y}_3+\omega_2^2y_3+d(y_3-y_2)=0 \end{aligned} \tag{2}$$

representing two strongly coupled linear oscillators that are weakly attached to an essentially nonlinear oscillator. In Fig. 2 we depict the transient response of this system for $\lambda=0.5$, $\omega_2^2=0.9$, $C=5.0$, $d=1.0$, $\varepsilon=0.1$, and zero initial conditions except $\dot{y}_3(0)\neq 0$. Whereas for low excitation no energy pumping to the nonlinear oscillator occurs (cf. Fig. 2(*a*)), as $\dot{y}_3(0)$ increases energy pumping takes place (cf. Figs. 2(*b*,*c*)). Hence the nonlinear energy pumping phenomenon can also be realized in multi-degree-of-freedom systems.

We now focus exclusively in the two-degrees-of-freedom system (1). Considering the transient responses depicted in Fig. 1 we note that, when energy pumping occurs the motion can be divided into two phases: In the initial phase energy is rigorously pumped from oscillator 2 to oscillator 1 in a *one-way (irreversible)* transfer, until oscillator 1 reaches a certain amplitude of oscillation; in the second phase of the motion, both oscillators perform decaying oscillations due to damping dissipation with oscillator 1 retaining most of the vibrational energy. Moreover, during the initial energy pumping phase (defined approximately for $0<t<40$ in Fig. 1(*b*), and $0<t<60$ for Fig. 1(*c*), the motion of oscillator 1 is composed of a ''fast'' oscillation with frequency nearly identical to the natural frequency of oscillator 2, and a ''slow'' envelope oscillation. This strongly suggests that a 1-1 *internal resonance* between oscillators 1 and 2 plays an important role in the energy pumping phase, although this still does not explain the one way energy pumping from oscillator 2 to oscillator 1.

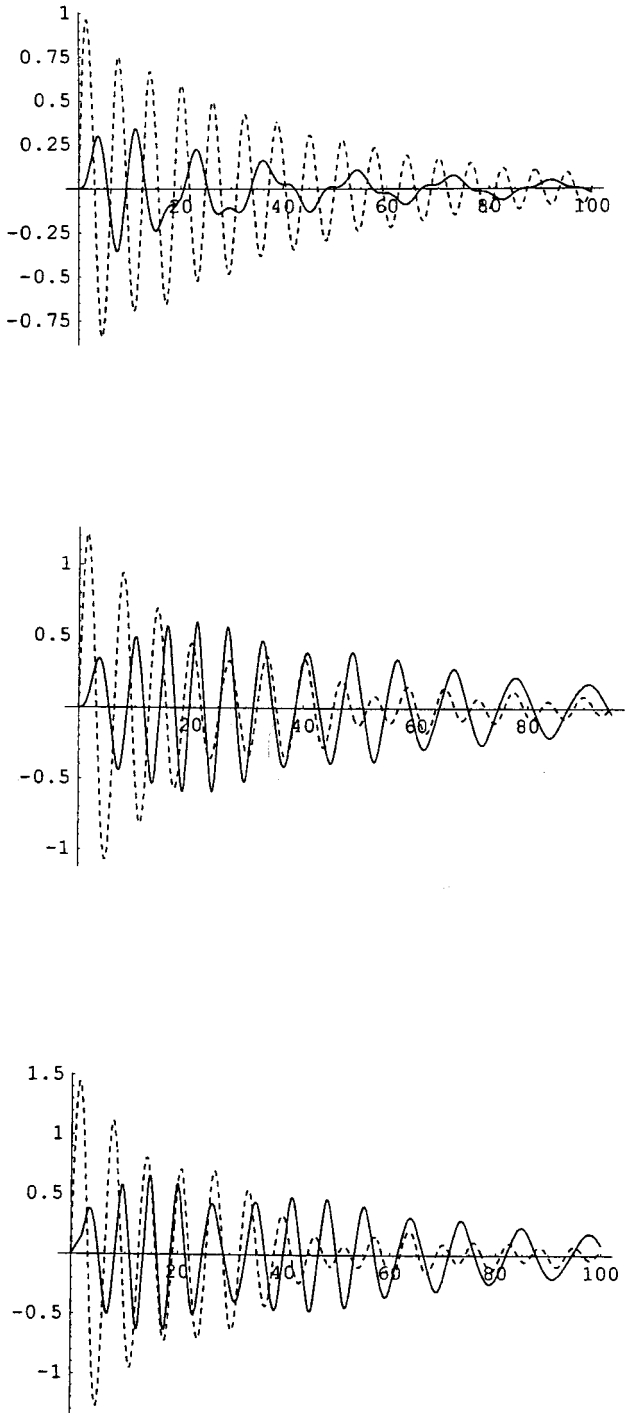


Fig. 1 Numerical transient responses $y_1(t)$ and $y_2(t)$ of system (1) for (*a*) $h=0.5$, (*b*) $h=0.8$, (*c*) $h=1.125$; —— oscillator 1, — — — oscillator 2

Motivated by these observations we now proceed to examine the periodic orbits of the underlying Hamiltonian system by eliminating damping from Eqs. (1). Since system (1) is weakly damped, one expects that, at least at the initial stages of the motion, the dynamics will be greatly influenced by the dynamics of the corresponding (undamped) Hamiltonian system. In turn, the undamped dynamics are dominated by periodic orbits. As a result, we expect that the topological structure of the periodic orbits (and their bifurcations) of the Hamiltonian system, will play a dominant role in the energy pumping phenomenon.

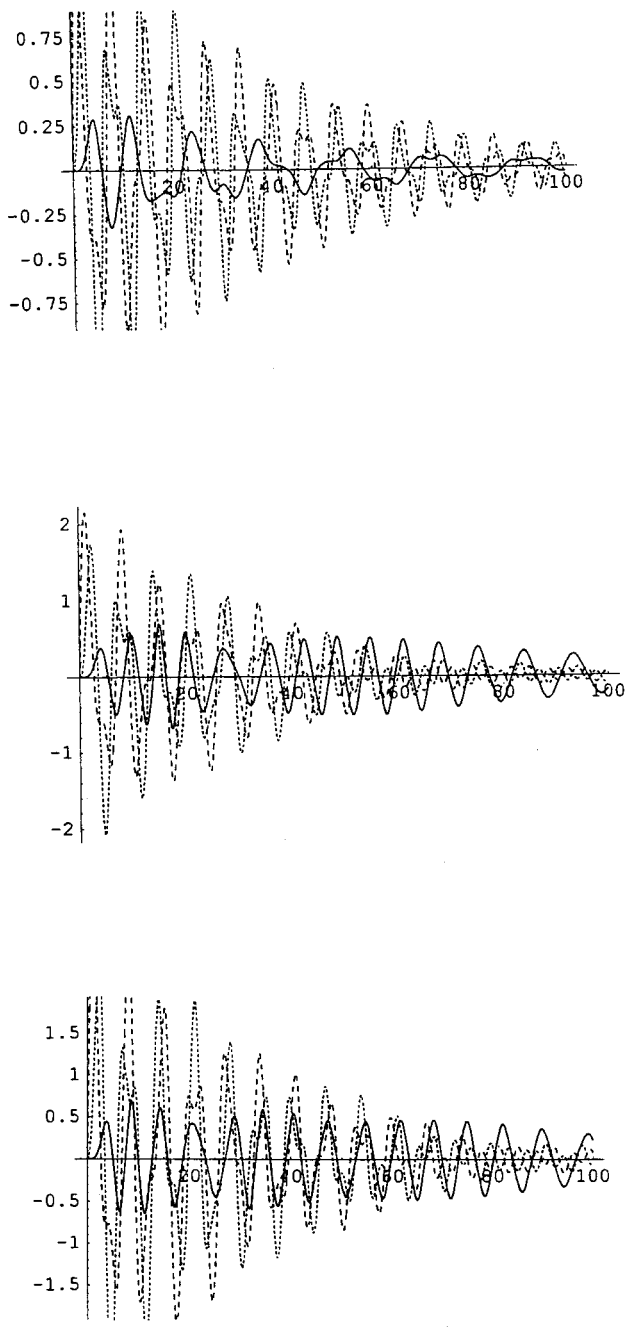


Fig. 2 Numerical transient response of the three-degrees-of-freedom system for (*a*) $\dot{y}_3(0)=2.0$, (*b*) $\dot{y}_3(0)=3.0$, and (*c*) $\dot{y}_3(0)=4.0$; —— oscillator 1, · · · · oscillator 2, — — — oscillator 3

3 Periodic Orbits of the Underlying Hamiltonian System

The underlying Hamiltonian two-degrees-of-freedom system is obtained by setting $\lambda=0$ in (1). At a fixed level of energy (Hamiltonian) we employ the reduction method outlined in [13] to reduce the undamped system (1) to a single-degree-of-freedom nonautonomous oscillator with periodic forcing. This is a standard reduction process by which an $(n+1)$-degree-of-freedom Hamiltonian system with symmetry (symmetry of time translations) is reduced to an n-degree-of-freedom nonautonomous system with no symmetry.

Introducing the action-angle variables $(I_2,\theta_2)\in(R^+\times S^1)$ for oscillator 2 defined by the relations $y_2=\sqrt{2I_2/\omega_2}\sin\theta_2$, $v_2\equiv\dot{y}_2=\sqrt{2I_2\omega_2}\cos\theta_2$, the Hamiltonian of the undamped system is expressed as

$$H^\varepsilon=F(y_1,v_1)+G(I_2)+\frac{\varepsilon}{2}H^1(y_1,v_1,\theta_2,I_2) \tag{3}$$

where

$$F(y_1,v_1)=(v_1^2/2)+(Cy_1^4/4),\quad G(I_2)=\omega_2 I_2,$$
$$H^1(y_1,v_1,\theta_2,I_2)=(y_1-\sqrt{2I_2/\omega_2}\sin\theta_2)^2.$$

The equations of motion can then be placed in the following form:

$$\dot{y}_1=\frac{\partial F}{\partial v_1}+\frac{\varepsilon}{2}\frac{\partial H^1}{\partial v_1},\quad \dot{v}_1=-\frac{\partial F}{\partial y_1}-\frac{\varepsilon}{2}\frac{\partial H^1}{\partial y_1},$$
$$\dot{\theta}_2=\omega_2+\frac{\varepsilon}{2}\frac{\partial H^1}{\partial I_2},\quad \dot{I}_2=-\frac{\varepsilon}{2}\frac{\partial H^1}{\partial\theta_2} \tag{4}$$

where $v_1=\dot{y}_1$. By fixing the Hamiltonian (total energy) to a constant level h, we can express the action I_2 in terms of the other variables of the system as follows:

$$H^\varepsilon=F(y_1,v_1)+G(I_2)+\frac{\varepsilon}{2}H^1(y_1,v_1,\theta_2,I_2)$$
$$=h\Rightarrow I_2=L^\varepsilon(y_1,v_1,\theta_2,h) \tag{5}$$

where L^ε is a complicated expression. As a reviewer pointed out, the inversion (5) is only possible if the system is nonsingular, i.e., if the condition $\partial H^\varepsilon/\partial I_2\neq 0$ is satisfied; clearly, this is the case in our problem. Taking into account (5), eliminating the time variable from (4), and combining the resulting first-order expressions into a single second-order one we obtain the reduced oscillator

$$y_1''+(C/\omega_2^2)y_1^3=\varepsilon\frac{1}{2\omega_2^4}\left\{-2\omega_2^2y_1+4Cy_1^3\sin^2\theta_2\right.$$
$$+\frac{\omega_2(4h-5Cy_1^4)\sin\theta_2}{\sqrt{2h-\frac{Cy_1^4}{2}-\omega_2^2y_1'^2}}+2\omega_2^2y_1'\left[-\sin 2\theta_2\right.$$
$$\left.\left.+\frac{\omega_2y_1\cos\theta_2}{\sqrt{2h-\frac{Cy_1^4}{2}-\omega_2^2y_1'^2}}\right]\right\}+O(\varepsilon^2)$$
$$\equiv\varepsilon g(y_1,y_1',\theta_2)+O(\varepsilon^2). \tag{6}$$

We note that the derived expression is approximate since it neglects $O(\varepsilon^2)$ terms; this approximation was imposed by the impossibility of finding an exact expression for L^ε in (5). As a result, the following analysis is valid only for undamped systems (1) with sufficiently weak coupling. In (6), $y_1=y_1(\theta_2)$, and primes denote differentiation with respect to θ_2. In addition, the ''forcing term'' on the right-hand side is 2π–periodic in θ_2. This completes the reduction process.

Employing the previous analysis, the problem of computing the periodic orbits of the undamped system (1) is equivalent to the problem of computing the periodic solutions of the reduced system (6). This equivalence holds since a periodic motion in y_1 and v_1 under a periodic change in θ_2 leads to a periodic motion for I_2 as well. Since the reduced system is essentially nonlinear we resort to an analytical/numerical technique to compute the periodic orbits and their bifurcations. In particular, we introduce a nonsmooth transformation of variables to transform the problem to a

set of nonlinear boundary value problems over finite domains. This technique was first developed by Pilipchuck [14,15] and then applied to smooth and nonsmooth problems in dynamics in a series of works ([16,17]). We refer the reader to these works for more technical details of the method.

We now compute the periodic solutions of (6) with period $T=4a$ (yet undetermined). We express the solution in the following form:

$$y_1(\theta_2)=X(\tau(\phi))+e(\phi)Y(\tau(\phi)),\quad \phi=\theta_2/a \tag{7}$$

where the new independent variables $\tau(\phi)$ and $e(\phi)$ are bounded nonsmooth functions of their argument ϕ:

$$\tau(\phi)=\frac{2}{\pi}\arcsin\left[\sin\left(\frac{\pi\phi}{2}\right)\right],\quad e(\phi)=\tau'(\phi). \tag{8}$$

The derivative in (8) should be understood in the context of the theory of distributions. Both nonsmooth variables are periodic in ϕ with (normalized) period equal to 4. We note that by (7) the solution is expressed in terms of two components. *X* depends only on τ and is termed the *R-component* of the solution. *Y* also depends solely on τ and is multiplied by *e*; it is termed the *I-component* of the solution. Interestingly, expression (7) has a phenomenological resemblance to complex variable representation with *e* playing the role of the imaginary constant *j* (note that $e^2=-j^2=1$).

Employing the transformation (7) we express the derivatives and powers of y_1 in (6) in terms of *X* and *Y*, and set separately the *R* and *I*-components of the resulting expression equal to zero. We then obtain the following two subproblems governing the *R* and *I*-components of the solution:

Subproblem 1.

$$Y''+(Ca^2/\omega_2^2)Y^3=\varepsilon a^2 g(y_1=Y,y_1'=Y'/a,\theta_2=m\pi\tau)+O(\varepsilon^2),$$
$$a=m\pi,\quad X=0\quad Y(\pm1)=0,\quad m=1,2,3,\ldots. \tag{9}$$

Subproblem 2.

$$X''+(Ca^2/\omega_2^2)X^3=\varepsilon a^2 g\left(y_1=X,y_1'=X'/a,\theta_2=\frac{(2n-1)\pi\tau}{2}\right)+O(\varepsilon^2),$$
$$a=\frac{(2n-1)\pi}{2},\quad Y=0\quad X'(\pm1)=0,\quad n=1,2,3,\ldots. \tag{10}$$

We note that the above subproblems were obtained by setting either the *R* or *I*-component of the solution equal to zero. Then, the solutions of each subproblem provide a distinct class of subharmonic motions of the problem. In general, the problem obtained by applying the previous method leads to a coupled system of equations in *X* and *Y*, however, this case will not be considered here.

Since no analytical solution exists for these nonlinear boundary value problems we need to resort to a numerical method to solve them. Before we perform this numerical computation, however, we make the following remarks concerning the method of nonsmooth transformations. The boundary conditions in (9) and (10) impose smoothness on the transformed derivatives of y_1; these boundary conditions define the domain of the solutions of the nonlinear boundary value problems $\tau\in[-1,1]$. In addition, the quarter-period of the solution, *a*, for each subproblem is allowed a countable infinity of values. Taking into account that the period of the periodic solution is $T=4a$, and that the nonhomogeneous term in (6) is 2π–periodic in θ_2, we conclude that *Subproblem 1 computes the* $2m$:1, $m=1,2,3,\ldots$ *subharmonic orbits of* (6), *whereas Subproblem 2 computes the* $(2n-1)$:1, $n=1,2,3,\ldots$ *subharmonic orbits*. A subharmonic orbit of order p:1 is a periodic orbit with period equal to $2p\pi$, i.e., *p*-times the period of the nonhomogeneous term *g*. Finally, we note that the periodic solution $y_1(t)$ is obtained from the solutions of the nonlinear boundary value problems either as $y_1(\theta_2)=e(\theta_2/a)Y(\tau(\theta_2/a))$ (Subproblem 1), or as $y_1(\theta_2)=X(\tau(\theta_2/a))$ (Subproblem 2), where $\theta_2=\omega_2 t+\theta_{20}+O(\varepsilon)$. Now, the nonlinear boundary value problems above provide the solution only in the normalized half-period $\tau\in[-1,1]$. To extend the result over a full normalized period (equal to 4) we need to add the component of the solution in the interval $\tau\in[1,3]$; to perform this we take into account the symmetry properties of the nonsmooth variables τ and *e*, and add either the antisymmetric image of the solution about the point $(Y,\tau)=(0,1)$ (for Subproblem 1, cf. Fig. 3(*a*)), or the mirror image of the solution about the line $\tau=1$ (for Subproblem 2, cf. Fig. 3(*b*)).

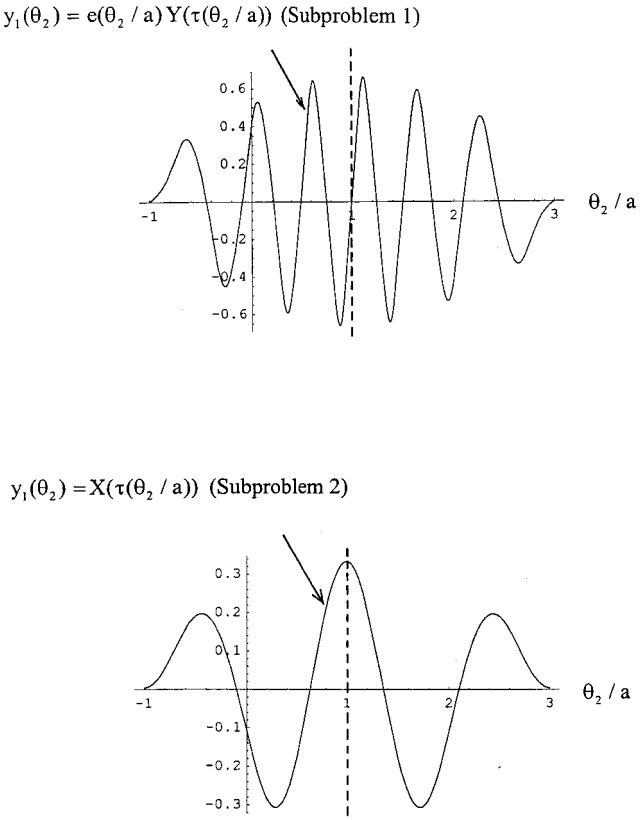


Fig. 3 Construction of the solution y_1 over an entire normalized period (equal to 4) from the half-normalized period solutions (*a*) $eY(\tau)$, and (*b*) $X(\tau)$

The nonlinear boundary value problems (9) and (10) were solved using a single-point numerical shooting method. In Figs. 4 and 5 we depict the leading low-order subharmonic orbits and their bifurcations for the undamped two-degrees-of-freedom system with $\omega_2^2=0.9$, $C=5.0$, $\varepsilon=0.1$ and varying values of the total energy *h*. In these figures we also present one-period representations of a number of subharmonic orbits. In the bifurcation plot of Fig. 4 we depict the values of $Y'(-1)$ at the subharmonic orbits as functions of *h* (recall that the solution domain of the above nonlinear boundary value problems is $-1\leqslant\tau\leqslant1$); in physical terms, *each point denotes the initial slope* ay_1' *of the subharmonic orbit, corresponding to zero initial displacement*, $y_1=0$. In Fig. 5 we plot $X(-1)$ as function of *h*; in physical terms, *each point denotes the initial displacement* y_1 *of the subharmonic orbit, corresponding zero initial slope*, $y_1'=0$. These plots depict only the responses of the unexcited oscillator 1; the corresponding responses of oscillator 2 are computed using relation (5) derived in

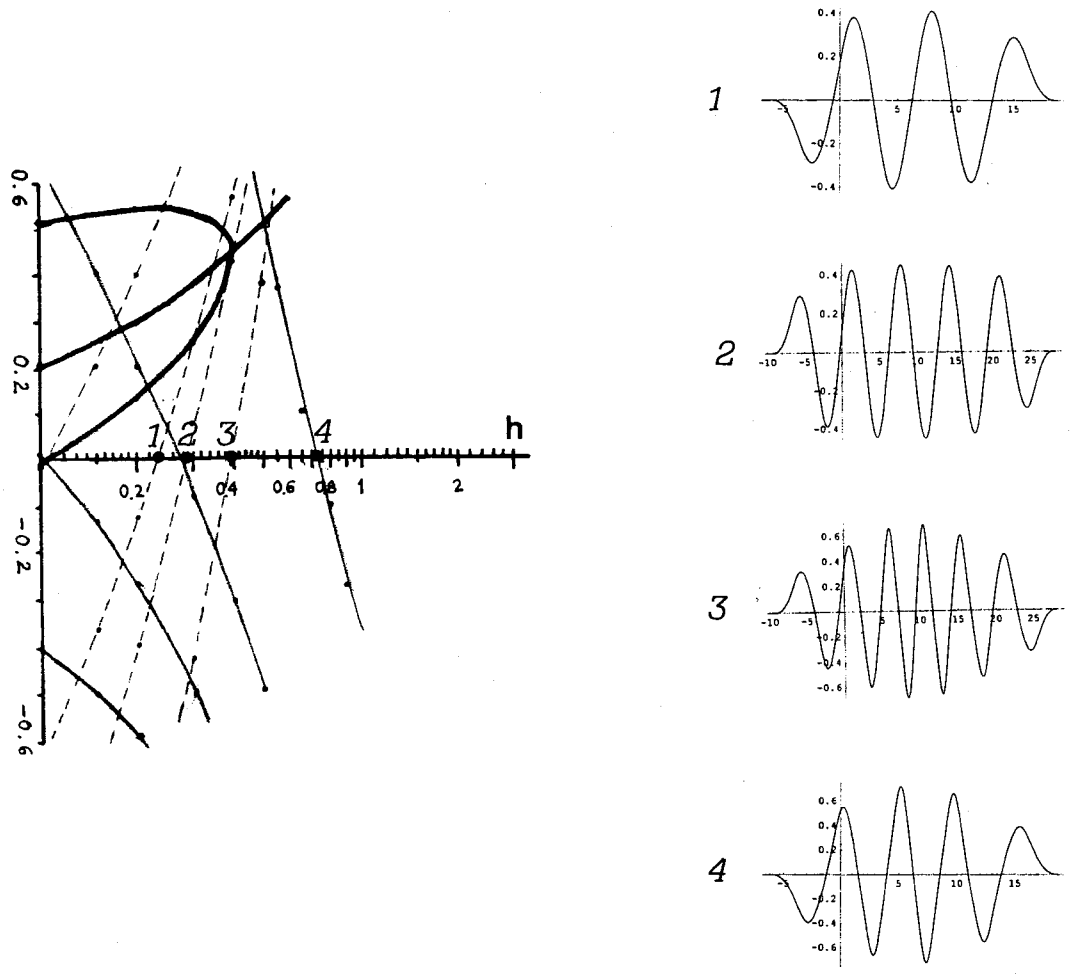


Fig. 4 Leading $2m$:1 subharmonic orbits as functions of h: ——$Y'(-1)$ for $m=1$, ——$Y'(-1)$ for $m=2$, — — —$Y'(-1)$ for $m=3$

the course of the reduction process, and the relations $y_2 = \sqrt{2I_2/\omega_2}\sin\theta_2$, $\theta_2 = \omega_2 t + \theta_{20} + O(\varepsilon)$. Finally, we emphasize that, since the reduced system (6) neglects $O(\varepsilon^2)$ terms, the results presented in Figs. 4 and 5 are approximate and valid only for sufficiently small values of ε. Of special interest are the 1:1 subharmonic orbits labeled A-D in Fig. 5. These orbits dominate the dynamics as shown below.

The domain of attraction and the stability of the 1:1 subharmonic orbits were determined by numerical Poincaré maps. These were constructed by considering the original undamped Eqs. (1). First, the four-dimensional phase space of the solutions of (1) was reduced to a three-dimensional isoenergetic manifold $\aleph$ by fixing the total energy to a constant level, $H^{\varepsilon}(y_1,\dot{y}_1,y_2,\dot{y}_2)=h$; $\aleph$ was then ''cut'' by the Poincaré section $\Sigma=\{(y_1,\dot{y}_1,y_2)\in\aleph/y_2=0,\dot{y}_2>0\}$. The Poincaré map P^{ε} was defined as, $P^{\varepsilon}:\Sigma\rightarrow\Sigma$, i.e., as a mapping of points on Σ to their images under the flow of the dynamical system on Σ, under orientation preserving restrictions. Stable periodic orbits of (1) appear as centers in the Poincaré map, whereas, unstable periodic orbits appear as saddle points.

In Fig. 6 we depict the Poincaré maps of the undamped system (1) with $C=5.0$, $\varepsilon=0.1$ and varying values of the energy h. The 1:1 subharmonic orbits labeled A-D correspond to the ones of the approximate bifurcation plot of Fig. 5. A common feature of all these plots is a large region of regular motion (smooth quasi-periodic orbits) in the upper regions of the plots, surrounding the stable 1:1 subharmonic orbit A. When damping is added to the system, the large regular region surrounding orbit A is expected to become a large region of attraction, with that orbit becoming an attractor. Additional stable and unstable 1:1 subharmonic orbits of the system are indicated in the Poincaré maps, confirming the approximate asymptotic results of Fig. 5. At the small energy level $h=0.05$ there are two stable subharmonic orbits; both orbits correspond to localized motions, with orbit A localizing in oscillator 1 and orbit B in oscillator 2. At higher values of h the low-energy bifurcation of 1:1 subharmonic orbits (predicted in the plot of Fig. 5) has occurred and there exist four orbits, three stable and one unstable. Note that as h increases orbit A gradually delocalizes from oscillator 1 and localizes in oscillator 2. The regions of chaotic motion (the ''stochastic sea'') in the maps is a well-documented feature in the dynamics of such strongly nonlinear systems.

The bifurcation diagrams and Poincaré maps of the 1:1 subharmonic orbits lead to a preliminary qualitative explanation of the energy pumping phenomenon, which, as shown in the previous section, occurs only above a certain level of the initial energy h. When energy pumping occurs, an initial transfer of energy occurs from the directly excited (linear) oscillator 2 to the unexcited (nonlinear) oscillator 1; moreover, the ''fast'' oscillation during this initial phase of the motion has a frequency nearly identical to the linearized natural frequency ω_2. Hence, it is logical to conclude *that the 1:1 subharmonic orbit A existing over the entire range of h and having a large domain of attraction, is mainly*

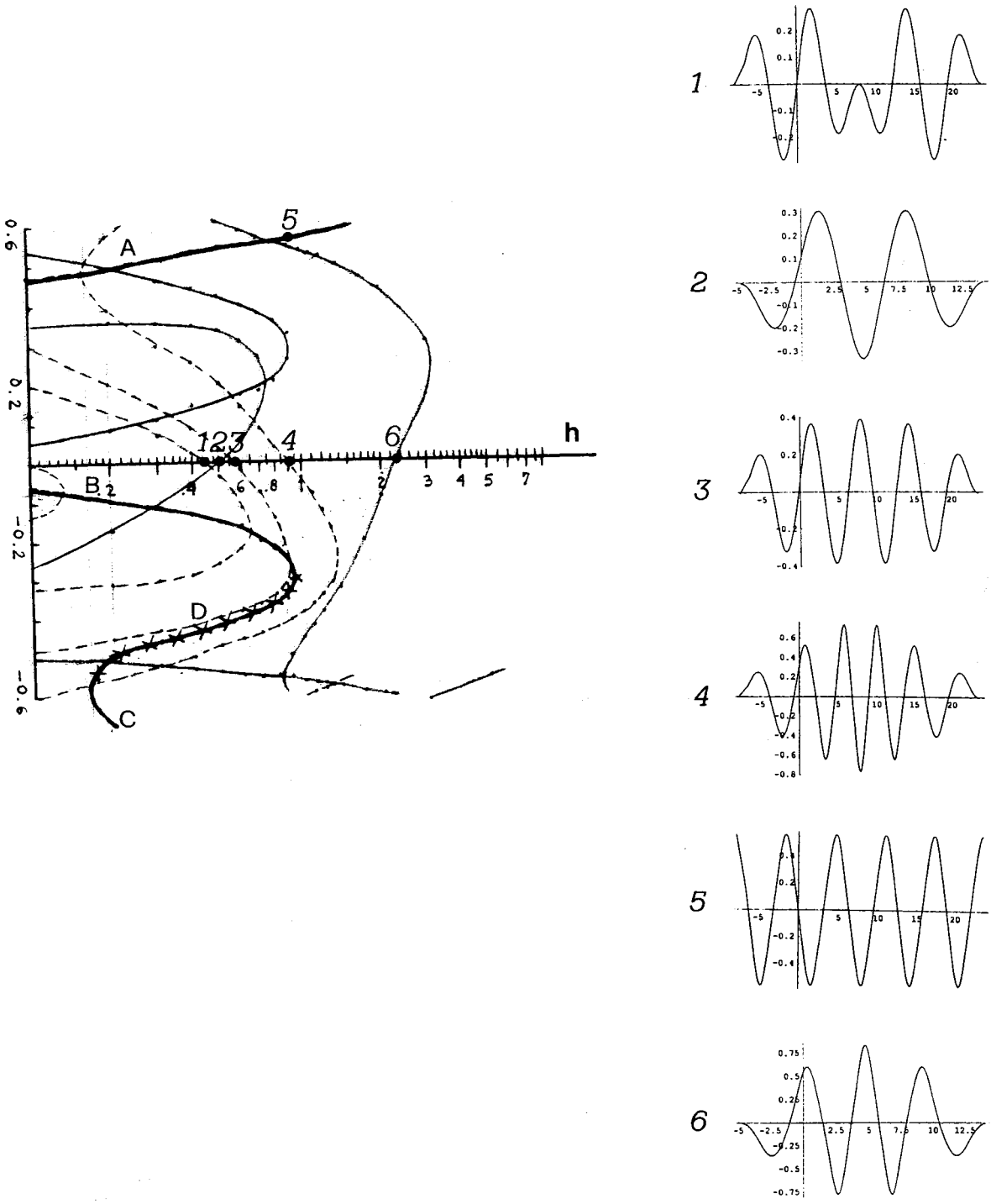


Fig. 5 Leading $(2n-1)$:1 subharmonic orbits as functions of h: ——$X(-1)$ for $n=1$, ——$-X(-1)$ for $n=2$, — — —$X(-1)$ for $n=3$; ×××× unstable 1:1 subharmonic orbits

responsible for the energy pumping phenomenon. However, this family of orbits can not be directly excited at $t=0$ since it cannot satisfy pointwise the initial condition $((y_1(0),\dot{y}(0))=(0,0)$ (this is the initial state of oscillator 1 when the energy pumping phenomenon is initiated); as a result, a *transient ''bridging'' orbit* must be initially excited, satisfying zero initial conditions and ultimately ''connecting'' with the 1:1 subharmonic orbit A. Under these conditions energy pumping occurs. Noting that the amplitude of the 1:1 orbit has a lower bound of approximately 0.47 (cf. Fig. 5) and considering the initial transients of the numerical simulations of Fig. 1, we conjecture that if the initial transient transients of oscillator 1 can not attain sufficiently large amplitudes, they cannot act as ''bridging'' orbits to excite the 1:1 subharmonic orbit A and no energy pumping can take place. Thus, for sufficiently low values of h no energy pumping is possible. For higher values of h the initial transients for oscillator 1 attain sufficiently large amplitudes to excite the 1:1 orbit and energy pumping can occur. This conjecture explains the lack of energy pumping for $h=0.5$ in the simulations of Fig. 1. Clearly, the previous arguments form merely a conjecture, but the more rigorous analysis in Part II of this work ([1]) validates these arguments.

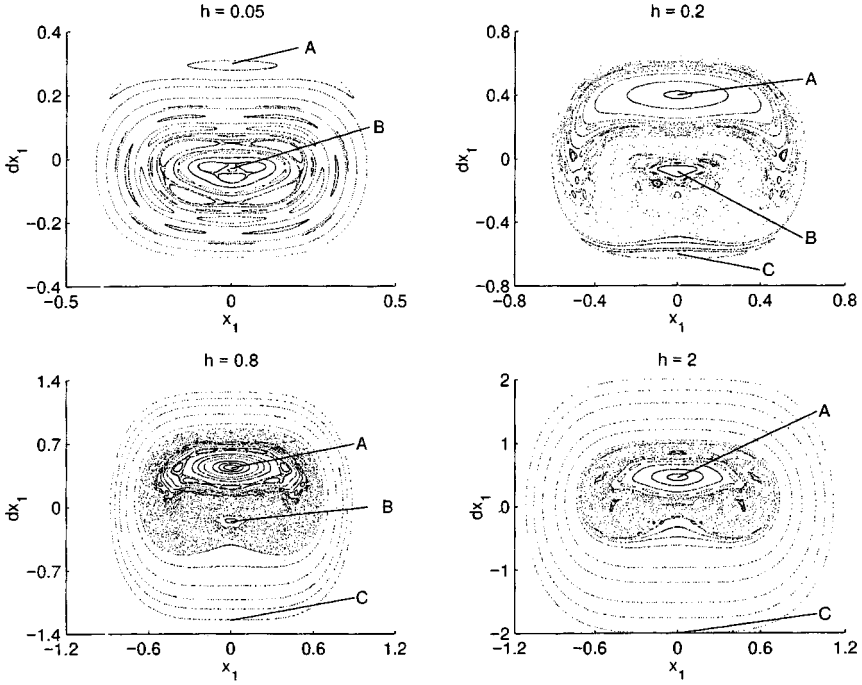


Fig. 6 Poincaré maps of the dynamics of the undamped system (1) at varying energy levels for $\omega_2^2=0.9$, $C=5.0$, $\varepsilon=0.1$: (*a*) $h=0.05$, (*b*) $h=0.2$, (*c*) $h=0.8$, (*d*) $h=2.0$

4 Analytical Approximations

As a final note, we now present an analytical technique to approximate the transient responses of the Hamiltonian system; in contrast to most standard techniques which are based on the assumption of weak nonlinearity, the method used here deals with the strong (nonlinearizable) nonlinearity of oscillator 1. An extension of this technique for the damped system in Part II of this work ([1]) will enable us to analytically approximate the transient responses during the initial phase of energy pumping of Fig. 1.

To this end, we express system (1) in the following form:

$$\begin{aligned} \ddot{y}_1+\varepsilon y_1+Cy_1^3-\varepsilon y_2=0 \\ \ddot{y}_2+\omega^2 y_2-\varepsilon y_1=0 \end{aligned} \tag{11}$$

where $\omega^2=\omega_2^2+\varepsilon$. A transformation to complex variables is now introduced,

$$\psi_1=\dot{y}_1+j\omega y_1, \quad \psi_2=\dot{y}_2+j\omega y_2, \tag{12}$$

and (12) are rewritten as

$$\begin{aligned} \dot{\psi}_1-\frac{j\omega}{2}(\psi_1+\psi_1^*)-\frac{j\varepsilon}{2\omega}(\psi_1-\psi_1^*)+\frac{jC}{8\omega^3}(\psi_1-\psi_1^*)^3 \\ +\frac{j\varepsilon}{2\omega}(\psi_2-\psi_2^*)=0 \\ \dot{\psi}_2-j\omega\psi_2+\frac{j\varepsilon}{2\omega}(\psi_1-\psi_1^*)=0 \end{aligned} \tag{13}$$

where the star denotes complex conjugate.

An approximate solution of (13) is sought, based on the assumption of fast oscillations at frequency ω:

$$\psi_1=\varphi_1 e^{j\omega t}, \quad \psi_2=\varphi_2 e^{j\omega t}. \tag{14}$$

Relations (14) indicate the presence of 1:1 internal resonance in the fast dynamics of oscillators 1 and 2. Substituting (14) into (13), and averaging over the fast periodic terms $e^{j\omega t}$ we obtain the following set of averaged equations governing the (complex) amplitudes φ_i, $i=1,2$:

$$\begin{aligned} \dot{\varphi}_1+\frac{j}{2}\left(\omega-\frac{\varepsilon}{\omega}\right)\varphi_1-\frac{3jC}{8\omega^3}|\varphi_1|^2\varphi_1+\frac{j\varepsilon}{2\omega}\varphi_2=0 \\ \dot{\varphi}_2+\frac{j\varepsilon}{2\omega}\varphi_1=0. \end{aligned} \tag{15}$$

Interestingly, in contrast to (11), the transformed system (15) is completely integrable, with the following two first integrals of motion:

$$\begin{aligned} |\varphi_1|^2+|\varphi_2|^2=N^2, \\ \frac{j\omega}{2}|\varphi_1|^2-\frac{3jC}{16\omega^3}|\varphi_1|^4+\frac{j\varepsilon}{2\omega}(\varphi_1\varphi_2^*+\varphi_1^*\varphi_2)=H. \end{aligned} \tag{16}$$

Employing these results, the complex amplitudes are expressed as

$$\varphi_1=N\sin\theta e^{j\delta_1}, \quad \varphi_2=N\cos\theta e^{j\delta_2}. \tag{17}$$

Substituting (17) into (15), we obtain the final set of equations on the 2-Torus governing the angle-variables θ and $\delta=\delta_1-\delta_2$:

$$\begin{aligned} \dot{\delta}+\frac{\omega}{2}-\frac{3CN}{8\omega^3}\sin^2\theta+\frac{\varepsilon}{\omega}\cot 2\theta\cos\delta=0 \\ \dot{\theta}+\frac{\varepsilon}{2\omega}\sin\delta=0. \end{aligned} \tag{18}$$

We note that the orbits of (18) can be analytically computed in terms of pseudo-elliptic quadratures by employing the change of variables (17) in (16), and then integrating the second of Eqs. (18) by quadratures.

By numerically integrating (18) we can study transient (nonperiodic) orbits in the neighborhoods of the 1:1 subharmonic orbits of the underlying Hamiltonian system. In Fig. 7 we present the phase plots of (18) for varying values of the energy-like first integral *N*, confirming the bifurcations of the 1:1 subharmonic orbits depicted in the analytical approximations of Fig. 5 and the numerical results of Fig. 6.

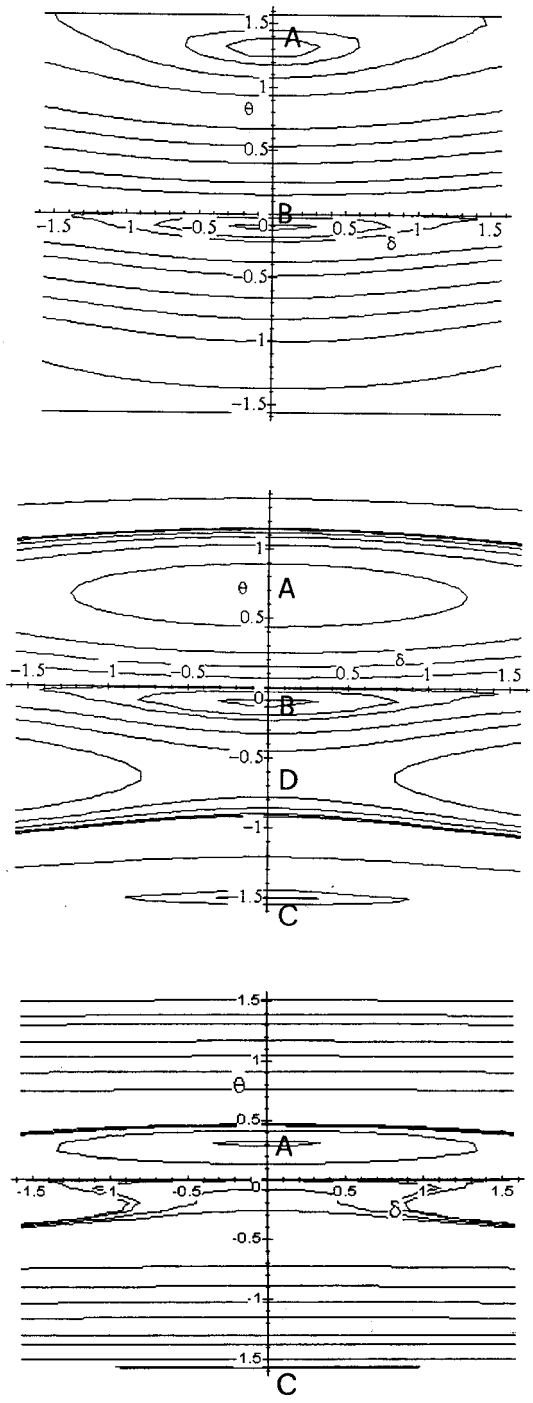


Fig. 7 Phase plots of the system of Eqs. (18) for $\omega_2^2=0.9$, $C=5.0$, $\varepsilon=0.1$, and varying values of the first integral of motion N: (*a*) $N=0.4$, (*b*) $N=0.8$, (*c*) $N=1.9$

5 Discussion

We presented numerical evidence of energy pumping in nonlinear mechanical oscillators. The systems considered herein were composed of weakly coupled, linear, and essentially nonlinear (nonlinearizable) parts. In such systems it is possible to induce one-way (irreversible) ''channeling'' of vibrational energy from the linear to the nonlinear part, provided that the imparted energy is above a critical level. Hence, energy can be ''pumped'' to a predetermined part of the system (the nonlinear oscillator), which, in essence acts as a passive nonlinear sink. Clearly, no such phenomenon is possible in linear systems.

A 1:1 stable subharmonic orbit of the underlying Hamiltonian system (obtained by setting damping equal to zero) was conjectured to be responsible for the energy pumping phenomenon. We conjecture that the reason for lack of energy pumping at low energies is due to the fact that the 1:1 subharmonic orbit can not be excited unless the energy of the system is above a critical level. The energy pumping phenomenon will be further studied in a companion paper ([1]) where it will be shown that it is caused by transient *resonance capture* on a 1:1 resonance manifold of the system.

We remark that the energy pumping phenomenon, which in this work was purely passive, could be enhanced using active control. Utilizing this phenomenon one can introduce passive or active nonlinear sinks in predominantly linear extended periodic structures where externally imparted energy is directed and locally eliminated. This can lead to enhanced vibration and shock isolation designs of extended mechanical systems.

Acknowledgments

This work was supported in part by NSF Young Investigator Award CMS-94-57750 (Dr. Allison Flatau is the Grant Monitor), and by an International Supplement to this Award (Ms. Cassandra Dudca is the Grant Monitor).

References

[1] Vakakis, A. F., and Gendelman, O., 2001, ''Energy 'Pumping' in Coupled Mechanical Oscillators II: Resonance Capture,'' ASME J. Appl. Mech., **68**, pp. 42–48.
[2] Hodges, C. H., 1982, ''Confinement of Vibration by Structural Irregularity,'' J. Sound Vib., **82**, No. 3, pp. 411–424.
[3] Pierre, C., and Dowell, E. H., 1987, ''Localization of Vibrations by Structural Irregularity,'' J. Sound Vib., **114**, No. 3, pp. 549–564.
[4] Bendiksen, O. O., 1987, ''Mode Localization Phenomena in Large Space Structures,'' AIAA J., **25**, No. 9, pp. 1241–1248.
[5] Pierre, C., and Cha, P., 1989, ''Strong Mode Localization in Nearly Periodic Disordered Structures,'' AIAA J., **27**, No. 2, pp. 227–241.
[6] Photiadis, D. M., 1992, ''Anderson Localization of One-Dimensional Wave Propagation on a Fluid-Loaded Plate,'' J. Acoust. Soc. Am., **91**, No. 2, pp. 771–780.
[7] Vakakis, A. F., and Cetinkaya, C., 1993, ''Mode Localization in a Class of Multi-Degree-of-Freedom Nonlinear Systems With Cyclic Symmetry,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **53**, pp. 265–282.
[8] Vakakis, A. F., Nayfeh, T. A., and King, M. E., 1993, ''A Multiple-Scales Analysis of Nonlinear Localized Modes in a Cyclic Periodic System,'' ASME J. Appl. Mech., **60**, No. 2, pp. 388–397.
[9] Vakakis, A. F., Manevitch, L. I., Mikhlin, Yu., Pilipchuck, V., and Zevin, A., 1996, *Normal Modes and Localization in Nonlinear Systems*, John Wiley and Sons, New York.
[10] Nayfeh, A. H., and Mook, D., 1984, *Nonlinear Oscillations*, John Wiley and Sons, New York.
[11] Nayfeh, S. A., and Nayfeh, A. H., 1994, ''Energy Transfer From High to Low-Frequency Modes in a Flexible Structure via Modulation,'' ASME J. Vibr. Acoust., **116**, pp. 203–207.
[12] Gendelman, O., 1999, ''Transition of Energy to Nonlinear Localized Mode in a Highly Asymmetric System of Two Oscillators,'' Nonlinear Dyn., submitted for publication.
[13] Holmes, P. J., and Marsden, J. E., 1982, ''Horseshoes in Perturbations of Hamiltonian Systems With Two Degrees of Freedom,'' Commun. Math. Phys., **82**, pp. 523–544.
[14] Pilipchuck, V. N., 1985, ''The Calculation of Strongly Nonlinear Systems Close to Vibration-Impact Systems,'' Prikl. Mat. Meck. (PMM), **49**, No. 6, pp. 572–578.
[15] Pilipchuck, V. N., 1988, ''A Transformation for Vibrating Systems Based on a Non-Smooth Periodic Pair of Functions,'' Dokl. Akad. Nauk SSSR, Ser. A, **4**, pp. 37–40 (in Russian).
[16] Pilipchuck, V. N., Vakakis, A. F., and Azeez, M. A. F., 1997, ''Study of a Class of Subharmonic Motions Using a Non-Smooth Temporal Transformation,'' Physica D, **100**, pp. 145–164.
[17] Vakakis, A. F., and Atanackovic, T. M., 1999, ''Buckling of an Elastic Ring Forced by a Periodic Array of Compressive Loads,'' ASME J. Appl. Mech., **66**, pp. 361–367.

Energy Pumping in Nonlinear Mechanical Oscillators: Part II—Resonance Capture

A. F. Vakakis
Department of Mechanical and Industrial Engineering,
University of Illinois,
1206 W. Green Street,
Urbana, IL 61801
e-mail: avakakis@uiuc.edu

O. Gendelman
Institute of Chemical Physics,
Russian Academy of Sciences,
Kosygin Street 4,
117977 Moscow, Russia
e-mail: ovgend@center.chph.ras.ru

We study energy pumping in an impulsively excited, two-degrees-of-freedom damped system with essential (nonlinearizable) nonlinearities by means of two analytical techniques. First, we transform the equations of motion using the action-angle variables of the underlying Hamiltonian system and bring them into the form where two-frequency averaging can be applied. We then show that energy pumping is due to resonance capture in the 1:1 resonance manifold of the system, and perform a perturbation analysis in an $O(\sqrt{\varepsilon})$ neighborhood of this manifold in order to study the attracting region responsible for the resonance capture. The second method is based on the assumption of 1:1 internal resonance in the fast dynamics of the system, and utilizes complexification and averaging to develop analytical approximations to the nonlinear transient responses of the system in the energy pumping regime. The results compare favorably to numerical simulations. The practical implications of the energy pumping phenomenon are discussed.
[DOI: 10.1115/1.1345525]

1 Introduction

In this paper we extend the results presented in [1] on energy pumping in weakly coupled nonlinear oscillators. The effects of damping are added to the analysis, and it is shown that energy pumping is caused by resonance capture on a 1:1 resonance manifold. The phenomenon of resonance capture occurs in nonconservative oscillators and leads to transient capture of trajectories in a domain of attraction on the resonance manifold.

Resonance capture, as well as single and multifrequency averaging techniques for analyzing nonlinear oscillators have been studied in previous publications (for example, [2–4]). General theorems on resonance capture in two-frequency systems (such as the ones considered herein) were given in the aforementioned references and in [5,6]. In [7–9] resonance capture in perturbed two-dimensional Hamiltonians is studied by perturbation techniques. Additional asymptotic techniques for analyzing transient resonant layers (passage through resonance) are given in [10–16].

In this work we study energy pumping by employing two analytical techniques. We note that the two-degrees-of-freedom system considered herein is *strongly nonlinear* and *weakly damped*; as a result, standard perturbation methods that are valid for weakly nonlinear systems are not applicable in this case. First, we transform the strongly nonlinear, weakly damped equations of motion into a system of four first-order equations in $(R^+\times R^+\times S^1\times S^1)$ using the action-angle variables of the underlying Hamiltonian system. The resulting equations are in a form amenable to multifrequency averaging, and resonance capture analysis. We then show that the transformed system satisfies the conditions for resonance capture on a 1:1 resonance manifold. In the second methodology followed in this work we extend the perturbation method based on complexification of the equations of motion developed in [1] for the weakly damped case, and provide analytic reconstructions of the transient responses of the system during energy pumping. We conclude with a discussion of the practical implications of the results of this work.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Sept. 29, 1999; final revision, May 2, 2000. Associate Editor: N. C. Perkins. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 1:1 Resonance Capture Analysis

We consider the following two-degrees-of-freedom system composed of two weakly coupled, weakly damped oscillators:

$$\begin{aligned}\ddot{y}_1+\varepsilon\lambda\dot{y}_1+Cy_1^3+\varepsilon(y_1-y_2)&=0\\ \ddot{y}_2+\varepsilon\lambda\dot{y}_2+\omega_2^2y_2+\varepsilon(y_2-y_1)&=0.\end{aligned}\tag{1}$$

The linear oscillator is labeled ''oscillator 2,'' whereas the (strongly) nonlinear one ''oscillator 1.'' This system was numerically integrated in [1] with initial conditions $y_1(0)=y_2(0)=0$, $\dot{y}_1(0)=0$, $\dot{y}_2(0)=\sqrt{2h}$, corresponding to impulsive loading of oscillator 2. It was shown that for sufficiently high values of h (impulse) energy pumping occurs: Vibrational energy ''flows'' to the unexcited oscillator 1 in an irreversible way. After energy pumping, the motion of the two oscillators is dissipated due to damping.

To analyze the energy pumping phenomenon in the strongly nonlinear system (1) we first transform the equations of motion utilizing the action-angle variables of the underlying Hamiltonian system (corresponding to $\lambda=0$). These are given by [17]:

$$\begin{aligned}y_1&=\Lambda I_1^{1/3}cn\left[\frac{2K(1/2)\theta_1}{\pi},\frac{1}{2}\right]\\ v_1\equiv\dot{y}_1&=-\Lambda I_1^{1/3}\Omega_1(I_1)\frac{2K(1/2)}{\pi}\\ &\quad\times sn\left[\frac{2K(1/2)\theta_1}{\pi},\frac{1}{2}\right]dn\left[\frac{2K(1/2)\theta_1}{\pi},\frac{1}{2}\right]\\ y_2&=\sqrt{\frac{2I_2}{\omega_2}}\sin\theta_2\\ v_2\equiv\dot{y}_2&=\sqrt{2I_2\omega_2}\cos\theta_2\end{aligned}\tag{2}$$

where $\Omega_1(I_1)=\Xi I_1^{1/3}$ is the instantaneous frequency of free oscillation of oscillator 1, $K(1/2)$ is the complete elliptic integral of the first kind, and the variables Λ and Ξ are defined as

$$\Lambda=\left(\frac{1}{4C}\right)^{1/6}\left(\frac{3\pi}{K(1/2)}\right)^{1/3},\quad \Xi=\left(\frac{3\pi^4C}{8K^4(1/2)}\right)^{1/3}.$$

The action angle variables $(I_1, I_2, \theta_1, \theta_2) \in (R^+ \times R^+ \times T^2)$, (where T^2 is the 2-torus) can be regarded as nonlinear polar coordinates for the underlying Hamiltonian system.

Introducing the transformations $(y_1, \dot{y}_1) \rightarrow (I_1, \theta_1)$ and $(y_2, \dot{y}_2) \rightarrow (I_2, \theta_2)$ into Eqs. (1), and expressing them as a set of four first-order equations, we obtain

$$\begin{aligned}
\dot{I}_1 &= \varepsilon \frac{3 I_1^{1/3} \pi}{2K(1/2)\Lambda \Xi [c n^4 + 2 s n^2 d n^2]} \\
&\quad \times \left\{ -\lambda \Lambda I_1^{1/3} \Omega_1(I_1) \frac{2K(1/2)}{\pi} s n^2 d n^2 + \Lambda I_1^{1/3} c n s n d n \right. \\
&\quad \left. - \sqrt{\frac{2I_2}{\omega_2}} \sin\theta_2 s n d n \right\} \equiv \varepsilon \tilde{F}_1(I_1, I_2, \theta_1, \theta_2) \\
\dot{I}_2 &= -\varepsilon \left\{ 2\lambda I_2 \cos^2\theta_2 + \sqrt{\frac{2I_2}{\omega_2}} \cos\theta_2 \left[\sqrt{\frac{2I_2}{\omega_2}} \sin\theta_2 - \Lambda I_1^{1/3} c n \right] \right\} \\
&\equiv \varepsilon \tilde{F}_2(I_1, I_2, \theta_1, \theta_2) \\
\dot{\theta}_1 &= \Omega_1(I_1) + \varepsilon \left[\Xi \Lambda I_1^{2/3} \left(\frac{2K(1/2)}{\pi} \right)^2 (c n^4 + 2 s n^2 d n^2) \right]^{-1} \\
&\quad \times \left\{ -\lambda \Lambda I_1^{1/3} \Omega_1(I_1) \frac{2K(1/2)}{\pi} c n s n d n + \Lambda I_1^{1/3} c n^2 \right. \\
&\quad \left. - \sqrt{\frac{2I_2}{\omega_2}} \sin\theta_2 c n \right\} \equiv \Omega_1(I_1) + \varepsilon \tilde{G}_1(I_1, I_2, \theta_1, \theta_2) \\
\dot{\theta}_2 &= \omega_2 + \varepsilon \frac{\sin\theta_2}{\sqrt{2I_2\omega_2}} \left\{ \lambda \sqrt{2I_2\omega_2} \cos\theta_2 + \sqrt{\frac{2I_2}{\omega_2}} \sin\theta_2 - \Lambda I_1^{1/3} c n \right\} \\
&\equiv \omega_2 + \varepsilon \tilde{G}_2(I_1, I_2, \theta_1, \theta_2).
\end{aligned} \tag{3}$$

In the expressions above, the arguments of all elliptic functions are given by $[2K(1/2)\theta_1/\pi, 1/2]$. We note that by construction the expressions on the right sides of expressions (3) are 2π–periodic in the angle variables θ_1 and θ_2, and the action variables are positive real numbers.

Equations (3) represent a two-frequency dynamical system in $(R^+ \times R^+ \times T^2)$, and are in a form that is directly amenable to two-frequency averaging ([3]). By applying straightforward averaging we obtain the following simplified system:

$$\begin{aligned}
\dot{J}_1 &= \varepsilon \hat{F}_1(J_1, J_2) = -\varepsilon \lambda J_1 \\
\dot{J}_2 &= \varepsilon \hat{F}_2(J_1, J_2) = -\varepsilon \lambda J_2
\end{aligned} \quad \text{(two-frequency averaged system)} \tag{4}$$

where $\hat{F}_p(J_1, J_2) \equiv 1/4\pi^2 \int_0^{2\pi} \int_0^{2\pi} \tilde{F}_p(J_1, J_2, \theta_1, \theta_2) d\theta_1 d\theta_2$, and $J_p > 0$, $p = 1,2$. The conditions under which the dynamics of the averaged system (4) accurately describes the dynamics of the full system (3) has been addressed in previous works (for example, [2,5,9]). Arnold's theorem ([2]) answers this question.

Theorem ([2]). If system (3) satisfies the following condition A,

$$\frac{d}{dt}\left(\frac{\Omega(I_1)}{\omega_2} \right) \neq 0 \quad \text{(along trajectories of the system)}$$

then the difference between the slow motion $(I_1(t), I_2(t))$ of the perturbed system (3) and $(J_1(t), J_2(t))$ of the averaged system (4), remains small over time $(1/\varepsilon)$:

$$\| I(t) - J(t) \| \leq \kappa \sqrt{\varepsilon}, \quad \text{if } I(0) = J(0), \quad 0 \leq t \leq 1/\varepsilon$$

This result is optimal.

Additional theorems on the relation between the trajectories of the full and averaged systems have been proven by Neishtadt and Arnold in the references cited. Condition A of the Theorem precludes any trajectory of (3) from being captured on a *resonance manifold*. For system (3), the conditions for the existence of an $(m:n)$ resonance manifold are given by

$$\begin{aligned}
& m\Omega_1(I_1) - n\omega_2 = 0 \\
& \int_0^{2\pi} \int_0^{2\pi} \tilde{F}_p(J_1, J_2, \theta_1, \theta_2) \exp[-j(m\theta_1 - n\theta_2)] d\theta_1 d\theta_2 \neq 0, \\
& p = 1,2
\end{aligned} \tag{5}$$

where m, n are integers. A basic feature of two-frequency systems, such as (3), is the possibility *of resonance capture* on a resonance manifold: This is a dynamic phenomenon where an orbit gets ''trapped'' by an attracting region of the system in an $O(\sqrt{\varepsilon})$ neighborhood (boundary layer) of the resonance manifold. After the trajectory gets ''captured'' on the resonance manifold, averaging is not justified anymore since the time average is not close to the space average over the entire 2-Torus $(\theta_1, \theta_2) \in T^2$; hence condition A of the Theorem.

From (4) it directly follows that in the absence of resonance capture one expects the action variables to decay (approximately) exponentially in time. In Fig. 1 we depict the numerical time decays of the action variables for system (1) corresponding to parameters $\lambda = 0.5$, $\omega_2^2 = 0.9$, $C = 5.0$, $\varepsilon = 0.1$, and initial conditions $y_1(0) = y_2(0) = 0$, $\dot{y}_1(0) = 0$, $\dot{y}_2(0) = \sqrt{2h}$ with varying h (the energy of the system at $t = 0+$) varies; these responses correspond to the transient responses of the physical coordinates depicted in Fig. 1 in Part I of this work ([1]). For $h = 0.5$ no energy pumping occurs and the actions decay nearly exponentially to zero indicating the absence of resonance capture. At the higher energy values $h = 0.8$ and 1.125 energy pumping from oscillator 2 to oscillator 1 occurs, which is indicated by the fact that as time progresses $I_1(t)$ surpasses $I_2(t)$. Moreover, at certain time intervals there is a ''flattening'' of the plot of $I_2(t)$, accompanied with oscillatory behavior of $I_1(t)$; these variations of the action plots from exponential decay indicate the occurrence of resonance capture at these higher energy values, a phenomenon which can be directly associated to energy pumping. In addition, examination of the transient response $y_1(t)$ during energy pumping indicates the presence of fast oscillations with frequency approximately equal to ω_2 ([1]). This observation, coupled with the previous findings suggests that *the energy pumping phenomenon in* (1) *is associated with resonance capture in a neighborhood of the 1:1 resonance manifold*.

Motivated by the above discussion, we analyze in detail possible resonance capture in (1) associated with the (1:1) resonance manifold (it is a straightforward task to show that conditions (5) hold for $m = n = 1$),

$$\Omega_1(I_1) = \omega_2 \Rightarrow I_1 = (\omega_2/\Xi)^3 \equiv I_1^{(1-1)} \tag{6}$$

and restrict the analysis in an $O(\sqrt{\varepsilon})$ neighborhood of this manifold. To study the dynamics in the boundary layer close to this manifold we introduce the (slow) combination angle $\psi = \theta_1 - \theta_2$, and introduce the change of angles $(\theta_1, \theta_2) \rightarrow (\psi, \theta_2)$ and the following coordinate transformation in (3):

$$I_1 = I_1^{(1-1)} + \sqrt{\varepsilon}\xi, \quad I_2 = \eta. \tag{7}$$

Transforming the last of Eqs. (3) using the previous coordinate changes, we express the fast angle θ_2 in terms of the independent time variable, as follows:

$$\dot{\theta}_2 = \omega_2 + \varepsilon G_2(I_1^{(1-1)}, \eta, \psi, \theta_2) + O(\varepsilon^{3/2}) \Rightarrow t = (\theta_2/\omega_2) + O(\varepsilon). \tag{8}$$

Employing (8) we establish θ_2 as the independent variable in the remaining three Eqs. (3), and express them as

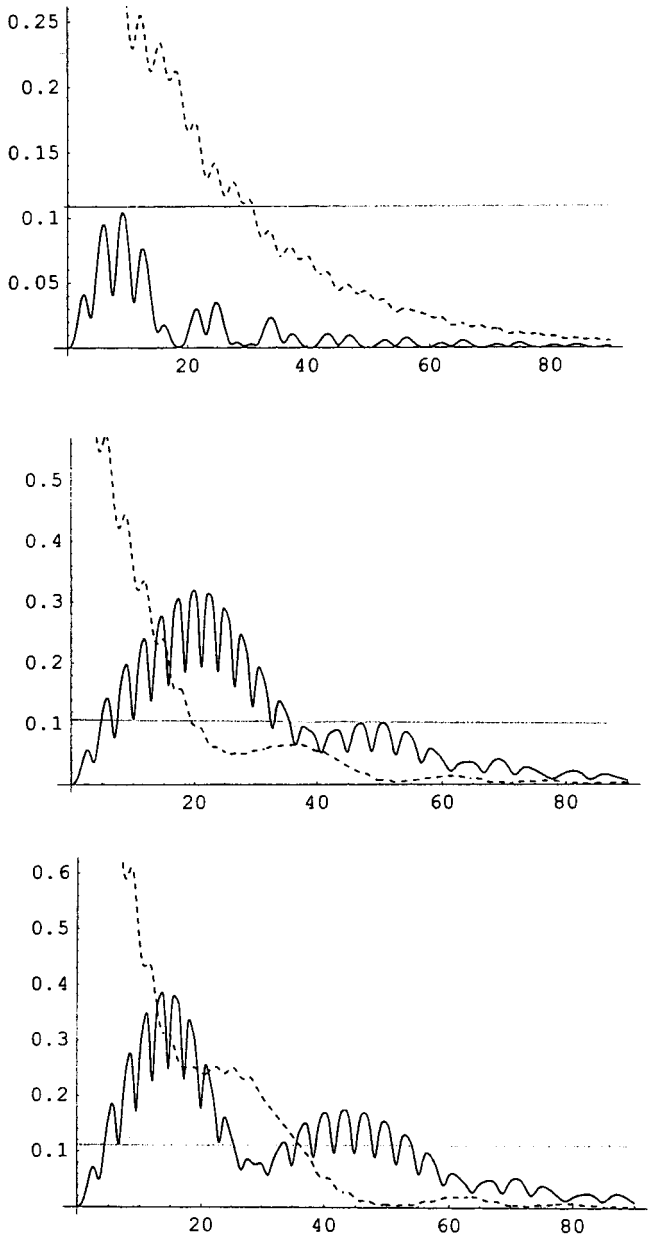


Fig. 1 Transient responses $I_1(t)$ and $I_2(t)$ of system (1) for, (*a*) $h=0.5$, (*b*) $h=0.8$, (*c*) $h=1.125$; —— oscillator 1, — — — oscillator 2

$$\xi' = \sqrt{\varepsilon}\omega_2^{-1}F_1(I_1^{(1-1)},\eta,\psi,\theta_2)+\varepsilon\omega_2^{-1}\frac{\partial F_1}{\partial I_1}(I_1^{(1-1)},\eta,\psi,\theta_2) + O(\varepsilon^{3/2})$$

$$\eta' = \varepsilon\omega_2^{-1}F_1(I_1^{(1-1)},\eta,\psi,\theta_2)+O(\varepsilon^{3/2}) \tag{9}$$

$$\psi' = \sqrt{\varepsilon}\Omega_1'(I_1^{(1-1)})\omega_2^{-1}\xi+\varepsilon\omega_2^{-1}\left[\Omega_1''(I_1^{(1-1)})\frac{\xi^2}{2} + G_1(I_1^{(1-1)},\eta,\psi,\theta_2)-G_2(I_1^{(1-1)},\eta,\psi,\theta_2)\right]+O(\varepsilon^{3/2})$$

where primes denote differentiation with respect to θ_2, and the following notation is adopted:

$$\widetilde{F}_p(I_1,I_2=\eta,\theta_1=\psi+\theta_2,\theta_2)\equiv F_p(I_1,\eta,\psi,\theta_2),$$

$$\widetilde{G}_p(I_1,I_2=\eta,\theta_1=\psi+\theta_2,\theta_2)\equiv G_p(I_1,\eta,\psi,\theta_2),\quad p=1,2.$$

We note that, in contrast to (3), (9) is a *local* model since it is valid only in an $O(\sqrt{\varepsilon})$ neighborhood of the (1:1) resonance manifold. Since the only remaining fast variable (θ_2) is used as independent variable, all dependent variables in (9) are slow varying, and, as a result, we can apply asymptotic techniques from the theory of nonlinear dynamics (such as, averaging or multiple-scales) to study the flow of the system close to the resonance manifold.

To this end, we replace the independent variable in (9) by two ''slow'' and ''fast'' variables, $\zeta=\sqrt{\varepsilon}\theta_2$, $\delta=\theta_2$, respectively. In addition, we express the dependent variables in series,

$$\begin{aligned}\xi&=\xi_0(\delta,\zeta)+\sqrt{\varepsilon}\xi_1(\delta,\zeta)+\varepsilon\xi_2(\delta,\zeta)+O(\varepsilon^{3/2})\\ \eta&=\eta_0(\delta,\zeta)+\sqrt{\varepsilon}\eta_1(\delta,\zeta)+\varepsilon\eta_2(\delta,\zeta)+O(\varepsilon^{3/2})\\ \psi&=\psi_0(\delta,\zeta)+\sqrt{\varepsilon}\psi_1(\delta,\zeta)+\varepsilon\psi_2(\delta,\zeta)+O(\varepsilon^{3/2}),\end{aligned} \tag{10}$$

and substitute into (9). Balancing the coefficients of the same order of ε we obtain a series of subproblems governing the approximations of different orders in (10).

***O*(ε^0) Approximations.** The zeroth-order approximations can be trivially solved:

$$\begin{aligned}\xi_{0\delta}=0&\Rightarrow\xi_0(\delta,\zeta)=A_0(\zeta)\\ \eta_{0\delta}=0&\Rightarrow\eta_0(\delta,\zeta)=B_0(\zeta)\\ \psi_{0\delta}=0&\Rightarrow\psi_0(\delta,\zeta)=C_0(\zeta)\end{aligned} \tag{11}$$

where the short-hand notation for partial differentiation, $\partial(\bullet)/\partial\delta\equiv(\bullet)_\delta$, is adopted, and the ζ-dependent functions are determined at the next order of approximation.

***O*($\varepsilon^{1/2}$) Approximations.** The subproblems governing the first-order approximations are

$$\begin{aligned}\xi_{1\delta}&=-\xi_{0\zeta}+\omega_2^{-1}F_1(I_1^{(1-1)},\eta_0,\psi_0,\delta)\\ \eta_{1\delta}&=-\eta_{0\zeta}\\ \psi_{1\delta}&=-\psi_{0\zeta}+\omega_2^{-1}\Omega_1'(I_1^{(1-1)})\xi_0.\end{aligned} \tag{12}$$

Substituting (11) into (12), and eliminating secular terms (i.e., right-hand side terms depending only on ζ), we obtain the following solvability relations that govern the unknown functions in (11):

$$\begin{aligned}&A_0'(\zeta)+(2\pi\omega_2)^{-1}\int_0^{2\pi}F_1(I_1^{(1-1)},B_0(\zeta),C_0(\zeta),\delta)d\delta=0\\ &B_0'(\zeta)=0\Rightarrow B_0(\zeta)=B_0\\ &C_0'(\zeta)-\omega_2^{-1}\Omega_1'(I_1^{(1-1)})A_0(\zeta)=0.\end{aligned} \tag{13}$$

Performing explicitly the integration in the first of the above relations and combining the three equations into a single second-order one we obtain a pendulum equation with constant forcing governing the combination angle,

$$\begin{aligned}&C_0''(\zeta)-\omega_2^{-1}\Omega_1'(I_1^{(1-1)})\\ &\quad\times(2\pi\omega_2)^{-1}\int_0^{2\pi}F_1(I_1^{(1-1)},B_0(\zeta),C_0(\zeta),\delta)d\delta\\ &\quad=0\Rightarrow C_0''(\zeta)+\mu\cos C_0(\tau)=-\nu,\end{aligned} \tag{14}$$

where $\mu=0.8987\Xi\sqrt{B_0}/\omega_2^{3/2}K(1/2)\Lambda$ and $\nu=(\lambda\omega_2)/3$. Depending on the relative values of μ and ν, the phase portrait of (14) possess (if $\mu>\nu$) or not (if $\mu\leqslant\nu$) a closed homoclinic loop of containing closed periodic orbits surrounding a stable equilibrium point (cf. Fig. 2). This loop when perturbed by higher order terms becomes the attracting region responsible for sustained resonance

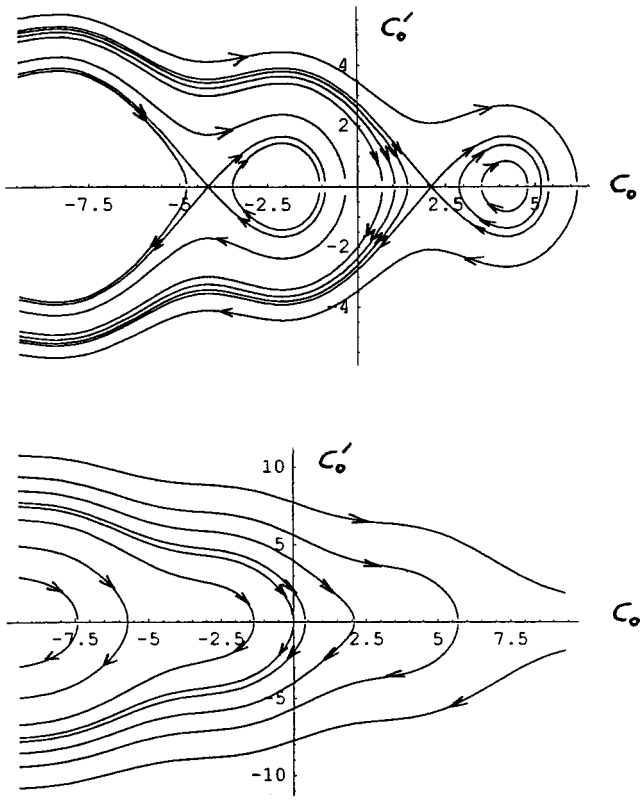


Fig. 2 Phase portraits of system (14) for, (*a*) $\mu>\nu$, and (*b*) $\mu\leqslant\nu$

capture in system (1): under certain initial conditions, trajectories of the system in an $O(\sqrt{\varepsilon})$ neighborhood of the 1:1 resonance manifold get attracted to the region of the loop where they perform multiple oscillations around the attractor. Under different initial conditions trajectories lie outside the homoclinic loop and get repelled away from the attracting region; in this case no resonance capture occurs. Note that sustained resonance capture is only possible if $\mu>\nu$, which leads to the following lower bound for B_0:

$$B_0>\left(\frac{\lambda\omega_2^{5/2}K(1/2)\Lambda}{2.696\Xi}\right)^2 \quad \text{(condition for } \textit{resonance capture}\text{)}. \tag{15}$$

This relation indicates that for resonance capture to occur the action variable (i.e., the energy) of the directly excited linear oscillator 1 must be above a certain threshold. This conclusion is in accordance with the conjecture made in Part I of this work ([1]).

After computing the ζ-dependent functions (11) by eliminating secular terms from (12), the $O(\sqrt{\varepsilon})$ approximations are computed as

$$\begin{aligned}\xi_1(\delta,\zeta)&=\omega_2^{-1}\hat{F}_1^0(I_1^{(1-1)},B_0,C_0(\zeta),\delta)+A_1(\zeta)\\ \eta_1(\delta,\zeta)&=B_1(\zeta)\\ \psi_1(\delta,\zeta)&=C_1(\zeta)\end{aligned} \tag{16}$$

where $\hat{F}_1^0(I_1^{(1-1)},B_0,C_0(\zeta),\delta)=\int^{\delta}\hat{F}_1(I_1^{(1-1)},B_0,C_0(\zeta),u)du$ is a 2π-periodic function in δ, and $\hat{F}_1$ denotes the zero-mean component of the function F_1 (i.e., the function minus the constant term appearing in the first of Eqs. (13)). The unknown ζ-dependent functions of the solutions are computed by eliminating secular terms at the next order of approximation.

***O*(ε) Approximations.** The subproblems governing the second-order approximations are rather involved and are omitted. The equations for eliminating the secular terms from these subproblems are given below:

$$\begin{aligned}A_1'(\zeta)&=T_1(\zeta)C_1(\zeta)+q_1(\zeta)\\ C_1'(\zeta)&=T_2A_1(\zeta)+q_2(\zeta)\\ B_1'(\zeta)&=(2\pi\omega_2)^{-1}\int_0^{2\pi}F_2(I_1^{(1-1)},B_0,C_0(\zeta),\delta)d\delta\end{aligned} \tag{17}$$

where

$$T_1(\zeta)=(2\pi\omega_2)^{-1}\int_0^{2\pi}\frac{\partial F_1}{\partial\psi}(I_1^{(1-1)},B_0,C_0(\zeta),\delta)d\delta,$$

$$T_2=\omega_2^{-1}\Omega_1'(I_1^{(1-1)})$$

$$\begin{aligned}q_1(\zeta)=(2\pi\omega_2)^{-1}\int_0^{2\pi}\Bigg[&-\frac{\partial\hat{F}_1^0}{\partial C_0(\zeta)}C_0'(\zeta)\\ &+\frac{\partial F_1}{\partial\eta}(I_1^{(1-1)},B_0,C_0(\zeta),\delta)B_1(\zeta)\\ &+\frac{\partial F_1}{\partial I_1}(I_1^{(1-1)},B_0,C_0(\zeta),\delta)A_0(\zeta)\Bigg]d\delta\end{aligned}$$

$$\begin{aligned}q_2(\zeta)=(2\pi\omega_2^2)^{-1}\Omega_1'(I_1^{(1-1)})&\int_0^{2\pi}\hat{F}_1^0(I_1^{(1-1)},B_0,C_0(\zeta),\delta)d\delta\\ &+(2\pi\omega_2)^{-1}\int_0^{2\pi}[G_1-G_2]_{(I_1^{(1-1)},B_0,C_0(\zeta),\delta)}d\delta\\ &+(2\omega_2)^{-1}\Omega_1''(I_1^{(1-1)})A_0^2(\zeta).\end{aligned}$$

The third of Eqs. (17) uncouples from the first two, which form a set of nonhomogeneous *linear* equations with a parameter-dependent coefficient. These equations govern the perturbations of the phase portraits of Fig. 2 and produce the attracting region for resonance capture.

Although algebraically involved, we now show that the solution of the linear set (17) can be written in explicit analytical form. To this end, we combine the first two equations into a single second-order equation as follows:

$$C_1''(\zeta)-T_2T_1(\zeta)C_1(\zeta)=T_2q_1(\zeta)+q_2'(\zeta). \tag{18}$$

We note that the parameter-dependent coefficient of (18) can be expressed as

$$\begin{aligned}T_2T_1(\zeta)=\frac{\partial}{\partial\psi}\Bigg\{&\omega_2^{-1}\Omega_1'(I_1^{(1-1)})(2\pi\omega_2)^{-1}\\ &\times\int_0^{2\pi}F_1(I_1^{(1-1)},B_0,\psi,\delta)d\delta\Bigg\}_{\psi=C_0(\zeta)}\end{aligned} \tag{19}$$

i.e., as the partial derivative in ψ of the second part of Eq. (14) governing $C_0(\zeta)$. It follows that one homogeneous solution of (18) is given by $C_1^{(1h)}=\alpha C_0'(\zeta)$. A second linearly independent homogeneous solution can be obtained by considering the equation for the Wronskian of (18), leading to, $C_1^{(2h)}=\beta C_0'(\zeta)\int^{\zeta}[C_0'(u)]^{-2}du$. These two linearly independent homogeneous solutions are used to compute a particular integral by the method of variation of parameters, which completes the solution. The final expression of the solution of (18) is given by

$$C_1(\zeta)=\left[\alpha-\int_0^{\zeta}C_1^{(2h)}(v)[T_2q_1(v)+q_2'(v)]dv\right]C_1^{(1h)}(\zeta)$$
$$+\left[\beta-\int_0^{\zeta}C_1^{(1h)}(v)[T_2q_1(v)+q_2'(v)]dv\right]C_1^{(2h)}(\zeta). \tag{20}$$

The coefficients α and β in (20) are determined by satisfiying limiting conditions of the solution as ζ increases or decreases. Examples of such calculations are given in [18].

In the next section we construct analytical approximations for the transient response of (1) in the initial phase of energy pumping (2). In the spirit of the previous section the analysis is based on the assumption of 1:1 internal resonance, and the results compare favorable to direct numerical simulations.

3 Analytical Approximations

We start the analysis by rewriting system (1) in the following form:

$$\ddot{y}_1+\varepsilon\lambda\dot{y}_1+\varepsilon y_1+Cy_1^3-\varepsilon y_2=0$$
$$\ddot{y}_2+\varepsilon\lambda\dot{y}_2+\omega^2y_2-\varepsilon y_1=0 \tag{21}$$

where $\omega^2=\omega_2^2+\varepsilon$. Note that in (21) the damping coefficients are assumed to be of $O(1)$, in contrast to (1). At this point we complexify the problem by introducing the new complex variables,

$$\psi_1=\dot{y}_1+j\omega y_1, \quad \psi_2=\dot{y}_2+j\omega y_2, \tag{22}$$

$j=(-1)^{1/2}$, and express (21) as

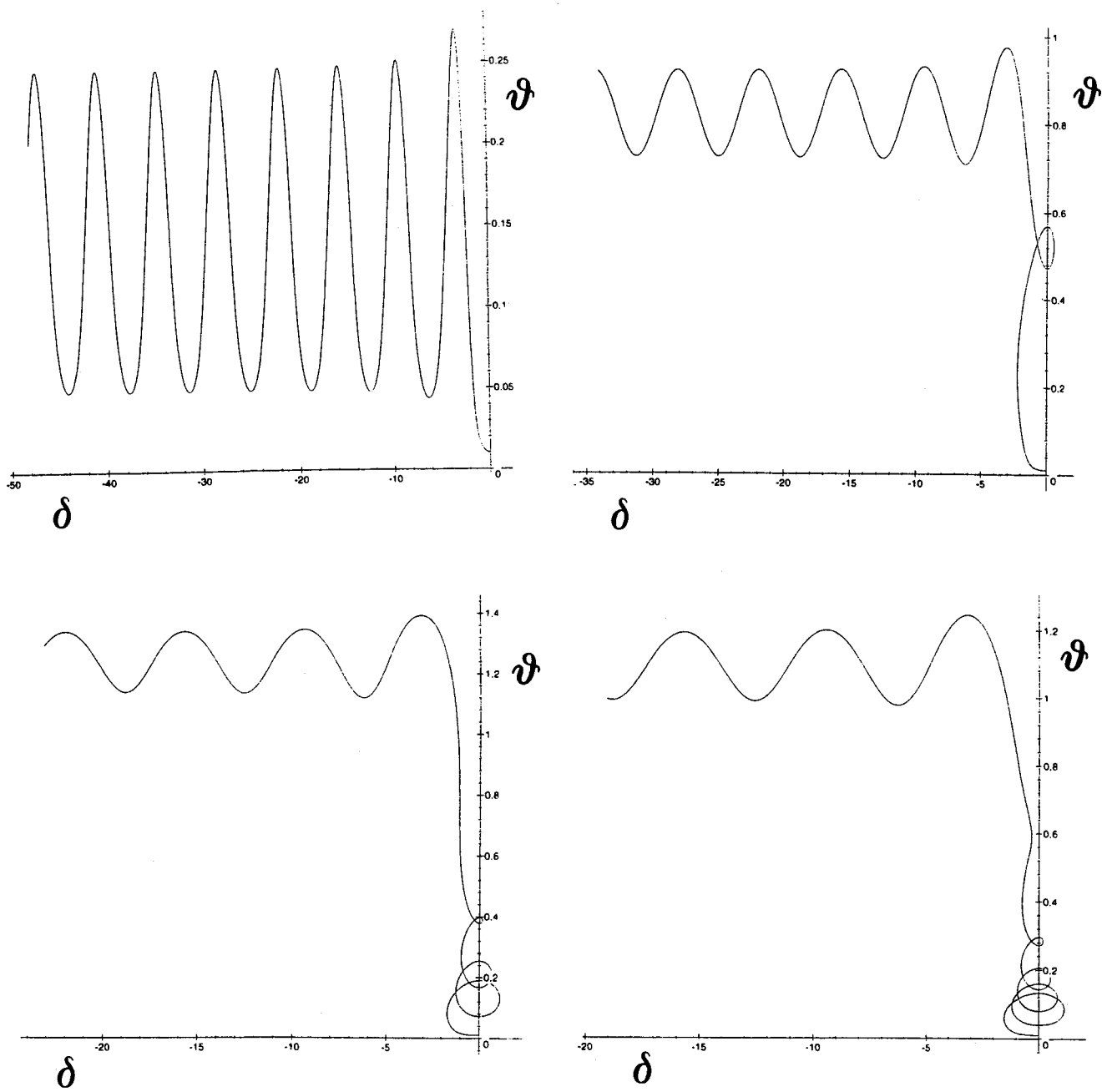


Fig. 3 Numerical solutions of system (30): (*a*) no resonance capture ($M=2.8$), (*b*, *c*, *d*) resonance capture ($M=4.0, 10.0, 15.0$)

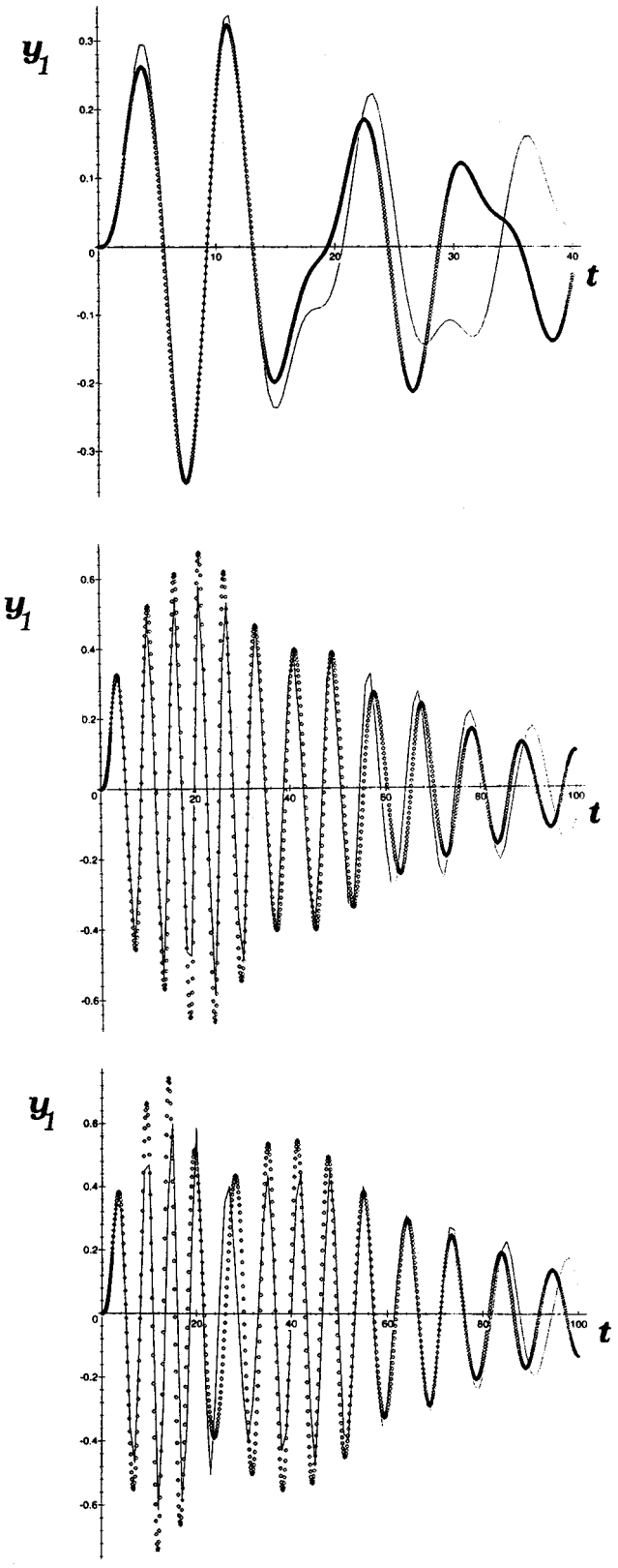


Fig. 4 Transient response $y_1(t)$ of system (1), (*a*) when no energy pumping occurs ($h=0.5$), and (*b*,*c*) when energy pumping takes place ($h=0.8, 1.125$). ◇ ◇ ◇ ◇ ◇ ◇ ◇ ◇ ◇ Analytical approximations based on (30), —— Numerical simulations.

$$\dot{\psi}_1-\frac{j\omega}{2}(\psi_1+\psi_1^*)+\frac{\varepsilon\lambda}{2}(\psi_1+\psi_1^*)-\frac{j\varepsilon}{2\omega}(\psi_1-\psi_1^*)+\frac{jC}{8\omega^3}(\psi_1+\psi_1^*)^3+\frac{j\varepsilon}{2\omega}(\psi_2-\psi_2^*)=0$$
$$\dot{\psi}_2-j\omega\psi_2+\frac{\varepsilon\lambda}{2}(\psi_2+\psi_2^*)+\frac{j\varepsilon}{2\omega}(\psi_1-\psi_1^*)=0. \tag{23}$$

Relations (23) are exact. We now seek an approximate solution of (23) is sought, based on the assumption of fast oscillations at frequency ω:

$$\psi_1=\varphi_1 e^{j\omega t},\quad \psi_2=\varphi_2 e^{j\omega t}. \tag{24}$$

Relations (24) signify 1:1 internal resonance condition in the fast dynamics of the system. Substituting (24) into (23), and averaging over the fast periodic variables $e^{j\omega t}$, we obtain the following averaged system:

$$\dot{\varphi}_1+\frac{j}{2}\left(\omega-\frac{\varepsilon}{\omega}\right)\varphi_1+\frac{\varepsilon\lambda}{2}\varphi_1-\frac{3jC}{8\omega^3}|\varphi_1|^2\varphi_1+\frac{j\varepsilon}{2\omega}\varphi_2=0$$
$$\dot{\varphi}_2+\frac{\varepsilon\lambda}{2}\varphi_2+\frac{j\varepsilon}{2\omega}\varphi_1=0. \tag{25}$$

For $\lambda=0$ the system is completely integrable and has been further analyzed in [1]. To account for the amplitude decays due to damping we introduce the new variables σ_1 and σ_2 defined by

$$\varphi_1=\sigma_1\exp(-\varepsilon\lambda t/2),\quad \varphi_2=\sigma_2\exp(-\varepsilon\lambda t/2) \tag{26}$$

and express (25) as

$$\dot{\sigma}_1+\frac{j}{2}\left(\omega-\frac{\varepsilon}{\omega}\right)\sigma_1-\frac{3jCe^{-\varepsilon\lambda t}}{8\omega^3}|\sigma_1|^2\sigma_1+\frac{j\varepsilon}{2\omega}\sigma_2=0$$
$$\dot{\sigma}_2+\frac{j\varepsilon}{2\omega}\sigma_1=0. \tag{27}$$

Manipulating the above set of equations it can be shown that it possesses the first integral,

$$|\sigma_1|^2+|\sigma_2|^2=M^2, \tag{28}$$

which enables one to express the amplitudes in the following way:

$$\sigma_1=M\sin\theta e^{j\delta_1},\quad \sigma_2=M\cos\theta e^{j\delta_2}. \tag{29}$$

Substituting (29) into (27), and performing algebraic manipulations we reduce the problem to a final set of parameter-dependent nonlinear equations on the 2-Torus:

$$\dot{\delta}+\frac{\omega}{2}-\frac{3CMe^{-\varepsilon\lambda t}}{8\omega^3}\sin^2\theta+\frac{\varepsilon}{\omega}\cot 2\theta\cos\delta=0$$
$$\dot{\theta}+\frac{\varepsilon}{2\omega}\sin\delta=0 \tag{30}$$

where $\delta=\delta_1-\delta_2$. We mention that although the set of equations above appears to be similar to the one derived in [1] for $\lambda=0$ (actually, the two sets become identical by setting $Me^{-\varepsilon\lambda t}=N$ in the notation of that work), the damped dynamics is dominated by the ''drifting'' of the ''instantaneous equilibrium points'' due to the exponentially decaying term in (30). We note that δ denotes the relative phase, while θ determines the instantaneous amplitudes of the motions of oscillators 1 and 2.

The numerical integrations of system (30) for varying values of the initial first integral M reveal clearly the energy pumping and resonance capture phenomenon occurring in system (1). These results are presented in the (δ,θ) phase plots of Fig. 3 for parameters $\omega=1$, $C=2.0$, $\varepsilon=0.1$, $\lambda=1.0$ and initial conditions $\delta(0)=0$ and $\theta(0)=0.01$. These numerical results were obtained by matching displacements and velocities of the following two solution branches: (a) the initial branch is obtained by Taylor expansions

of the original equations of motion (1) close to $t=0$ and takes fully into account the initial conditions of the system; (b) the later branch consists of the numerical solution of the analytic approximation (30). For $M=2.8$ the initial energy imparted in oscillator 2 remains confined to that oscillator since the damped oscillation corresponds to small values of θ (cf. relations (29)). At higher values of M we note energy pumping to oscillator 1, corresponding to trajectories which start with small values of θ and after some transients settle to damped oscillations with θ close to $\pi/2$. Of particular interest is the fact that the analytical results capture nicely not only the resonance capture of trajectories, but also transient capture, whereby an initially transient trajectory is temporally captured and then released by the attracting region of the system. This last dynamical phenomenon can be realized by suitably choosing the initial conditions in (30), and is depicted in Figs. 3(*b–d*).

To compare the analytical predictions with the numerical results reported in [1] in Fig. 4 we depict analytical approximations for the transient responses of (1) based on Eqs. (30) for parameters $C=5.0$, $\omega_2^2=0.9$, $\varepsilon=0.1$, $\lambda=0.5$; the initial conditions are given zero values except for the initial velocity of the (linear) oscillator 2 which assumes the value $\dot{y}_2(0)=\sqrt{2h}$. For comparison purposes in the same pictures we depict the results of direct numerical simulations of system (1). In Fig. 4(*a*) where no energy pumping occurs there is poor agreement between the two results. In Figs. 4(*b,c*) where energy pumping takes place there is satisfactory agreement between the predicted and actual transient responses, although some overshooting or undershooting can be noted over certain time intervals. These errors can be attributed to the averaging approximations introduced in deriving Eqs. (25), and to the strong nonlinearities of the system considered. These results indicate that when 1:1 resonance capture occurs the analytical method presented in this section can satisfactorily predict the nonlinear transient responses of the system. This should be expected, since by relations (24) the analytical constructions were based on an 1:1 internal resonance in the ''fast'' dynamics.

4 Discussion

We analyzed the energy pumping phenomenon in the impulsively excited, damped system (1) by means of two analytical techniques. The first technique, is based on the perturbation analysis of the dynamics in an $O(\sqrt{\varepsilon})$ neighborhood of the 1:1 resonance manifold of the system. Analytical results are derived for the attractive region responsible for the resonance capture. The second method is based on the assumption of 1:1 internal resonance in the fast dynamics, and utilizes complexification and averaging of the equations of motion. This leads to satisfactory analytical approximations of the nonlinear transient responses of the system in the regime of energy pumping.

The energy pumping phenomenon considered in Parts I and II of this work can lead to the design of structures with local passive or active sinks where externally imparted energy is initially directed and then eliminated. Of particular interest to the authors is the potential enhancement of the energy pumping phenomenon through either nonlinear coupling stiffness elements, or active control. This would pave the way for practical implementation of energy pumping in vibration and shock isolation designs of engineering systems.

Acknowledgments

This work was supported in part by NSF Young Investigator Award CMS-94-57750 (Dr. Allison Flatau is the Grant Monitor), and by an International Supplement to this Award (Ms. Cassandra Dudca is the Grant Monitor).

References

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

[2] Arnold, V. I., ed., 1988, *Dynamical Systems III* (Encyclopaedia of Mathematical Sciences), Vol. 3, Springer-Verlag, Berlin.

[3] Lochak, P., and Meunier, C., 1988, *Multiphase Averaging for Classical Systems* (Series on Applied Mathematical Sciences), Vol. 72, Springer-Verlag, Berlin.

[4] Morozov, A. D., 1998, *Quasi-conservative Systems, Cycles, Resonances and Chaos* (Series on Nonlinear Science, Series A), Vol. 30, World Scientific, Singapore.

[5] Neishtadt, A. I., 1975, ''Passage Through a Resonance in the Two-Frequency Problem,'' Dokl. Akad. Nauk, **221**, pp. 301–304.

[6] Neishtadt, A. I., 1976, ''Averaging in Multifrequency Systems II,'' Dokl. Akad. Nauk SSSR, **226**, pp. 1295–1298.

[7] Neishtadt, A. I., 1975, ''Passage Through a Separatrix in a Resonance Problem With a Slowly-Varying Parameter,'' Prikl. Mat. Meck. (PMM), **39**, No. 4, pp. 621–632.

[8] Haberman, P., 1983, ''Energy Bounds for the Slow Capture by a Center in Sustained Resonance,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **43**, No. 2, pp. 244–256.

[9] Morozov, A. D., and Shil'nikov, L. P., 1984, ''On Nonconservative Periodic Systems Close to Two-Dimensional Hamiltonian,'' Prikl. Mat. Meck. (PMM), **47**, No. 3, pp. 327–334.

[10] Lewin, L., and Kevorkian, J., 1978, ''On the Problem of Sustained Resonance,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **35**, No. 4, pp. 738–754.

[11] Kath, W. L., 1983, ''Necessary Conditions for Sustained Reentry Roll Resonance,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **43**, No. 2, pp. 314–324.

[12] Kath, W. L., 1983, ''Conditions for Sustained Resonance II,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **43**, No. 3, pp. 579–583.

[13] Quinn, D., and Rand, R., 1995, ''The Dynamics of Resonance Capture,'' Nonlinear Dyn., **8**, pp. 1–20.

[14] Rand, R., and Quinn, D., 1995, ''Resonant Capture in a System of Two Coupled Homoclinic Oscillators,'' J. Vib. Control, **1**, pp. 41–56.

[15] Bosley, D. L., and Kevorkian, J., 1995, ''On the Asymptotic Solution of Non-Hamiltonian Systems Exhibiting Sustained Resonance,'' Stud. Appl. Math., **94**, pp. 83–130.

[16] Bosley, D. L., 1996, ''An Improved Matching Procedure for Transient Resonance Layers in Weakly Nonlinear Oscillatory Systems,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math., **56**, No. 2, pp. 420–445.

[17] Percival, I., and Richards, D., 1982, *Introduction to Dynamics*, Cambridge University Press, Cambridge, UK.

[18] Vakakis, A. F., 1994, ''Exponentially Small Splittings of Manifolds in a Rapidly Forced Duffing System,'' J. Sound Vib., **170**, No. 1, pp. 119–129.

Parametric Instability of Axially Moving Media Subjected to Multifrequency Tension and Speed Fluctuations

R. G. Parker[1]
Mem. ASME
e-mail: parker.242@osu.edu

Y. Lin
Graduate Student

Department of Mechanical Engineering,
The Ohio State University,
206 W. 18th Avenue,
Columbus, OH 43210-1107

This work investigates the stability of axially moving media subjected to parametric excitation resulting from tension and translation speed oscillations. Each of these excitation sources has spectral content with multiple frequencies and arbitrary phases. Stability boundaries for primary parametric instabilities, secondary instabilities, and combination instabilities are determined analytically through second-order perturbation. The classical result that primary instability occurs when one of the excitation frequencies is close to twice a natural frequency changes as a result of multiple excitation frequencies. Unusual interactions occur for the practically important case of simultaneous primary and secondary instabilities. While sum type combination instabilities occur, no difference type instabilities are detected. The nonlinear limit cycle amplitude that occurs under primary instability is derived using the method of multiple scales. [DOI: 10.1115/1.1343914]

Introduction

The transverse vibration of axially moving materials subjected to parametric excitations has received considerable attention from many researchers. Most studies have addressed the stability under parametric excitation with a single frequency component. Practical systems, however, are subjected to multifrequency excitations that may significantly impact the dynamic behavior. In vehicle serpentine belt drives, for example, the engine drives a crankshaft pulley that powers a single belt, which in turn supplies power to multiple automotive accessories. Engine firing pulses cause belt translation speed fluctuations. Additionally, these engine firing pulses, in combination with dynamic accessory torques, excite pulley rotational vibrations that lead to tension oscillations in the individual belt spans. The speed and tension fluctuations both parametrically excite the belt spans. The speed oscillations have spectral content related to harmonics of the engine speed, and the tension oscillations have multifrequency spectral content associated with the engine speed and dynamic accessory load frequencies.

This study investigates the stability of parametrically excited, moving media subjected to dynamic tension and speed fluctuations with arbitrary spectral content. A discretization/perturbation method yields the excitation frequency-amplitude parameter plane boundaries separating stable and unstable regions for the single mode primary and secondary resonances and the combination resonances for any two modes. These boundaries are determined analytically in closed form through second-order perturbation. Nonlinear limit cycles that occur in the parametric resonance regions are determined analytically and numerically.

This work builds on that of Mockensturm et al. [1], who determined closed-form analytical expressions for all primary and the first sum type combination resonance regions of a moving string with tension fluctuation. Similarly, Pakdemirli and Ulsoy [2] determined stability boundaries for the moving string with speed fluctuation. Both of these analyses are limited to monofrequency parametric excitation and first-order approximation. They do not address secondary resonances. The present work shows unique and practically important behaviors associated with multifrequency excitation, secondary resonances, and second-order approximation.

The two works noted above give a good review of prior studies on parametrically excited moving media. Particularly relevant studies include the work of Mahalingam [3], Mote [4,5], Naguleswaran and Williams [6], and Asokanthan and Ariaratnam [7]. Recent studies include Oz et al. [8] and Chakraborty and Mallik [9]. Ulsoy et al. [10] motivated the studies for automotive belt drives by showing a primary source of transverse belt vibration to be parametric instability caused by tension fluctuation.

Problem Formulation

The system is a beam/string of length L moving with time-dependent transport velocity $c(T)$. The equation of motion for transverse vibration is

$$\rho A(V_{TT}+c_T V_x+2cV_{TX}+c^2V_{XX})-(P_d+P_s)V_{XX}+EIV_{XXXX}=0 \tag{1}$$

where ρA is the mass per unit length, EI is the bending stiffness, V is the transverse displacement, T is the time, X is the spatial coordinate, P_s is the mean belt tension, and $P_d(T)$ is the dynamic tension. The dynamic tension results from longitudinal belt motion and midplane stretching from transverse deflection. Under the assumption of quasi-static stretching ([11]), the dynamic tension is

$$P_d=\frac{EA}{L}\left[U(L,T)-U(0,T)+\frac{1}{2}\int_0^L V_x^2\,dX\right] \tag{2}$$

where EA is the longitudinal stiffness modulus and U is the longitudinal displacement. Use of the dimensionless parameters

$$x,v,u=\frac{X,V,U}{L},\quad t=\sqrt{\frac{P_s}{\rho AL^2}}\,T,\quad \gamma=c\Big/\sqrt{\frac{P_s}{\rho A}},$$

$$\zeta=\frac{EA}{P_s},\quad \alpha=\frac{EI}{P_sL^2} \tag{3}$$

[1]To whom correspondence should be addressed

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Aug. 26, 1999; final revision, June 27, 2000. Associate Editor: A. K. Mal. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

gives

$$v_{tt}+2\gamma v_{tx}+\gamma_t v_x-(1-\gamma^2)v_{xx}+\alpha v_{xxxx} -\zeta\left[u(1,t)-u(0,t)+\frac{1}{2}\int_0^1 v_x{}^2 dx\right]v_{xx}=0. \quad (4)$$

The relative longitudinal motion of the end points, which results from rotational pulley oscillations, consists of multifrequency excitation of the form

$$\zeta[u(1,t)-u(0,t)]=\zeta\sum_{i=1}^{k} u_i\cos(\Omega_i t+\theta_i)=\sum_{i=1}^{k}\varepsilon_i\cos(\Omega_i t+\theta_i). \quad (5)$$

$\varepsilon_i=(EAu_i)/P_s<1$ represents the ratio of the dynamic tension fluctuation caused by the ith spectral component of the relative endpoint motion to the mean span tension. The relative longitudinal motion of the endpoints $u(1,t)-u(0,t)$ is specified. In serpentine belt drives, it is calculated from dynamic analysis of the discrete pulley rotations induced by crankshaft excitations and dynamic accessory torques. Engine firing pulses cause a speed ''ripple'' on the mean crankshaft rotation speed. The associated belt speed fluctuations are

$$\gamma=\gamma_0+\sum_{i=1}^{k'}\varepsilon_i'\sin(\Omega_i' t+\theta_i'). \quad (6)$$

Typically, the dominant speed and tension fluctuation frequency equals $N/2$ times the engine speed, where N is the number of cylinders, though higher harmonics of this frequency and accessory torque frequencies are also present. The dimensionless frequencies are related to the dimensional ones (Ω_i^*) by $\Omega_i=\sqrt{\rho AL^2/P_s}\Omega_i^*$. From (4), the linearized equation of motion with tension and speed fluctuations is

$$v_{tt}+2\gamma_0 v_{tx}-(1-\gamma_0^2)v_{xx}+\alpha v_{xxxx}-\sum_{i=1}^{k}\varepsilon_i\cos(\Omega_i t+\theta_i)v_{xx} +\sum_{i=1}^{k'}\varepsilon_i'\{2v_{tx}\sin(\Omega_i' t+\theta_i')+2\gamma_0 v_{xx}\sin(\Omega_i' t+\theta_i') +\Omega_i' v_x\cos(\Omega_i' t+\theta_i')\}+\left(\sum_{i=1}^{k'}\varepsilon_i'\sin(\Omega_i' t+\theta_i')\right)^2 v_{xx}=0. \quad (7)$$

For subsequent discretization, it is convenient to rewrite (7) in state space form as

$$AW_t+BW+\sum_{i=1}^{k'}\varepsilon_i'\{\sin(\Omega_i' t+\theta_i')C+\Omega_i'\cos(\Omega_i' t+\theta_i')D\}W -\sum_{i=1}^{k}\varepsilon_i\cos(\Omega_i t+\theta_i)EW+\left(\sum_{i=1}^{k'}\varepsilon_i'\sin(\Omega_i' t+\theta_i')\right)^2 EW=0 \quad (8)$$

$$A=\begin{bmatrix}1 & 0\\ 0 & -(1-\gamma_0^2)\dfrac{\partial^2}{\partial x^2}+\alpha\dfrac{\partial^4}{\partial x^4}\end{bmatrix},$$

$$B=\begin{bmatrix}2\gamma_0\dfrac{\partial}{\partial x} & -(1-\gamma_0^2)\dfrac{\partial^2}{\partial x^2}+\alpha\dfrac{\partial^4}{\partial x^4}\\ (1-\gamma_0^2)\dfrac{\partial^2}{\partial x^2}-\alpha\dfrac{\partial^4}{\partial x^4} & 0\end{bmatrix},$$

$$C=\begin{bmatrix}2\dfrac{\partial}{\partial x} & 2\gamma_0\dfrac{\partial^2}{\partial x^2}\\ 0 & 0\end{bmatrix},\quad D=\begin{bmatrix}0 & \dfrac{\partial}{\partial x}\\ 0 & 0\end{bmatrix},$$

$$E=\begin{bmatrix}0 & \dfrac{\partial^2}{\partial x^2}\\ 0 & 0\end{bmatrix},\quad W=\begin{bmatrix}v_t\\ v\end{bmatrix}. \quad (9)$$

The inner product in the state space is $\langle W,V\rangle=\int_0^1 W^T\overline{V}dx$, where the overbar denotes complex conjugate and superscript T denotes transpose.

The Galerkin basis consists of the state-space eigenfunctions for the nonparametrically excited system (8) ([12])

$$\Phi_n=\begin{bmatrix}j\omega_n\psi_n\\ \psi_n\end{bmatrix}=\begin{bmatrix}\lambda_n\psi_n\\ \psi_n\end{bmatrix} \quad (10)$$

where ψ_n are the complex eigenfunctions of (7) ($\varepsilon_i=\varepsilon_i'=0$) and ω_n are the natural frequencies. The Φ_n possess the orthonormality properties $\langle A\Phi_n,\Phi_m\rangle=\delta_{mn}$, $\langle B\Phi_n,\Phi_m\rangle=-\lambda_n\delta_{mn}=-j\omega_n\delta_{mn}$. For the moving string model ($\alpha=0$) the eigensolutions are

$$\psi_n=\frac{1}{n\pi\sqrt{1-\gamma_0^2}}e^{jn\pi\gamma_0 x}\sin(n\pi x),\quad \lambda_n=j\omega_n=jn\pi(1-\gamma_0^2) \quad (11)$$

for fixed pulleys at the string supports. Eigensolutions for a traveling beam can not be expressed in closed form and require numerical solution ([4]).

Multifrequency Parametric Instabilities

To investigate primary parametric instabilities, we use a single-term Galerkin discretization for the nth mode obtained by use of one traveling system basis function

$$W=\xi_n(t)\Phi_n(x)+\overline{\xi_n}(t)\overline{\Phi}_n(x)=2\,\mathrm{Re}[\xi_n(t)\Phi_n(x)]. \quad (12)$$

Mockensturm et al. [1] demonstrated the excellent convergence achieved with this single term expansion for a string model. Substituting (12) into (8) and taking the inner product with Φ_n yields the complex, time-varying equation (the notation $E_{nm}=\langle E\Phi_n,\Phi_m\rangle$, $E_{\overline{nm}}=\langle E\overline{\Phi}_n,\Phi_m\rangle$, $\overline{E}_{nm}=\overline{\langle E\Phi_n,\Phi_m\rangle}$, etc., and similar relations for the C and D operators are used throughout)

$$\dot{\xi}_n-j\omega_n\xi_n-\left[\varepsilon\sum_{i=1}^{k}f_i\cos(\Omega_i t+\theta_i)\right](\xi_n E_{nn}+\overline{\xi_n}E_{\overline{nn}}) +\left[\varepsilon\sum_{i=1}^{k'}f_i'\sin(\Omega_i' t+\theta_i')\right](\xi_n C_{nn}+\overline{\xi_n}C_{\overline{nn}}) +\left[\varepsilon\sum_{i=1}^{k'}f_i'\Omega_i'\cos(\Omega_i' t+\theta')\right](\xi_n D_{nn}+\overline{\xi_n}D_{\overline{nn}})=0,$$
$$n=1,2,\cdots \quad (13)$$

where all the excitations are taken to be of the same order, that is

$$\varepsilon_i=\varepsilon f_i,\quad \varepsilon_i'=\varepsilon f_i'\quad f_i,f_i'=0(1). \quad (14)$$

Based on the Floquet theory ([13]), combinations of parametric excitation frequency and amplitude for which (13) has periodic solutions separate the regions of bounded and unbounded motions. These stability boundaries are sought in the form of perturbation expansions ([11]),

$$\xi_n=p_0+\varepsilon p_1+\varepsilon^2 p_2,\quad \omega_n=\tilde{\omega}_n+\varepsilon r_1+\varepsilon^2 r_2. \quad (15)$$

For simplicity, the body of the paper examines the case of tension excitation alone ($\varepsilon_i'=0$), where (13) reduces to

$$\dot{\xi}_n - j\omega_n\xi_n - \left[\varepsilon\sum_{i=1}^{k} f_i\cos(\Omega_i t+\theta_i)\right](\xi_n E_{nn}+\overline{\xi}_n E_{\overline{nn}})=0$$

$$n=1,2,\cdots. \tag{16}$$

Stability results for speed fluctuations and simultaneous tension and speed fluctuations are given in the Appendix. Substitution of (15) into (16) gives the sequence of perturbation problems

$$\dot{p}_0 - j\tilde{\omega}_n p_0 = 0 \tag{17}$$

$$\dot{p}_1 - j\tilde{\omega}_n p_1 = jr_1 p_0 + \left[\sum_{i=1}^{k} f_i\cos(\Omega_i t+\theta_i)\right][p_0 E_{nn}+\overline{p}_0 E_{\overline{nn}}] \tag{18}$$

$$\dot{p}_2 - j\tilde{\omega}_n p_2 = jr_1 p_1 + jr_2 p_0 + \left[\sum_{i=1}^{k} f_i\cos(\Omega_i t+\theta_i)\right][p_1 E_{nn}+\overline{p}_1 E_{\overline{nn}}]. \tag{19}$$

The periodic solution of (17) is

$$p_0 = ae^{i\tilde{\omega}_n t}. \tag{20}$$

Substitution of (20) into (18) yields

$$\dot{p}_1 - j\tilde{\omega}_n p_1 = jr_1 ae^{j\tilde{\omega}_n t} + \sum_{i=1}^{k}\frac{f_i}{2}\{aE_{nn}[e^{j[(\Omega_i+\tilde{\omega}_n)t+\theta_i]} + e^{-j[(\Omega_i-\tilde{\omega}_n)t+\theta_i]}] + \overline{a}E_{\overline{nn}}[e^{j[(\Omega_i-\tilde{\omega}_n)t+\theta_i]}+e^{-j[(\Omega_i+\tilde{\omega}_n)t+\theta_i]}]\}. \tag{21}$$

1 Primary Instability. In general, the sole secular term in (21) is $jr_1 ae^{j\tilde{\omega}_n t}$ and its elimination leads to the trivial solution ($a=0$) or $r_1=0$ (secondary instability, considered later). When any excitation frequency is near $2\omega_n$, however, additional secular terms exist. In this case, $\Omega_l \approx 2\omega_n$ (that is, $\Omega_l = 2\tilde{\omega}_n$) and $\Omega_i \neq 2\omega_n$ for $i\neq l$. The periodicity condition demands that secular terms vanish, yielding

$$jr_1 a + \frac{f_l}{2}\overline{a}E_{\overline{nn}}e^{j\theta_l}=0. \tag{22}$$

Separating the real and imaginary parts leads to

$$\begin{bmatrix} -r_1+\frac{f_l}{2}\mathrm{Im}(E_{\overline{nn}}e^{j\theta_l}) & \frac{f_l}{2}\mathrm{Re}(E_{\overline{nn}}e^{j\theta_l}) \\ -\frac{f_l}{2}\mathrm{Re}(E_{\overline{nn}}e^{j\theta_l}) & r_1+\frac{f_l}{2}\mathrm{Im}(E_{\overline{nn}}e^{j\theta_l}) \end{bmatrix}\begin{bmatrix}\mathrm{Im}(a)\\ \mathrm{Re}(a)\end{bmatrix}=0. \tag{23}$$

For a nontrivial solution of (23) to exist,

$$r_1 = \pm\frac{f_l}{2}|E_{\overline{nn}}|. \tag{24}$$

With $r_1=\pm(f_l/2)|E_{\overline{nn}}|$, a solution of (21) is

$$p_1 = be^{j\tilde{\omega}_n t} + \sum_{i=1}^{k} j\frac{f_i aE_{nn}}{2\Omega_i}[-e^{j[(\Omega_i+\tilde{\omega}_n)t+\theta_i]}+e^{-j[(\Omega_i-\tilde{\omega}_n)t+\theta_i]}] - \sum_{i=1,i\neq l}^{k} j\frac{f_i\overline{a}E_{\overline{nn}}}{2(\Omega_i-\Omega_l)}e^{j[(\Omega_i-\tilde{\omega}_n)t+\theta_i]} + \sum_{i=1}^{k} j\frac{f_i\overline{a}E_{\overline{nn}}}{2(\Omega_i+\Omega_l)}e^{-j[(\Omega_i+\tilde{\omega}_n)t+\theta_i]}. \tag{25}$$

Substitution of (20) and (25) into (19) yields

$$\dot{p}_2 - j\tilde{\omega}_n p_2 = jr_2 ae^{j\tilde{\omega}_n t} + jr_1 be^{j\tilde{\omega}_n t} + \frac{f_l}{2}E_{\overline{nn}}\overline{b}e^{j(\tilde{\omega}_n t+\theta_l)} + j\sum_{i=1,i\neq l}^{k}\frac{f_i^2 a|E_{\overline{nn}}|^2}{4(\Omega_i-\Omega_l)}e^{j\tilde{\omega}_n t} - j\sum_{i=1}^{k}\frac{f_i^2 a|E_{\overline{nn}}|^2}{4(\Omega_i+\Omega_l)}e^{j\tilde{\omega}_n t} + \text{N.S.T.} \tag{26}$$

where N.S.T. denotes nonsecular terms. Elimination of secular terms from p_2 requires

$$jr_1 b + \frac{f_l}{2}\overline{b}E_{\overline{nn}}e^{j\theta_l} = -ja\left\{r_2+|E_{\overline{nn}}|^2\left[\sum_{i=1,i\neq l}^{k}\frac{f_i^2}{4(\Omega_i-\Omega_l)} - \sum_{i=1}^{k}\frac{f_i^2}{4(\Omega_i+\Omega_l)}\right]\right\}. \tag{27}$$

Considering Re(b) and Im(b) as the unknowns, (22)–(24) show that the coefficient matrix in (27) is singular. The solvability condition for (27) leads to r_2, and the final boundary curves are obtained from (15) as

$$\omega_n = \frac{\Omega_l}{2}\pm\frac{\varepsilon_l}{2}|E_{\overline{nn}}|+|E_{\overline{nn}}|^2\left[-\sum_{i=1,i\neq l}^{k}\frac{\varepsilon_i^2}{4(\Omega_i-\Omega_l)} + \sum_{i=1}^{k}\frac{\varepsilon_i^2}{4(\Omega_i+\Omega_l)}\right]. \tag{28}$$

Using $\Omega_l = 2\omega_n + O(\varepsilon)$, (28) is converted to

$$\Omega_l = 2\omega_n \pm \varepsilon_l|E_{\overline{nn}}| - |E_{\overline{nn}}|^2\left[-\sum_{i=1,i\neq l}^{k}\varepsilon_i^2\frac{2\omega_n}{\Omega_i^2-4\omega_n^2}+\frac{\varepsilon_l^2}{8\omega_n}\right]. \tag{29}$$

Equation (29) applies for moving, tensioned beams. When specialized to the moving string ($\alpha=0$), the result (29) can be expressed entirely in terms of system parameters with the following expressions obtained from (11):

$$E_{\overline{nn}} = (1-e^{-2jn\pi\gamma_0})/(4\gamma_0),\quad E_{nn} = jn\pi(1+\gamma_0^2)/2. \tag{30}$$

Up to the first order of perturbation, the stability boundaries as given by (29) are determined solely by the root cause parametric excitation $\Omega_l \approx 2\omega_n$ with no effect from excitations at other frequencies (see Eq. (24)). Changes in the stability boundaries caused by the presence of multiple parametric excitation terms are evident at higher orders of perturbation. Note that the primary instability boundaries (29) are not affected by the phase angles θ_i between the multiple excitations. These features are also reflected in the stability boundaries for speed excitation and simultaneous speed and tension excitation derived from (13) ((52) and (55) in the Appendix). In the simultaneous tension and speed excitation case ((55)–(57)), the tension and speed fluctuations share a common frequency component that excites instability. This is typical of automotive belt drives where the tension and speed both fluctuate at the engine firing frequency.

Figure 1 compares the stability boundaries for a moving string ($\alpha=0$) under two simultaneous tension excitations ($\varepsilon_i'=0$) obtained by three methods: first-order perturbation, second-order perturbation, and numerical methods. $\Omega_1 \approx 2\omega_1$ is the root cause of primary instability. Numerical boundaries are determined by examining the eigenvalues of the numerically integrated fundamental matrix of (16) for varying Ω_1 and ε_1. The first-order stability boundaries (dotted lines) do not capture the effects of the second parametric excitation (Ω_2) and are the same as for mono-frequency excitation. The second-order boundaries reflect the impact of the second excitation, and the entire instability region shifts (solid lines). The classical result that parametric instability

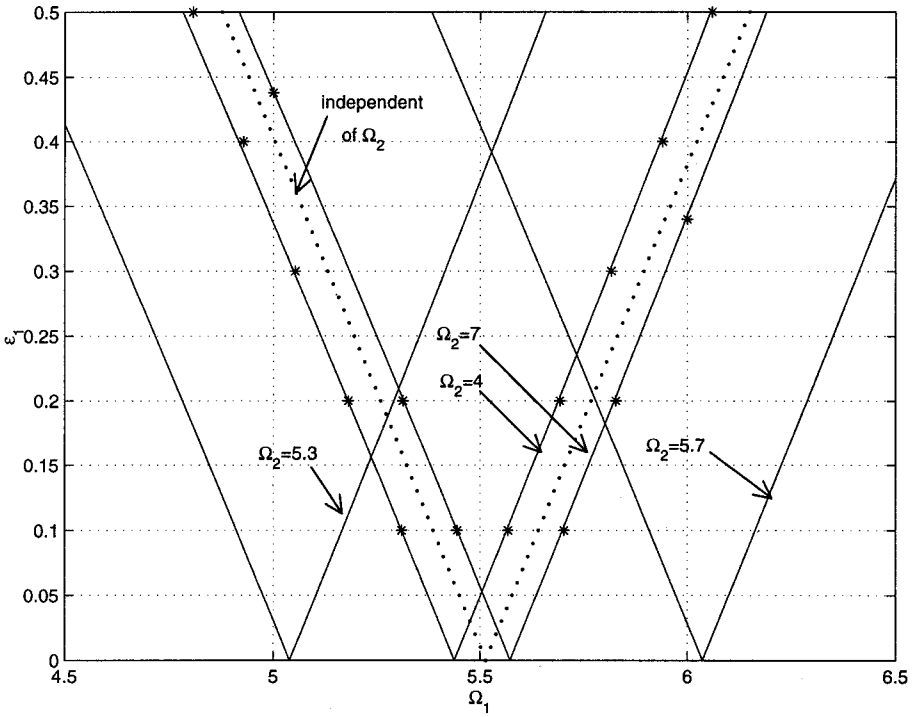


Fig. 1 Comparison among numerical results (*), first-order perturbation (dots) and second-order perturbation (solid curves) of the first mode ($\omega_1=2.76$) primary instability region of an axially moving string under two parametric tension excitations. $\gamma=0.35$, $\varepsilon_2=0.35$.

occurs when an excitation frequency is twice a natural frequency does not hold when multiple parametric excitations are present. To see this, consider (29) when two excitations exist,

$$\Omega_1=2\omega_n\pm\varepsilon_1|E_{nn}^-|-|E_{nn}^-|^2\left[\frac{\varepsilon_1^2}{8\omega_n}-\frac{2\varepsilon_2^2\omega_n}{(\Omega_2^2-4\omega_n^2)}\right].$$

Because of the excitation at frequency Ω_2, the cusp ($\varepsilon_1\to 0$) moves from $\Omega_1=2\omega_n$ to $\Omega_1=2\omega_n+2|E_{nn}^-|^2\varepsilon_2^2\omega_n/(\Omega_2^2-4\omega_n^2)$. The whole instability region shifts accordingly, and parametric instability occurs at higher excitation frequency ($\Omega_1>2\omega_n$) when $\Omega_2>2\omega_n$ and lower excitation frequency ($\Omega_1<2\omega_n$) when $\Omega_2<2\omega_n$. Note that a separate analysis including additional secular terms is required for the case $\Omega_2\approx\Omega_1\approx 2\omega_n$.

The continuous dependence of the first mode primary instability region with the two excitation amplitudes $\varepsilon_{1,2}$ is illustrated in Fig. 2. Ω_1 causes the primary instability. The shift of the instability region away from $\Omega_1=2\omega_1$ due to the second excitation is apparent from the drift in the cusp at $\varepsilon_1=0$.

2 Secondary Instability. Primary instabilities are characterized by a response frequency of half the parametric excitation frequency. In automotive belt drives, however, transverse belt vibration frequently occurs where the belt frequency is the same as the engine firing frequency. This is characteristic of secondary instability where a parametric excitation frequency is close to one of the system natural frequencies ($\Omega_l\approx\omega_n$).

In the absence of primary instability, the only secular term in (21) is $jr_1ae^{j\tilde{\omega}_nt}$. Specifying $r_1=0$, (25) is again a solution of (21) except the second summation allows any value of i. Substitution of this solution into (19) yields

$$\begin{aligned}\dot{p}_2-j\tilde{\omega}_np_2=jr_2ae^{j\tilde{\omega}_nt}&\\ -j\sum_{i=1}^{k}\sum_{p=1}^{k}\frac{f_if_p\overline{a}E_{nn}^-E_{nn}}{4(\Omega_p-2\tilde{\omega}_n)}e^{j[(\Omega_i+\Omega_p-\tilde{\omega}_n)t+(\theta_i+\theta_p)]}&\\ -j\sum_{i=1}^{k}\sum_{p=1}^{k}\frac{f_if_p\overline{a}E_{nn}^-\overline{E}_{nn}}{4\Omega_p}e^{j[(\Omega_i+\Omega_p-\tilde{\omega}_n)t+(\theta_i+\theta_p)]}&\\ +j\sum_{i=1}^{k}\frac{f_i^2a|E_{nn}^-|^2}{4(\Omega_i-2\tilde{\omega}_n)}e^{j\tilde{\omega}_nt}-j\sum_{i=1}^{k}\frac{f_i^2a|E_{nn}^-|^2}{4(\Omega_i+2\tilde{\omega}_n)}e^{j\tilde{\omega}_nt}&\\ +\text{N.S.T.}\end{aligned}\tag{31}$$

Elimination of secular terms in (31) for the case when $\Omega_l\approx\omega_n$ and $\Omega_i\neq\omega_n$ for $i\neq l$ leads to the secondary instability boundaries

$$\begin{aligned}\Omega_l=\omega_n\pm\frac{\varepsilon_l^2}{\omega_n}\text{Im}(E_{nn})|E_{nn}^-|&\\ -|E_{nn}^-|^2\left[-\sum_{i=1,i\neq l}^{k}\varepsilon_i^2\frac{\omega_n}{\Omega_i^2-4\omega_n^2}+\frac{\varepsilon_l^2}{3\omega_n}\right].&\end{aligned}\tag{32}$$

Analogous results for speed and tension/speed excitation are given in the Appendix.

3 Simultaneous Primary and Secondary Instability. In a system under multiple parametric excitations, a mode may be simultaneously excited to primary instability by one excitation and secondary instability by another. This situation is expected in automotive belt drives as discussed later. With simultaneous instability, the instability boundaries may be significantly different from those of either excitation acting individually. Figure 3(*a*) shows the first mode primary and secondary instability boundaries for a moving string for a single excitation (Ω_1). The secondary

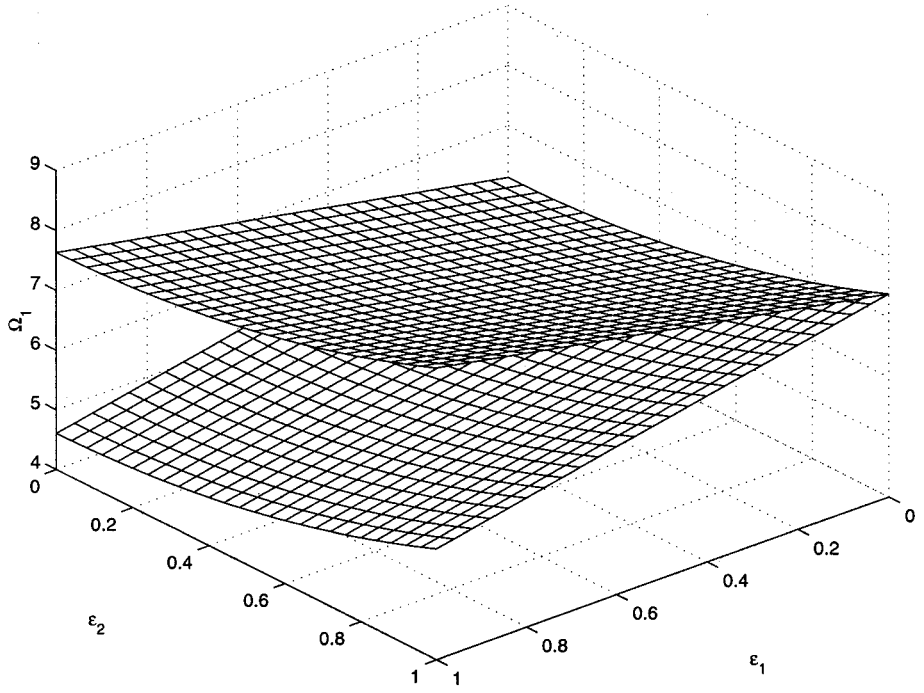


Fig. 2 Continuous dependence on excitation amplitudes of the first mode ($\omega_1=3.07$) primary instability region of an axially moving string under two parametric tension excitations. $\Omega_2=7, \gamma=0.15$.

instability region is characteristically much narrower than the primary one. When a second tension excitation exists with frequency $\Omega_2=2\Omega_1\approx 2\omega_n$, primary instability from Ω_2 occurs near the secondary instability from Ω_1. Perturbation analysis of this dual excitation case gives the instability region

$$\Omega_1=\omega_n\pm\frac{\varepsilon_2}{2}|E_{nn}^{-}|-|E_{nn}^{-}|^2\left[\frac{\varepsilon_1^2}{3\omega_n}+\frac{\varepsilon_2^2}{16\omega_n}\right]\mp\frac{\varepsilon_1^2}{\omega_n}\mathrm{Im}(E_{nn})|E_{nn}^{-}|. \tag{33}$$

As shown in Fig. 3(*b*), the coincidence of the primary instability caused by Ω_2 and the secondary instability of Ω_1 significantly widens the secondary instability region. Notice that the unstable region is *wider* for small ε_1. Considering the dual excitation case with $\Omega_2=(1/2)\Omega_1$, the presence of a simultaneous secondary instability from Ω_2 impacts the primary instability at $\Omega_1\approx 2\omega_n$ as seen by comparing Figs. 3(*a*) and 3(*c*). Here, the primary instability region narrows overall and *closes* for nonzero amplitude of the excitation causing primary instability ($\varepsilon_1\approx 0.1$). This phenomenon is further depicted in Figs. 4 and 5 for instability in the second mode. Figures 4(*a*) and 4(*b*) contrast the dependence of the second mode primary instability region on excitation amplitude and speed for the cases with and without simultaneous secondary instability. In Fig. 4(*b*), notice that the instability region closes for nonzero $\varepsilon_1\approx 0.3$ even though each of the Ω_1 and Ω_2 excitations induce instability individually. While the width of the primary instability regions widen with excitation amplitude (Fig. 4(*a*)), the width of the simultaneous primary/secondary region may decrease with excitation amplitude (Fig. 4(*b*) for small ε_1). One can see similar interplay between the primary and secondary instabilities in Fig. 5, which is analogous to Fig. 4 except the focus is on the secondary instability.

The widening of the secondary instability region ($\Omega_1\approx\omega_n$) when a second excitation $\Omega_2=2\Omega_1\approx 2\omega_n$ is present (Fig. 3(*b*)) has implications for practical systems. In automotive belt drives (and other systems), the excitation is periodic but not sinusoidal. Because of the integer harmonics of the fundamental frequency (the firing frequency in automotive belt drives) in the excitation

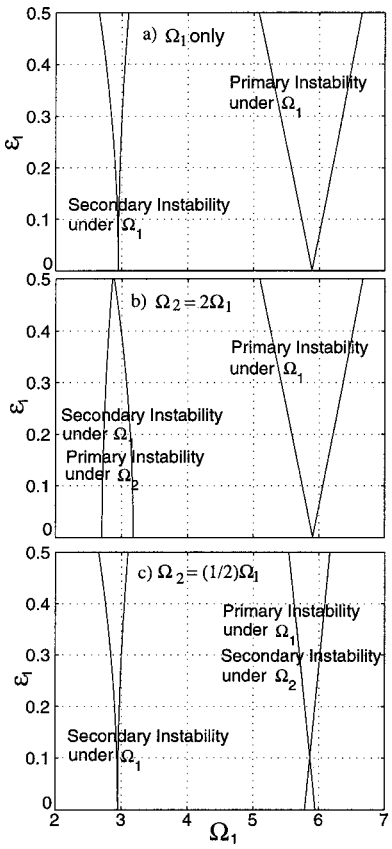


Fig. 3 First mode ($\omega_1=2.95$) stability boundaries of an axially moving string caused by three parametric excitation combinations for $\gamma=0.25$: (*a*) single excitation Ω_1, (*b*) two excitations $\Omega_2=2\Omega_1, \varepsilon_2=0.3$, and (*c*) two excitations $\Omega_2=(1/2)\Omega_1, \varepsilon_2=0.3$

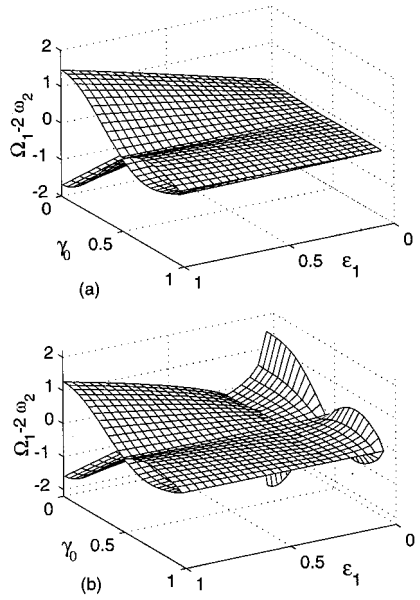


Fig. 4 Dependence of the second mode ($\omega_2=2\pi(1-\gamma_0^2)$) moving string principal instability region on translation speed for (*a*) single excitation, $\Omega_1\approx 2\omega_2$, (*b*) two excitations, $\Omega_1\approx 2\omega_2$ and $\Omega_2=(1/2)\Omega_1\approx\omega_2, \varepsilon_2=0.3$

spectrum, simultaneous primary and secondary parametric instability is likely. This may explain the common observation of belt span vibration at the engine firing frequency in automotive drives. To explain these observations with a monofrequency excitation model of secondary instability, large excitation amplitudes are required because of the narrowness of the secondary instability region (Fig. 3(*a*)) and the inherent damping. The behavior is more plausibly understood with a multifrequency excitation model for realistic excitation amplitudes.

As the translation speed increases, both of the secondary and primary instability regions narrow and even close at some speeds

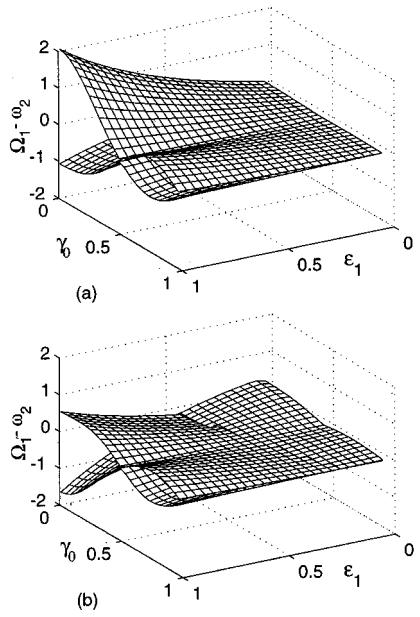


Fig. 5 Dependence of the second mode ($\omega_2=2\pi(1-\gamma_0^2)$) moving string secondary instability region on translation speed for (*a*) single excitation, $\Omega_1\approx\omega_2$, (*b*) two excitations, $\Omega_1\approx\omega_2$ and $\Omega_2=2\Omega_1\approx 2\omega_2, \varepsilon_2=0.3$

(Figs. 4(*a*) and 5(*a*)). As pointed out by Mockensturm et al. [1] for primary instability, there are n subcritical translation speeds where the nth mode primary instability region closes. The same holds true for secondary instability.

4 Combination Instability. This section addresses sum and difference type combination instabilities involving two modes. Taking the nth and mth modes, the discretized equations are obtained from the expansion

$$W=\xi_n(t)\Phi_n(x)+\overline{\xi}_n(t)\overline{\Phi}_n(x)+\xi_m(t)\Phi_m(x)+\overline{\xi}_m(t)\overline{\Phi}_m(x)$$
$$=2\,\mathrm{Re}[\,\xi_n(t)\Phi_n(x)+\xi_m(t)\Phi_m(x)\,]. \tag{34}$$

Considering tension fluctuations only, use of (34) in Galerkin discretization of (8) yields

$$\dot{\xi}_n-j\omega_n\xi_n-\left[\varepsilon\sum_{i=1}^{k} f_i\cos(\Omega_i t+\theta_i)\right]$$
$$\times(\xi_n E_{nn}+\overline{\xi}_n E_{n\bar{n}}+\xi_m E_{mn}+\overline{\xi}_m E_{m\bar{n}}) \tag{35}$$

$$\dot{\xi}_m-j\omega_m\xi_m-\left[\varepsilon\sum_{i=1}^{k} f_i\cos(\Omega_i t+\theta_i)\right]$$
$$\times(\xi_n E_{nm}+\overline{\xi}_n E_{n\bar{m}}+\xi_m E_{mm}+\overline{\xi}_m E_{m\bar{m}}). \tag{36}$$

Motivated by the expected sum-type instability when $\Omega_l\approx\omega_n+\omega_m$, the solution forms are chosen as

$$\Omega_l=(\omega_n+\omega_m)-2\varepsilon r_1-2\varepsilon^2 r_2$$
$$=(\omega_n-\varepsilon r_1-\varepsilon^2 r_2)+(\omega_m-\varepsilon r_1-\varepsilon^2 r_2)=\tilde{\omega}_n+\tilde{\omega}_m$$
$$\Rightarrow \begin{matrix}\xi_n=p_0+\varepsilon p_1+\varepsilon^2 p_2, & \xi_m=q_0+\varepsilon q_1+\varepsilon^2 q_2\\ \omega_n=\tilde{\omega}_n+\varepsilon r_1+\varepsilon^2 r_2, & \omega_m=\tilde{\omega}_m+\varepsilon r_1+\varepsilon^2 r_2.\end{matrix} \tag{37}$$

Substitution of (37) into (35) and (36) yields

$$\dot{p}_0-j\tilde{\omega}_n p_0=0,\ \ \dot{q}_0-j\tilde{\omega}_m q_0=0 \tag{38}$$

$$\dot{p}_1-j\tilde{\omega}_n p_1=jr_1 p_0+\left[\sum_{i=1}^{k} f_i\cos(\Omega_i t+\theta_i)\right][p_0 E_{nn}+\overline{p}_0 E_{\bar{n}\bar{n}}$$
$$+q_0 E_{mn}+\overline{q}_0 E_{m\bar{n}}] \tag{39}$$

$$\dot{q}_1-j\tilde{\omega}_m q_1=jr_1 q_0+\left[\sum_{i=1}^{k} f_i\cos(\Omega_i t+\theta_i)\right][p_0 E_{nm}+\overline{p}_0 E_{\bar{n}m}$$
$$+q_0 E_{mm}+\overline{q}_0 E_{m\bar{m}}] \tag{40}$$

and similar equations for p_2 and q_2. The periodic solutions of (38) are $p_0=a_n e^{j\tilde{\omega}_n t}$, $q_0=a_m e^{j\tilde{\omega}_m t}$. With these solutions, elimination of secular terms in (39) and (40) for $\Omega_l\approx\omega_n+\omega_m$ requires

$$jr_1 a_n+\frac{f_l}{2}\overline{a}_m E_{m\bar{n}} e^{j\theta_l}=0 \quad jr_1 a_m+\frac{f_l}{2}\overline{a}_n E_{n\bar{m}} e^{j\theta_l}=0. \tag{41}$$

After solution of (41) and use of $\Omega_l=\omega_n+\omega_m+O(\varepsilon)$, (37) gives

$$\Omega_l=\omega_n\pm\omega_m+\varepsilon_l\sqrt{E_{n\bar{m}}\overline{E}_{m\bar{n}}} \tag{42}$$

A natural extension to second-order perturbation was also calculated. As with primary and secondary instabilities, these boundaries are independent of the phasing between the different excitation frequencies. Figure 6 shows the sum type instability region of a moving string under tension fluctuation obtained by first and second-order perturbation. Note the scaling of Fig. 6; the combination instability region is much narrower than the primary instability region (Fig. 1). The effects from multiple excitation frequencies, including the shift of the entire instability region, are minimal, and first-order approximations that do not capture these

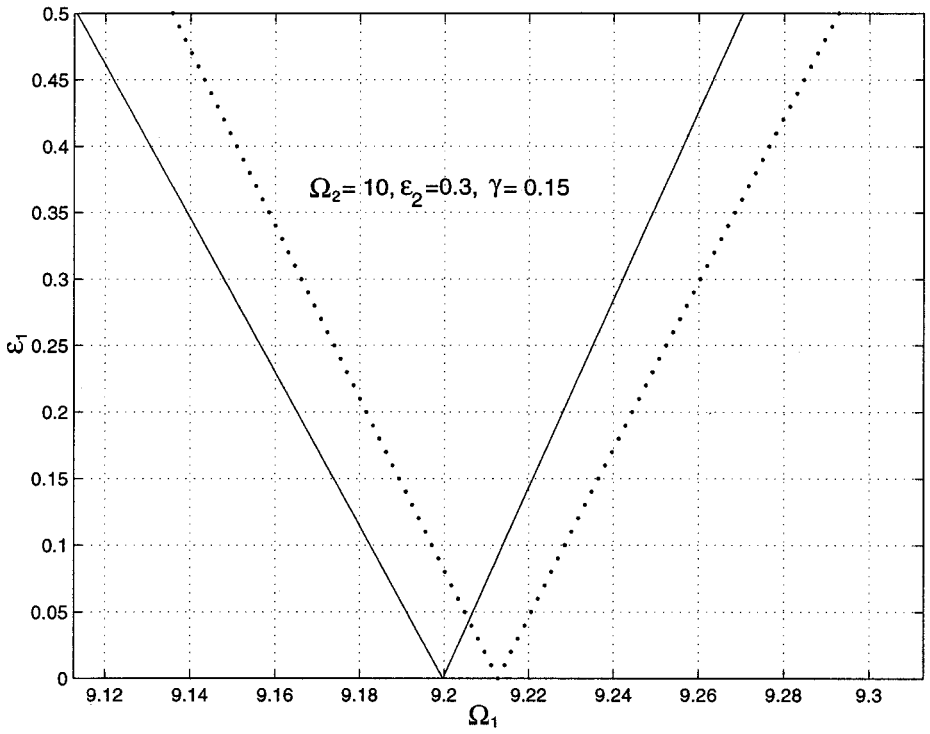


Fig. 6 Moving string stability boundaries of first (ω_1=3.071) and second mode (ω_2 =6.142) sum-type combination instability ($\Omega_1 \approx \omega_1+\omega_2$) for two parametric tension excitations. Dotted curves denote first-order perturbation, and solid curves denote second-order perturbation.

effects appear justified for practical systems. See the Appendix for speed excitation and simultaneous speed and tension excitation results.

To examine possible difference type combination instabilities where $\Omega_l \approx \omega_n - \omega_m$, the approximate solutions are constructed as

$$\begin{aligned}\Omega_l &= (\omega_n-\omega_m)-2\varepsilon r_1-2\varepsilon^2 r_2\\ &=(\omega_n-\varepsilon r_1-\varepsilon^2 r_2)-(\omega_m+\varepsilon r_1+\varepsilon^2 r_2)=\tilde{\omega}_n+\tilde{\omega}_m\\ &\Rightarrow \begin{matrix}\xi_n=p_0+\varepsilon p_1+\varepsilon^2 p_2, & \xi_m=q_0+\varepsilon q_1+\varepsilon^2 q_2\\ \omega_n=\tilde{\omega}_n+\varepsilon r_1+\varepsilon^2 r_2, & \omega_m=\tilde{\omega}_m-\varepsilon r_1-\varepsilon^2 r_2.\end{matrix}\end{aligned} \tag{43}$$

In this case, the first-order stability boundaries are

$$\Omega_l=(\omega_n-\omega_m)\pm\varepsilon_l\sqrt{E_{mn}E_{nm}}. \tag{44}$$

Closed-form evaluation of the inner products in (44) gives complex values for Ω_l. This implies that there are no difference type instabilities up to first-order perturbation.

The results given in (29), (32), (33), (42), and the Appendix generalize those of Mockensturm et al. [1], where tension fluctuation is examined, and Pakdemirli and Ulsoy [2], where speed fluctuation is considered. In those analyses, the parametric excitation is restricted to a single harmonic term of either tension or speed excitation, only first-order approximations are derived, and secondary instabilities are not investigated.

Nonlinear Response Amplitude for Primary Instability

By including the midplane stretching nonlinearity in (4), transverse vibration amplitudes are determined for the principal parametric resonance regions. Allowing moderate displacements with the ordering $v=O(\sqrt{\varepsilon})$ ([1]), the nonlinear form of (7) is

$$\begin{aligned}&AW_t+BW+\varepsilon\sum_{i=1}^{k'} f_i' \cos(\Omega_i' t+\theta_i')\{\sin(\Omega_i' t+\theta_i')C\\ &\quad+\Omega_i' \cos(\Omega_i' t+\theta_i')D\}W\\ &\quad-\varepsilon\left(\sum_{i=1}^{k} f_i \cos(\Omega_i t+\theta_i)+\frac{1}{2}\zeta\int_0^1 {v_x}^2 dx\right)EW\\ &\quad+\left(\varepsilon\sum_{i=1}^{k'} f_i' \sin(\Omega_i' t+\theta_i')\right)^2 EW=0.\end{aligned} \tag{45}$$

Galerkin discretization of (45) using (12) yields

$$\begin{aligned}&\dot{\xi}_n-j\omega_n\xi_n-\varepsilon\left[\sum_{i=1}^{k} f_i \cos(\Omega_i t+\theta_i)+\frac{1}{2}(d_1\xi_n^2+d_2\xi_n\overline{\xi}_n+d_3\overline{\xi}_n^2)\right]\\ &\times(\xi_n E_{nn}+\overline{\xi}_n E_{\overline{nn}})+\varepsilon\sum_{i=1}^{k'} f_i'\{\sin(\Omega_i' t+\theta_i')(\xi_n C_{nn}+\overline{\xi}_n C_{\overline{nn}})\\ &+\Omega_i' \cos(\Omega_i' t+\theta_i')(\xi_n D_{nn}+\overline{\xi}_n D_{\overline{nn}})\}\\ &+\left(\varepsilon\sum_{i=1}^{k'} \sin(\Omega_i' t+\theta_i')\right)^2(\xi_n\langle E\Phi_n,\Phi_n\rangle+\overline{\xi}_n\langle E\overline{\Phi}_n,\Phi_n\rangle)\\ &\qquad=0 \quad n=1,2,\ldots\end{aligned} \tag{46}$$

$$d_1=\int_0^1\left(\frac{d\psi_n}{dx}\right)^2 dx,\quad d_2=\int_0^1\left(\frac{d\psi_n}{dx}\frac{d\overline{\Psi}_n}{dx}\right)dx,$$

$$d_3=\int_0^1\left(\frac{d\overline{\psi}_n}{dx}\right)^2 dx. \tag{47}$$

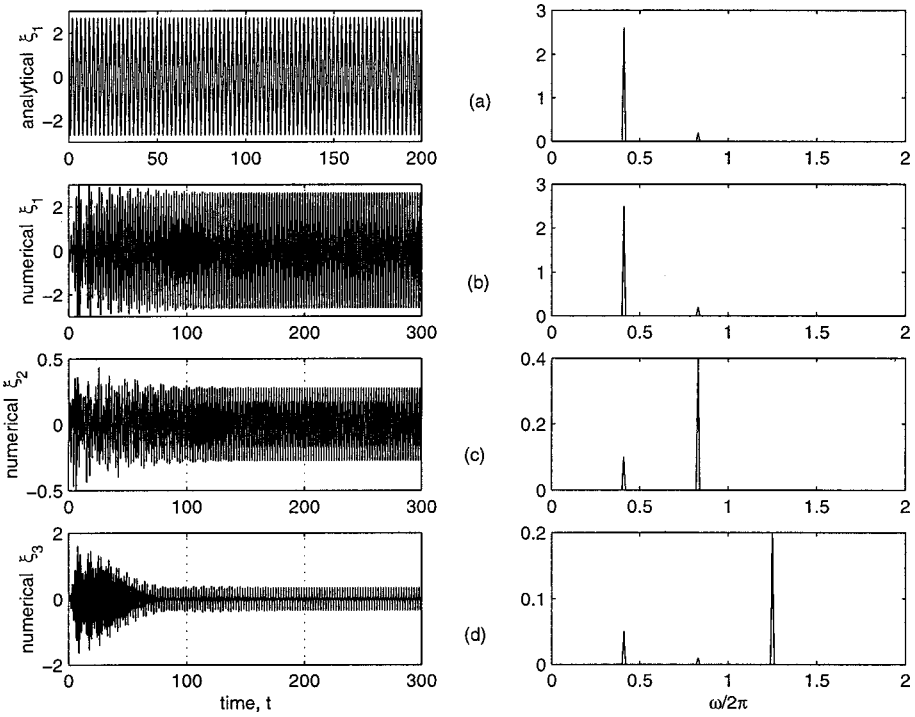


Fig. 7 Time histories and spectra of the modal response under first-mode primary instability with single frequency tension excitation. (*a*) analytical approximation, (*b*)–(*d*) numerical integration of coupled equations from a three-mode Galerkin discretization. $\gamma = 0.4$, $\omega_1 = 0.84\pi$, $\varepsilon_1 = 0.35$, $\sigma = -0.2$, $\Omega_1 = 2\omega_1 - 2\varepsilon_1\sigma = 1.68\pi$.

The method of multiple scales is applied to (46) with the expansion

$$\xi_n(t;\varepsilon) = p_0(t,\tau) + \varepsilon p_1(t,\tau) \tag{48}$$

where t and $\tau = \varepsilon t$ are the fast and slow time scales. The time derivative is defined as $d/dt \rightarrow \partial/\partial t + \varepsilon(\partial/\partial\tau)$. Insertion of (48) into (46) gives $\partial p_0/\partial t - j\omega_n p_0 = 0$ and a similar inhomogeneous equation for p_1.

The problem of interest is that when speed and tension fluctuation share a common frequency component that simultaneously causes primary instability, that is $\Omega_l = \Omega_l' \approx 2\omega_n$. The nearness of Ω_l and Ω_l' to $2\omega_n$ is represented by $\Omega_l = \Omega_l' = 2\omega_n - 2\varepsilon\sigma$, where $\sigma = O(1)$. With the solution $p_0 = K_n(\tau)e^{j\omega_n t} = \rho_n(\tau)e^{j(\beta_n(\tau)+\sigma\tau)}e^{j\omega_n t}$, elimination of secular terms in the differential equation for p_1 yields the conditions

$$\begin{aligned} \frac{d\rho_n}{d\tau} &= \varepsilon\rho_n[Q_n \sin(2\beta_n) + P_n \cos(2\beta_n) + R_n\rho_n^2] \\ \frac{d\beta_n}{d\tau} &= \varepsilon\left[-\frac{\sigma}{2} - P_n \sin(2\beta_n) + Q_n \cos(2\beta_n) + S_n\rho_n^2\right] \end{aligned} \tag{49}$$

where P_n, Q_n, respectively, are the real and imaginary parts of $[-(f_l'/2)(jC_{nn}^{-} + \Omega_i' D_{nn}^{-}) + (f_l/2)E_{nn}^{-}]$ and R_n, S_n are the real and imaginary parts of $(\zeta/2)[d_2 E_{nn} + d_1 E_{nn}^{-}]$. The nontrivial equilibria of (49) are

$$(\rho_n^0)^2 = \frac{\sigma S_n \pm \sqrt{(\sigma S_n)^2 - 4(R_n^2 + S_n^2)\left(\frac{\sigma^2}{4} - P_n^2 - Q_n^2\right)}}{2(R_n^2 + S_n^2)}. \tag{50}$$

Stability analyses reveal that the limit cycle with larger amplitude is stable, and the one with lower amplitude is unstable. When specialized to the moving string and only tension excitation, the results of Mockensturm et al. [1] are recovered.

The foregoing single mode analysis ignores the possibility of other modes being excited through nonlinear coupling. Figure 7 compares the nonlinear response from (50) with numerical integration of the coupled nonlinear equations from a three-term Galerkin discretization of (45). Figures 7(*a*) and 7(*b*) show that the amplitude of the first mode response is accurately predicted by single-mode analysis. There is, however, considerable energy transfer into other modes that is not captured in the single mode analysis (Figs. 7(*c*) and 7(*d*)).

Conclusions

Closed-form expressions are derived for the stability of axially moving media subjected to multifrequency parametric excitation from simultaneous tension and speed fluctuations.

1 The effects of the parametric excitations at frequencies other than the one that is the root cause of an instability are evident only in a second order perturbation. These effects, however, can be substantial. In a first-order solution, the set of primary and combination instability regions for multifrequency excitation are the superposition of the instability regions for the individual monofrequency excitations.

2 The primary instability region that one expects when an excitation frequency is twice a natural frequency shifts as a result of the multiple parametric excitations. The classical 2:1 ratio between excitation and natural frequencies no longer holds, and primary parametric instability occurs at different excitation frequencies higher or lower than $2\omega_n$ (Fig. 1).

3 Secondary resonances that are typically considered benign widen substantially when a second parametric excitation simultaneously excites a primary instability in the same mode (Fig. 3). Such conditions occur naturally when the fundamental frequency of periodic (but not sinusoidal) parametric excitation drives secondary instability and the first harmonic drives primary instability.

This provides a plausible explanation for the practically important case of a serpentine belt span oscillating at the same frequency as the engine firing frequency.

4 Instability regions for combination resonances of the sum type are significantly narrower than those for primary instability. For practical system damping and realistic excitation amplitudes, a first-order approximation appears to be sufficient. Difference type combination resonances do not occur even with multiple excitation frequencies.

5 The nonlinear response amplitude under primary instability is determined solely by the excitation causing the instability and is independent of other excitations in a first-order approximation. Transfer of energy from the unstable mode to other modes as a result of nonlinear coupling is apparent in a numerical solution (Fig. 7), though not captured in a first-order approximation.

Acknowledgment

The authors thank Mark IV Automotive/Dayco Corporation for their support of this project.

Appendix

Results in this Appendix hold for traveling strings and beams. For traveling strings ($\alpha=0$), the following relations are helpful (also see (30)):

$$C_{\bar{n}\bar{n}}=(1-e^{-2jn\pi\gamma_0})/2, \quad D_{\bar{n}\bar{n}}=0. \tag{51}$$

A Primary Instability Caused by Speed Fluctuation: $\Omega_l' \approx 2\omega_n$, $\varepsilon_i=0$

$$\Omega_l'=2\omega_n\pm\varepsilon_l'|-jC_{\bar{n}\bar{n}}+2\omega_nD_{\bar{n}\bar{n}}| -|-jC_{\bar{n}\bar{n}}+2\omega_nD_{\bar{n}\bar{n}}|^2\left[-\sum_{i=1,i\neq l}^{k'}\varepsilon_i'^2\frac{2\omega_n}{\Omega_i'^2-4\omega_n^2}+\frac{\varepsilon_l'^2}{8\omega_n}\right] +\sum_{i=1}^{k'}\left(\frac{\varepsilon_i'}{2}\right)^2|E_{nn}| \tag{52}$$

B Secondary Instability Caused by Speed Fluctuation: $\Omega_l' \approx \omega_n$, $\varepsilon_i=0$

$$\Omega_l'=\omega_n\pm\frac{\varepsilon_l'^2}{\omega_n}\mathrm{Im}(E_{nn})|-jC_{\bar{n}\bar{n}}+\omega_nD_{\bar{n}\bar{n}}| -|-jC_{\bar{n}\bar{n}}+\omega_nD_{\bar{n}\bar{n}}|^2\left[-\sum_{i=1,i\neq l}^{k'}\varepsilon_i'^2\frac{\omega_n}{\Omega_i'^2-4\omega_n^2}+\frac{\varepsilon_l'^2}{3\omega_n}\right] +\sum_{i=1}^{k'}\left(\frac{\varepsilon_i'}{2}\right)^2|E_{nn}| \tag{53}$$

C Combination Instability Caused by Speed Fluctuation: $\Omega_l' \approx \omega_n+\omega_m$, $\varepsilon_i=0$

$$\Omega_l'=\omega_n+\omega_{nn} \pm\varepsilon_l'\sqrt{(-jC_{\bar{n}\bar{m}}+(\omega_n+\omega_m)D_{\bar{n}\bar{m}})\overline{(-jC_{\bar{m}\bar{n}}+(\omega_n+\omega_m)D_{\bar{m}\bar{n}})}} \tag{54}$$

D Primary Instability Caused by Tension and Speed Fluctuation: $\Omega_l=\Omega_l' \approx 2\omega_n$

$$\Omega_l=\Omega_l'=2\omega_n\pm\sqrt{(\varepsilon_l'|-jC_{\bar{n}\bar{n}}+2\omega_nD_{\bar{n}\bar{n}}|)^2+(\varepsilon_l|E_{\bar{n}\bar{n}}|)^2} -|E_{\bar{n}\bar{n}}|^2\left[-\sum_{i=1,i\neq l}^{k}\varepsilon_i^2\frac{2\omega_n}{\Omega_i^2-4\omega_n^2}+\frac{\varepsilon_l^2}{8\omega_n}\right] -|-jC_{\bar{n}\bar{n}}+2\omega_nD_{\bar{n}\bar{n}}|^2\left[-\sum_{i=1,i\neq l}^{k'}\varepsilon_i'^2\frac{2\omega_n}{\Omega_i'^2-4\omega_n^2}+\frac{\varepsilon_l'^2}{8\omega_n}\right] +\sum_{i=1}^{k'}\left(\frac{\varepsilon_i'}{2}\right)^2|E_{nn}| \tag{55}$$

E Secondary Instability Caused by Tension and Speed Fluctuation: $\Omega_l=\Omega_l' \approx \omega_n$

$$\Omega_l=\Omega_l'=\omega_n\pm\sqrt{\left(\frac{\varepsilon_l^2}{\omega_n}\mathrm{Im}(E_{nn})|E_{\bar{n}\bar{n}}|\right)^2+\left(\frac{\varepsilon_l'^2}{\omega_n}\mathrm{Im}(-jC_{nn}+\omega_nD_{nn})|-jC_{\bar{n}\bar{n}}+\omega_nD_{\bar{n}\bar{n}}|\right)^2} -|E_{\bar{n}\bar{n}}|^2\left[-\sum_{i=1,i\neq l}^{k}\varepsilon_i^2\frac{\omega_n}{\Omega_i^2-4\omega_n^2}+\frac{\varepsilon_l^2}{3\omega_n}\right]-|-jC_{\bar{n}\bar{n}}+\Omega_l'D_{\bar{n}\bar{n}}|^2\left[-\sum_{i=1,i\neq l}^{k}\varepsilon_i'^2\frac{\omega_n}{\Omega_i'^2-4\omega_n^2}+\frac{\varepsilon_l'^2}{3\omega_n}\right]+\sum_{i=1}^{k'}\left(\frac{\varepsilon_i'}{2}\right)^2|E_{nn}| \tag{56}$$

F Combination Instability Caused by Tension and Speed Fluctuation: $\Omega_l=\Omega_l' \approx \omega_n+\omega_m$

$$\Omega_l=\Omega_l'=\omega_n+\omega_m\pm\sqrt{(\varepsilon_l')^2(-jC_{\bar{n}\bar{m}}+(\omega_n+\omega_m)D_{\bar{n}\bar{m}})\overline{(-jC_{\bar{m}\bar{n}}+(\omega_n+\omega_m)D_{\bar{m}\bar{n}})}+(\varepsilon_l)^2E_{\bar{n}\bar{m}}\overline{E}_{\bar{m}\bar{n}}}. \tag{57}$$

References

[1] Mockensturm, E. M., Perkins, N. C., and Ulsoy, A. G., 1996, ''Stability and Limit Cycles of Parametrically Excited, Axially Moving Strings,'' ASME J. Vibr. Acoust., **118**, July, pp. 346–351.

[2] Pakdemirli, M., and Ulsoy, A. G., 1997, ''Stability Analysis of an Axially Accelerating String,'' J. Sound Vib., **203**, No. 5, pp. 815–832.

[3] Mahalingam, S., 1957, ''Transverse Vibrations of Power Transmission Chains,'' Br. J. Appl. Phys., **8**, pp. 145–148.

[4] Mote, C. D. Jr., 1965, ''A Study of Band Saw Vibrations,'' J. Franklin Inst., **279**, No. 6, pp. 430–444.

[5] Mote, C. D. Jr., 1968, ''Dynamic Stability of an Axially Moving Band,'' J. Franklin Inst., **285**, No. 5, pp. 329–346.

[6] Naguleswaran, S., and Williams, C. J. H., 1968, ''Lateral Vibrations of Band Saw Blades, Pulley Belts, and the Like,'' Int. J. Mech. Sci., **10**, pp. 239–250.

[7] Asokanthan, S., and Ariaratnam, S., 1994, ''Flexural Instabilities in Axially Moving Bands,'' ASME J. Vibr. Acoust., **116**, No. 3, pp. 275–279.

[8] Oz, H. R., Pakdemirli, M., and Ozkaya, E., 1998, ''Transition Behavior From String to Beam for an Axially Accelerating Material,'' J. Sound Vib., **215**, No. 3, pp. 571–576.

[9] Chakraborty, G., and Mallik, A. K., 1998, ''Parametrically Excited Non-linear Traveling Beams With and Without External Forcing,'' Nonlinear Dyn., **17**, No. 4, pp. 301–324.

[10] Ulsoy, A. G., Whitesell, J. E., and Hooven, M. D., 1985, ''Design of Belt-Tensioner Systems for Dynamic Stability,'' ASME J. Vibr. Acoust., **107**, No. 3, pp. 282–290.

[11] Nayfeh, A. H., and Mook, D. T., 1979, *Nonlinear Oscillations*, John Wiley and Sons, New York.

[12] Wickert, J. A., and Mote, C. D. Jr., 1990, ''Classical Vibration Analysis of Axially Moving Continua,'' ASME J. Appl. Mech., **57**, pp. 738–744.

[13] McLachlan, N. W., 1947, *Theory and Applications of Mathieu Functions*, Oxford University Press, New York.

Dynamic Analysis of Rectilinear Motion of a Self-Propelling Disk With Unbalance Masses

T. Das
Graduate Student

R. Mukherjee
Associate Professor
Assoc. Mem. ASME

Department of Mechanical Engineering,
Michigan State University,
2555 Engineering Building,
East Lansing, MI 48824-1226

This paper investigates the dynamics of a rolling disk with three unbalance masses that can slide along radial spokes equispaced in angular orientation. The objective is to design trajectories for the masses that satisfy physical constraints and enable the disk to accelerate or move with constant velocity. The disk is designed to remain vertically upright and is constrained to move along a straight line. We design trajectories for constant acceleration, first using a static model, and then through detailed analysis using a dynamic model. The analysis based on the dynamic model considers two separate cases; one where the potential energy of the system is conserved, and the other where it continually varies. Whereas trajectories conserving potential energy are quite similar to those obtained from the static model, the variable potential energy trajectories are the most general. A number of observations related to the system center-of-mass are made with respect to both trajectories. Following the strategy for constant acceleration maneuvers, we give a simple approach to tracking an acceleration profile and provide some simulation results. [DOI: 10.1115/1.1344903]

1 Introduction

Ever since its invention, the wheel has been used primarily as a quasi-static device. Enhanced mobility and stability have been achieved using multiple wheels, large wheels, broad wheel bases, multiwheel drives, etc. Only in the recent past, researchers proposed wheels that are dynamical entities. The Gyrover proposed by Brown and Xu [1], for example, is a dynamically stabilized single-wheel robot that uses gyroscopic forces for steering and stability. The Gyrover, in which the wheel and the vehicle are one and the same, has a number of advantages over multiwheeled vehicles. Before the introduction of Gyrover, a few designs were proposed for spherical wheels with internal propulsion mechanisms. An omnidirectional robot comprised of a spherical wheel, an arch-shaped body, and an arm-like mechanism, was proposed by Koshiyama and Yamafuji [2]. In two different designs proposed by Halme et al. [3] and Bicchi et al. [4], a device constrained to roll inside the spherical cavity creates unbalance and generates motion. A change in heading is produced by turning the wheel axis. Both designs complicate the control problem by imposing nonholonomic constraints, internal and external to the spherical shell.

To simplify the control problem and from practical considerations, we proposed a spherical robot design ([5]), where the propulsion mechanism is fixed to the outer skeleton. The propulsion mechanism is comprised of four unbalance masses that are controlled along radial spokes; the extremities of the spokes define the vertices of a regular tetrahedron. The control of the four masses to achieve a desired motion of the spherical wheel poses a complicated and challenging problem in dynamics and control. To get insight into this problem, we study the planar case in this paper. We investigate the dynamics of a rolling disk with three unbalance masses, constrained to slide along radial spokes, configured 120 deg apart from one another. For this system, shown in Fig. 1, we propose to control the motion of the masses such that the disk center can have a specified constant acceleration. We also investigate optimal transition of the disk acceleration from one value to another. These problems are relevant since a trajectory of the disk can be specified in terms of an acceleration profile.

Besides the mechanism comprised of reciprocating masses, shown in Fig. 1, a number of other mechanisms can be designed for propelling the disk. For example, a heavy mass constrained to roll on the inner perimeter can create mass eccentricity and cause the disk to roll. Planar versions of the designs by Halme et al. [3] and Bicchi et al. [4] conform to this category. Also, a spinning rotor mounted on the disk can generate reaction torque and cause the disk to roll. This mechanism, with an additional rotor that allows control of the disk inclination, has been extensively studied by Ehlers et al. [6] and Yavin [7,8]. In their studies the complete dynamics of the disk and rotors were considered, and the nonlinear control problems of tracking and point-to-point stabilization were addressed. A number of other authors (Getz [9] and Rui and McClamroch [10], for example) have also addressed the stabilization problem in the rolling disk but few have investigated the dynamics of viable propulsion mechanisms.

In this paper, we first describe a propulsion mechanism for a vertically upright rolling disk, constrained to move along a straight line. In Section 3 we present preliminary analysis of the mechanism based on a static model. Two solutions are presented in this section for uniform acceleration of the disk along a straight line. The first solution identifies via-points and interpolates sinusoids to generate approximately constant acceleration trajectories. The trajectories conserve potential energy and suggest further analysis on the basis of potential energy. The second solution is optimal in the sense that it minimizes a component of the kinetic energy. It also renders the moment of inertia of the disk invariant with orientation and provides greater freedom in trajectory design. A dynamic model of the system is developed in Section 4 and uniform acceleration maneuvers that conserve potential energy are once again investigated. The variable potential energy trajectories are studied in Section 5. Similar to the static model, the trajectories conserving potential energy are limacons, and identical for the unbalance masses. The variable potential energy trajectories, which are not limacons, are described by five constants of motion and present the most general solution. In Section 6 we present an optimal approach to tracking an acceleration profile along with simulation results. Section 7 provides concluding remarks.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, June 24, 1999; final revision, April 16, 2000. Associate Editor: N. C. Perkins. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

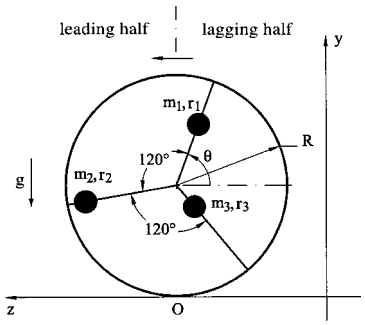


Fig. 1 The disk with reciprocating masses

2 Description of Propulsion Mechanism

A schematic description of the self-propelled disk is shown in Fig. 1. The radial lines from center of the disk to the circumference represent spokes, each of which carries a lumped mass. The masses, denoted by m_1, m_2, m_3, are of equal magnitude, m, and slide along their respective spokes. The angular position of mass m_1 is measured counter clockwise from the negative z-axis, and is denoted by θ. The masses m_2, m_3, are located 120 deg and 240 deg apart with respect to m_1. The radial positions of the masses are denoted by r_1, r_2, and r_3, respectively. For ease of explanation, we divide the disk area into two distinct halves: the ''leading half'' and the ''lagging half.'' The gravitational force of a mass in the leading half contributes positive moment and causes the disk to accelerate; the gravitational force of a mass in the lagging half causes the disk to deccelerate. Now consider a static model of the mechanism, which ignores the inertia forces of the unbalance masses. For this model, which is valid for small velocities and accelerations, we have

$$I(\theta)\ddot{\theta} = -mg[r_1 \cos\theta + r_2 \cos(\theta + 120°) + r_3 \cos(\theta - 120°)] \quad (1)$$

where, $I(\theta)$ is the mass moment of inertia of the entire assembly about the instantaneous center of rotation, O, which can be expressed as

$$I(\theta) \triangleq I_{ds} + 3mR^2 + m(r_1^2 + r_2^2 + r_3^2) + 2mR[r_1 \sin\theta + r_2 \sin(\theta + 120°) + r_3 \sin(\theta - 120°)]. \quad (2)$$

It is implicitly assumed in Eq. (1) that the friction force between the disk and the ground prevents the disk from slipping. In Eq. (2), I_{ds} represents the combined moment of inertia of the disk and spokes about O. The unbalance masses are constrained by the relation $0 \leqslant r_1, r_2, r_3 \leqslant R$, where R is the length of each spoke.

We designed our mechanism with three masses since fewer masses cannot maintain constant acceleration. For a single mass, this is evident from the equation of motion

$$r_1 = -\frac{I_1(\theta)\ddot{\theta}}{mg\cos\theta}, \quad I_1(\theta) \triangleq I_{ds} + m(R^2 + r_1^2 + 2r_1 R \sin\theta)$$

which indicates that constant $\ddot{\theta}$ cannot be achieved with r_1 satisfying $0 \leqslant r_1 \leqslant R$. For two masses, on spokes that are separated by an angle α, the equation of motion has the form

$$I_2(\theta)\ddot{\theta} = -mg[r_1 \cos\theta + r_2 \cos(\theta + \alpha)]$$

$$I_2(\theta) \triangleq I_{ds} + 2mR^2 + m(r_1^2 + r_2^2) + 2mR[r_1 \sin\theta + r_2 \sin(\theta + \alpha)].$$

On simplification, we have

$$r_{eq} = -\frac{I_2(\theta)\ddot{\theta}}{mg\cos(\theta + \psi)} \quad r_{eq} \triangleq \sqrt{r_1^2 + r_2^2 + 2r_1 r_2 \cos\alpha},$$

$$\psi \triangleq \arctan\left(\frac{r_2 \sin\alpha}{r_1 + r_2 \cos\alpha}\right)$$

where $0 \leqslant \psi \leqslant \alpha$. Since $(\theta + \psi)$ will assume all angular positions during motion, r_{eq} cannot remain bounded for a constant acceleration. Therefore, two masses will not suffice. In the next three sections, we will show that three masses equispaced in angular orientation, as shown in Fig. 1, is capable of maintaining constant acceleration.

3 Preliminary Analysis Using Static Model

3.1 An Approximate Solution. In this section we present an approximate solution to the constant acceleration maneuver problem. We use the static model in Eq. (1) but assume the moment of inertia of the system to be constant. We divide the leading half into three phases, shown in Fig. 2. The presence of a mass in phase 1 necessitates the second mass to be present in phase 3 and the third mass in the lagging half. The presence of a mass in phase 2 necessitates the other two masses to be confined to the lagging half. Now consider the configuration in Fig. 3 where mass m_1 is on the boundary between phase 1 and phase 2. We use this configuration to determine the maximum acceleration of the disk that can be maintained for all values of θ. To achieve maximum acceleration in this configuration, we must have $r_1 = R$ and $r_3 = 0$. The instantaneous value of r_2 is not important since m_2 does not have a moment arm. The maximum acceleration in this configuration can be obtained from Eq. (1), as follows:

$$I\ddot{\theta} = -mgR\cos 150°.$$

To maintain this acceleration for all values of θ, the mass positions should satisfy

$$R\cos 150° = r_1 \cos\theta + r_2 \cos(\theta + 120°) + r_3 \cos(\theta - 120°).$$

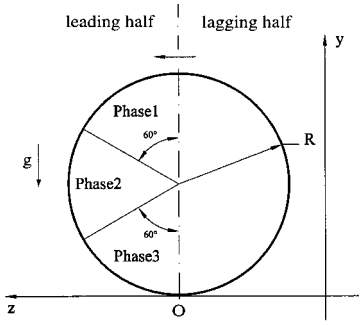


Fig. 2 Different phases in the leading half of the disk

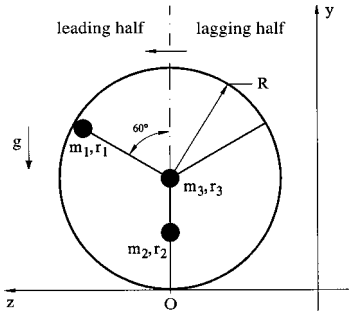


Fig. 3 A particular configuration of the reciprocating masses

In the neighborhood of $\theta=150°$, we have $r_3\approx 0$ and $r_1\approx R$. Hence, r_2 can be obtained as

$$r_2(150°)=\lim_{\theta\to 150°} R\frac{(\cos 150°-\cos\theta)}{\cos(\theta+120°)}=\frac{R}{2}.$$

A similar analysis can be carried out at $\theta=210°$ where m_1 is between phase 2 and phase 3. At this configuration, where $r_1=R$, and $r_2=0$, we can show that r_3 satisfies $r_3=R/2$.

Since the spokes are symmetrically located, we assume the mass trajectories to be identical with 120 deg phase shift from one another. The above analysis then implies that there are six via-points on the trajectory, namely

$$r_1=\begin{cases}0 & \text{for } \theta=30°,-30°\\ R/2 & \text{for } \theta=90°,270°\\ R & \text{for } \theta=150°,210°\end{cases}$$

By fitting sinusoids between these via-points, the approximate solution is obtained as

$$r_1(\theta)=\begin{cases}0 & \text{for } -30°\leqslant\theta\leqslant 30°\\ R[1-\cos(\theta-30°)] & \text{for } 30°\leqslant\theta\leqslant 90°\\ -R\cos(\theta+30°) & \text{for } 90°\leqslant\theta\leqslant 150°\\ R & \text{for } 150°\leqslant\theta\leqslant 210°\\ -R\cos(\theta-30°) & \text{for } 210°\leqslant\theta\leqslant 270°\\ R[1-\cos(\theta+30°)] & \text{for } 270°\leqslant\theta\leqslant 330°.\end{cases} \tag{3}$$

The trajectories of r_2 and r_3 can be simply obtained as

$$r_2(\theta)=r_1(\theta+120°),\quad r_3(\theta)=r_1(\theta-120°). \tag{4}$$

For example, $r_2(\theta)$ can be expressed as

$$r_2(\theta)=\begin{cases}0 & \text{for } 210°\leqslant\theta\leqslant 270°\\ R(1+\sin\theta) & \text{for } 270°\leqslant\theta\leqslant 330°\\ -R\cos(\theta+150°) & \text{for } -30°\leqslant\theta\leqslant 30°\\ R & \text{for } 30°\leqslant\theta\leqslant 90°\\ R\sin\theta & \text{for } 90°\leqslant\theta\leqslant 150°\\ R[1-\cos(\theta+150°)] & \text{for } 150°\leqslant\theta\leqslant 210°.\end{cases}$$

The following observations can now be made for the approximate solution. The trajectories of the masses are piecewise smooth with first derivative continuity. Also, they satisfy

$$r_1\sin\theta+r_2\sin(\theta+120°)+r_3\sin(\theta-120°)=0 \tag{5}$$

which implies conservation of potential energy. The moment of inertia of the system is, however, not constant. From Eq. (2), it can be expressed as

$$I_a=I_{ds}+3mR^2+m(r_1^2+r_2^2+r_3^2).$$

From the above equation it can be shown that I_a is comprised of a constant term and a periodic term. Therefore, the angular acceleration of the disk

$$\ddot{\theta}=-\frac{mgR\cos 150°}{I_a}\approx 0.866\frac{mgR}{I_a} \tag{6}$$

is not constant, but varies periodically. To complete the analysis, we note that a lower magnitude of acceleration can be generated by scaling down the trajectories of r_1, r_2, and r_3. This can be done by replacing R in Eq. (3) with R^*, $0\leqslant R^*\leqslant R$.

3.2 An Optimal Solution. In this section, we seek an optimal solution; one optimal in some sense of energy consumption. We impose the constraint that the potential energy is conserved and therefore Eq. (5) holds. This condition, which was satisfied by the approximate solution, is convenient and reduces the number of variable terms in the moment of inertia expression in Eq. (2).

By differentiating Eq. (5) with respect to θ and substituting Eq. (1), we obtain

$$r_1'\sin\theta+r_2'\sin(\theta+120°)+r_3'\sin(\theta-120°)=K,\quad K\triangleq\frac{I\ddot{\theta}}{mg} \tag{7}$$

where r_1', r_2', and r_3', are the derivatives of r_1, r_2, and r_3, respectively, with respect to θ, I represents the moment of inertia of the system given by Eq. (2), and $\ddot{\theta}$ denotes the specified constant acceleration of the disk. We now make the assumption that I is constant. In the ensuing analysis, we will show that I can be maintained at a constant value. By differentiating Eq. (1) with respect to θ and substituting Eq. (5), we get

$$r_1'\cos\theta+r_2'\cos(\theta+120°)+r_3'\cos(\theta-120°)=0. \tag{8}$$

Using Eqs. (7) and (8), we can express r_2' and r_3' in terms of r_1' as follows:

$$\begin{aligned}r_2'&=r_1'-\frac{2}{\sqrt{3}}K\cos(\theta-120°)\\ r_3'&=r_1'+\frac{2}{\sqrt{3}}K\cos(\theta+120°)\end{aligned} \tag{9}$$

where K was defined in Eq. (7). With the objective of designing the optimal trajectory, we now define the cost functional

$$J=\int_0^{2\pi}L\,d\theta,\quad L\triangleq\frac{1}{2}m(r_1'^2+r_2'^2+r_3'^2) \tag{10}$$

where the integrand represents the pseudo-kinetic energy, or the kinetic energy that accounts for the motion of the masses in the reference frame of the spokes. With respect to an inertial frame, the masses will undergo both translation and rotation that will depend on the instantaneous angular velocity of the disk. During constant acceleration maneuvers, the angular velocity of the disk will increase linearly with time and the true kinetic energy will be a function of time. To investigate optimal trajectories that are functions of θ, rather than both θ and time, we choose the cost functional as the integral of the pseudo-kinetic energy.

To proceed with the optimization, we substitute the expressions for r_2' and r_3' into Eq. (9) to rewrite the integrand in Eq. (10) as follows:

$$L=\frac{1}{2}m\left(3r_1'^2+\frac{4}{3}K^2[\cos^2(\theta-120°)+\cos^2(\theta+120°)]-4Kr_1'\sin\theta\right).$$

Using the Euler-Lagrange equation ([11]) from calculus of variations

$$\frac{d}{d\theta}\left(\frac{\partial L}{\partial r_1'}\right)-\frac{\partial L}{\partial r_1}=0$$

for the stationary value of J, we obtain the trajectory of r_1 as follows:

$$r_1=C_1\theta-\frac{2}{3}K\cos\theta+C_2$$

where C_1 and C_2 are constants of integration. Since r_1 has to satisfy $0\leqslant r_1\leqslant R$, C_1 must be zero, and C_2 and K must lie in the shaded region of Fig. 4. The optimal trajectory will therefore be a limacon, of the form

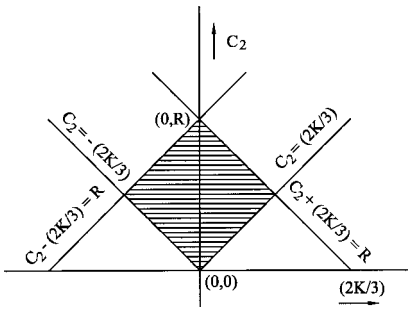


Fig. 4 Shaded region indicates feasible parameter values for the solution in Section 3.2

$$r_1 = C_2 - \frac{2}{3}K\cos\theta. \tag{11}$$

The optimal trajectories of r_2, r_3, can be derived using Eq. (4). The moment of inertia, which was assumed constant, can now be shown to be constant. From Eq. (2) we can show that

$$I_o = I_{ds} + m\left(3R^2 + 3{C_2}^2 + \frac{2}{3}K^2\right).$$

The constant acceleration of the disk can therefore be expressed as

$$\ddot{\theta} = \frac{mgK}{I_o}$$

where $-0.75R \leq K \leq 0.75R$, limits the maximum value of acceleration to $\ddot{\theta} = 0.75mgR/I_o$. The feasible range of values of K can be verified from Fig. 4.

Unlike the approximate solution, the optimal solution results in constant acceleration of the disk. The optimal solution is also smooth whereas the approximate solution is piecewise smooth with first derivative continuity. A single parameter R^* describes the family of approximate solutions. The particular solution where $R^* = R$ is shown in Eq. (3). Two parameters, C_2 and K, describe the family of optimal solutions. Clearly, the optimal solution provides greater freedom in trajectory selection. Despite differences, the trajectories for the approximate solution and the optimal solution, shown in Fig. 5, are strikingly similar.

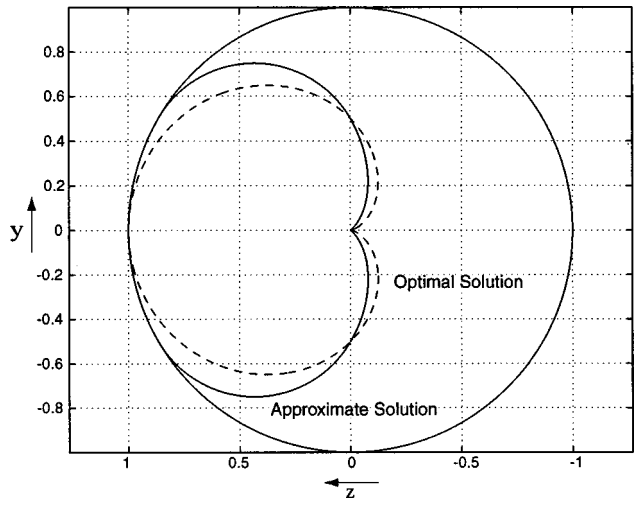


Fig. 5 Comparison of the approximate and optimal solutions for a disk of unity radius

4 Dynamic Model: Constant Potential Energy Maneuvers

4.1 Lagrangian Formulation. In this section we perform a detailed analysis of the problem using a dynamic model. To obtain Lagrange's equations, we first express the kinetic energy of the system as

$$T = \frac{1}{2}I_{ds}\dot{\theta}^2 + \frac{1}{2}m(v_1^2 + v_2^2 + v_3^2)$$

where v_1, v_2, and v_3, are the velocities of masses m_1, m_2, and m_3, respectively, and given by the relations

$$v_1^2 = R^2\dot{\theta}^2 + \dot{r}_1^2 + {r_1}^2\dot{\theta}^2 - 2R\dot{\theta}\dot{r}_1\cos\theta + 2Rr_1\dot{\theta}^2\sin\theta$$

$$v_2^2 = R^2\dot{\theta}^2 + \dot{r}_2^2 + {r_2}^2\dot{\theta}^2 - 2R\dot{\theta}\dot{r}_2\cos(\theta+120°) + 2Rr_2\dot{\theta}^2\sin(\theta+120°)$$

$$v_3^2 = R^2\dot{\theta}^2 + \dot{r}_3^2 + {r_3}^2\dot{\theta}^2 - 2R\dot{\theta}\dot{r}_3\cos(\theta-120°) + 2Rr_3\dot{\theta}^2\sin(\theta-120°).$$

In the kinetic energy expression above, it is implicitly assumed that the friction force between the disk and the ground prevents the disk from slipping. The potential energy of the system is expressed as

$$V = mg[r_1\sin\theta + r_2\sin(\theta+120°) + r_3\sin(\theta-120°)]. \tag{12}$$

Using the expressions for kinetic and potential energies, Lagrange's equation ([12]) for the generalized coordinate θ can be written as

$$\begin{aligned} &I_{ds}\ddot{\theta} + 3mR^2\ddot{\theta} + m\ddot{\theta}(r_1^2 + r_2^2 + r_3^2) + 2m\dot{\theta}(r_1\dot{r}_1 + r_2\dot{r}_2 + r_3\dot{r}_3) \\ &\quad - mR(\ddot{r}_1\cos\theta + \ddot{r}_2\cos(\theta+120°) + \ddot{r}_3\cos(\theta-120°)) \\ &\quad + 2mR\ddot{\theta}(r_1\sin\theta + r_2\sin(\theta+120°) \\ &\quad + r_3\sin(\theta-120°)) \\ &\quad + 2mR\dot{\theta}(\dot{r}_1\sin\theta + \dot{r}_2\sin(\theta+120°) \\ &\quad + \dot{r}_3\sin(\theta-120°)) + mR\dot{\theta}^2(r_1\cos\theta \\ &\quad + r_2\cos(\theta+120°) + r_3\cos(\theta-120°)) + mg(r_1\cos\theta \\ &\quad + r_2\cos(\theta+120°) + r_3\cos(\theta-120°)) = 0. \end{aligned} \tag{13}$$

The approximate and optimal solutions in Section 3 indicate that constant angular acceleration of the disk can be generated by periodic trajectories of the unbalance masses. This motivates us to seek periodic solutions from the dynamic analysis as well. We assume r_1, r_2, and r_3 to be periodic functions of the form

$$r_1 = Rf_1(\theta), \quad r_2 = Rf_2(\theta), \quad r_3 = Rf_3(\theta),$$
$$0 \leq f_1(\theta), f_2(\theta), f_3(\theta) \leq 1 \tag{14}$$

where f_1, f_2, f_3 are dimensionless variables. In the sequel we will establish that there exists a class of periodic trajectories for r_1, r_2, and r_3, that impart constant angular acceleration to the disk. From an implementation point of view, radial forces provided by suitable actuators will guarantee that the unbalance masses track their periodic trajectories. Since our main objective is to investigate the effect of the periodic trajectories on the overall motion of the disk, we do not pursue further analysis of the radial forces. One can easily determine these radial forces or control inputs from the right-hand sides of Lagrange's equations for the generalized coordinates r_1, r_2, and r_3. To continue with our analysis, we use Eq. (14) to rewrite Eq. (13) in the form

$$a(\theta)\ddot{\theta} + b(\theta)\dot{\theta}^2 + c(\theta) = 0$$

$$a(\theta)=I_{ds}+mR^2[3+f_1^2+f_2^2+f_3^2+2(f_1\sin\theta+f_2\sin(\theta+120°)+f_3\sin(\theta-120°))-(f_1'\cos\theta+f_2'\cos(\theta+120°)+f_3'\cos(\theta-120°))]$$

$$b(\theta)=mR^2[2(f_1f_1'+f_2f_2'+f_3f_3')-(f_1''\cos\theta+f_2''\cos(\theta+120°+f_3''\cos(\theta-120°))+2(f_1'\sin\theta+f_2'\sin(\theta+120°)+f_3'\sin(\theta-120°))+(f_1\cos\theta+f_2\cos(\theta+120°)+f_3\cos(\theta-120°))]$$

$$c(\theta)=mgR(f_1\cos\theta+f_2\cos(\theta+120°)+f_3\cos(\theta-120°)) \tag{15}$$

For a constant acceleration maneuver, $\ddot{\theta}=\lambda_1$, we will have $\dot{\theta}=\lambda_1 t+\lambda_2$, where λ_1 and λ_2 are constants. We can then argue that $b(\theta)=0$, or else $\ddot{\theta}$ will increase with time according to the relation

$$\ddot{\theta}=-\frac{b}{a}(\lambda_1 t+\lambda_2)^2-\frac{c}{a}. \tag{16}$$

From Eqs. (15) one can readily show that $b(\theta)=0$ implies

$$(f_1^2+f_2^2+f_3^2)+(f_1\sin\theta+f_2\sin(\theta+120°)+f_3\sin(\theta-120°))-(f_1'\cos\theta+f_2'\cos(\theta+120°)+f_3'\cos(\theta-120°))=\alpha_0 \tag{17}$$

where α_0 is a constant of integration. Apart from $b(\theta)=0$, we also need (c/a) to be constant, or both c and a to be constant, for a constant acceleration maneuver.

4.2 Uniform Acceleration With Constant Potential Energy

Consider the case where both c and a are constant. Rewriting $a(\theta)$ in Eq. (15) as

$$a(\theta)=I_{ds}+mR^2[3+\alpha_0+f_1\sin\theta+f_2\sin(\theta+120°)+f_3\sin(\theta-120°)] \tag{18}$$

and from the expression of $c(\theta)$ in Eq. (15), we get the identities

$$f_1\cos\theta+f_2\cos(\theta+120°)+f_3\cos(\theta-120°)=\alpha_1 \tag{19}$$

$$f_1\sin\theta+f_2\sin(\theta+120°)+f_3\sin(\theta-120°)=\alpha_2 \tag{20}$$

where α_1 and α_2 are constants. From Eqs. (12), (14), and (20) it is established that constant values of c and a lead to constant acceleration maneuvers with potential energy conservation. To proceed further, we differentiate Eq. (19) and substitute Eq. (20) to get

$$f_1'\cos\theta+f_2'\cos(\theta+120°)+f_3'\cos(\theta-120°)=\alpha_2.$$

Substituting this result and Eq. (20) into Eq. (17), we get

$$(f_1^2+f_2^2+f_3^2)=\alpha_0. \tag{21}$$

Using Eqs. (19), (20), and (21), the following expressions for the dimensionless variables f_1, f_2, and f_3 can be obtained:

$$f_1=X\cos(\theta-\phi)+Y \quad X\triangleq(2/3)\{\alpha_1^2+\alpha_2^2\}^{1/2}$$

$$f_2=X\cos(\theta-\phi+120°)+Y \quad Y\triangleq(1/3)\{3\alpha_0-2(\alpha_1^2+\alpha_2^2)\}^{1/2} \tag{22}$$

$$f_3=X\cos(\theta-\phi-120°)+Y \quad \phi\triangleq\arctan(\alpha_2/\alpha_1)$$

where X, Y, ϕ, are constants. From the expressions of X and Y it is clear that both were chosen to be positive square roots. This does not cause any loss of generality and is explained as follows. Irrespective of the sign of X, $X\cos(\theta-\phi)$ takes both positive and negative values. To satisfy the constraint $f_1\geqslant 0$ in Eq. (14), Y must therefore be positive. A change in sign of X is equivalent to a phase shift in ϕ by 180 deg. Therefore X is arbitrarily chosen to be positive. The above analysis confirms that f_1,f_2,f_3, and accordingly r_1,r_2,r_3, have identical trajectories, shifted in phase.

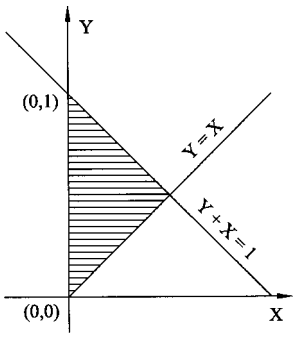


Fig. 6 Shaded region indicates feasible parameter values for the solution in Section 4.2

The trajectories in Eq. (22) are limacons, similar to those obtained in Section 3.2. This is surprising since the analysis in Section 3.2 was carried out with a static model whereas the complete dynamic model was employed in this section. The main difference between the two trajectories are in the number of defining parameters. The limacons in Section 3.2 are defined by two parameters, C_2 and K, as seen from Eq. (11). In this section the limacons are defined by three parameters, namely X, Y, and ϕ. The advantage of having three independent parameters is that any initial configuration of the three masses can determine the trajectories and the corresponding acceleration. From Eqs. (15), (16), (19), and (20), the disk acceleration can be expressed as

$$\ddot{\theta}=-\left[\frac{mgR\alpha_1}{I_{ds}+mR^2(3+\alpha_0+\alpha_2)}\right] \tag{23}$$

where $\alpha_0,\alpha_1,\alpha_2$ can be expressed in terms of trajectory parameters X, Y, ϕ as follows:

$$\alpha_0\triangleq 1.5X^2+3.0Y^2, \quad \alpha_1\triangleq 1.5X\cos\phi, \quad \alpha_2\triangleq 1.5X\sin\phi. \tag{24}$$

While arbitrary initial conditions can uniquely define a trajectory, not all trajectories will satisfy the physical constraints of Eq. (14). For feasibility, X,Y must lie in the shaded region, shown in Fig. 6. This is quite similar to the constraint imposed on the parameters of the trajectory in Section 3.2, shown in Fig. 4.

Although initial values of f_1, f_2, f_3, and θ uniquely define the trajectory parameterized by $\alpha_0,\alpha_1,\alpha_2$, or X,Y,ϕ, and uniquely define the acceleration of the disk $\ddot{\theta}$, the converse is not true: A given acceleration of the disk can be generated through various trajectories.

4.3 Effect of Variation of Path Parameters

Effect of Varying Phase Angle. Using Eqs. (23) and (24), the disk acceleration can be written as

$$\ddot{\theta}=-\left[\frac{1.5mgRX\cos\phi}{I_{ds}+1.5mR^2(2+X\sin\phi+X^2+2Y^2)}\right]=\frac{\beta\cos\phi}{\mu+\sin\phi} \tag{25}$$

where β and μ are constants, given by the relations

$$\beta\triangleq-\frac{g}{R}, \quad \mu\triangleq\frac{I_{ds}+1.5mR^2(2+X^2+2Y^2)}{1.5mR^2X}. \tag{26}$$

From the expression of μ, and feasible values of X in Fig. 6, it can be shown that the denominator in Eq. (25) is always positive. The

sign of $\ddot{\theta}$ therefore depends on the sign of $\cos\phi$, or $X\cos\phi$ since X is always positive. Specifically, $\ddot{\theta}$ is positive if $X\cos\phi$ is negative and vice versa. With Eqs. (19) and (24) it can be established that the z-coordinate of the center of mass, relative to the center of the disk, is located at

$$z_{cm}=\frac{R}{3}[f_1\cos(\pi+\theta)+f_2\cos(\pi+\theta+120°)+f_3\cos(\pi+\theta-120°)]=-\frac{R}{3}\alpha_1=-\frac{R}{2}X\cos\phi.$$

The above equation implies that the magnitude and direction of acceleration depends primarily on the z-coordinate of the center of mass, relative to the center of disk.

The rolling disk can acquire a range of acceleration by varying phase angle ϕ. From Eq. (25), the maximum and minimum values of the acceleration can be shown to be

$$\ddot{\theta}=\begin{cases}\ddot{\theta}_{\max} & \text{if } \phi=\pi+\arcsin(1/\mu)\\ \ddot{\theta}_{\min} & \text{if } \phi=-\arcsin(1/\mu).\end{cases} \tag{27}$$

Of course, motion with zero acceleration or constant velocity requires $\phi=\pm\pi/2$. From an implementation point of view, the rolling disk can change ϕ during its motion by suppressing the motion of the masses for an appropriate interval of time.

Effect of Varying Parameters ***X*** *and* ***Y****.* Each point in the shaded region of Fig. 6 corresponds to a certain trajectory of the disk; each of these trajectories has a certain acceleration. Clearly, a variation in X and Y is expected to provide a range of accelerations. By treating ϕ as constant, we partially differentiate $\ddot{\theta}$ in Eq. (25) with respect to X and Y. Equating these expressions to zero, the maximum $\ddot{\theta}$ is observed to occur at the following coordinate:

$$X=\sqrt{2(1+I_{ds}/3mR^2)},\quad Y=0$$

independent of the value of ϕ. Unfortunately, this coordinate lies outside the shaded region in Fig. 6. Since the above coordinate is the only location where $\ddot{\theta}$ is an extremum, we conclude that the maximum feasible $\ddot{\theta}$ occurs at a point, or a set of points on the boundary of the shaded region. Through numerical simulation we determined the maximum to occur at $(X,Y)=(0.5,0.5)$. Combining this result with the result in Eq. (27), we conclude that the trajectories that produce maximum and minimum acceleration are

$$\ddot{\theta}=\begin{cases}\ddot{\theta}_{\max} & \text{if } f_1=-0.5\cos[\theta-\arcsin(1/\bar{\mu})]+0.5\\ \ddot{\theta}_{\min} & \text{if } f_1=0.5\cos[\theta+\arcsin(1/\bar{\mu})]+0.5\end{cases},\qquad \bar{\mu}\triangleq\frac{I_{ds}+4.125mR^2}{0.75mR^2}$$

where $\bar{\mu}$, obtained from Eq. (26), is the value of μ evaluated at $X=Y=0.5$. Instead of numerical simulation, the above result for maximum acceleration can also be obtained through constrained optimization.

5 Dynamic Model: Variable Potential Energy Maneuvers

5.1 Uniform Acceleration With Variable Potential Energy It was shown in Section 4.1 that constant acceleration maneuvers require (c/a) to be constant. The analysis in Sections 4.2 and 4.3 was carried out assuming both c and a as constants, which leads to conservation of potential energy. In an effort to generalize the results, we investigate the case of varying potential energy in this section. We treat c and a as variables whose ratio is constant. We begin our analysis with the expression for the potential energy. Using Eqs. (12) and (14), V can be expressed as

$$V=mgR[f_1\sin\theta+f_2\sin(\theta+120°)+f_3\sin(\theta-120°)].$$

It can be shown that V satisfies $-mgR\leqslant V\leqslant mgR$. In compliance with these limits and without loss of generality, we consider a sinusoidal variation in V, given by the relation

$$V=mgR[A_1+B_1\sin(\theta-\psi)]$$

where A_1, B_1, and ψ are constants. By comparing the above two equations, we can write

$$f_1\sin\theta+f_2\sin(\theta+120°)+f_3\sin(\theta-120°)=A_1+B_1\sin(\theta-\psi). \tag{28}$$

We have $b(\theta)=0$ for constant acceleration maneuvers. Using the expressions for $c(\theta)$ and $a(\theta)$ in Eqs. (15) and (18), the equation of motion in Eq. (16) reduces to the form

$$\ddot{\theta}\{I_{ds}+mR^2[3+\alpha_0+f_1\sin\theta+f_2\sin(\theta+120°)+f_3\sin(\theta-120°)]\}=-mgR[f_1\cos\theta+f_2\cos(\theta+120°)+f_3\cos(\theta-120°)]. \tag{29}$$

Substituting Eq. (28) in Eq. (29), we get

$$\ddot{\theta}\{K_1+K_2[A_1+B_1\sin(\theta-\psi)]\}=f_1\cos\theta+f_2\cos(\theta+120°)+f_3\cos(\theta-120°)$$

$$K_1\triangleq-\frac{I_{ds}+mR^2(3+\alpha_0)}{mgR},\quad K_2\triangleq-\frac{R}{g}. \tag{30}$$

For a constant magnitude of acceleration, $\ddot{\theta}=G$, we can then write

$$f_1\cos\theta+f_2\cos(\theta+120°)+f_3\cos(\theta-120°)=A_2+B_2\sin(\theta-\psi)$$

$$A_2=G(K_1+A_1K_2),\quad B_2=GB_1K_2. \tag{31}$$

Using Eqs. (28) and (31), Eq. (17) can now be written as

$$f_1^2+f_2^2+f_3^2=\alpha_0+B_2\cos(\theta-\psi) \tag{32}$$

where it is obvious that $\alpha_0\geqslant|B_2|$. From Eqs. (28), (31), and (32), f_1, f_2, and f_3 can be solved as follows:

$$f_1=\frac{2}{3}P_1\pm\frac{1}{3}\sqrt{3(\alpha_0+B_2\cos(\theta-\psi))-2S}$$

$$f_2=\frac{2}{3}P_2\pm\frac{1}{3}\sqrt{3(\alpha_0+B_2\cos(\theta-\psi))-2S}$$

$$f_3=\frac{2}{3}P_3\pm\frac{1}{3}\sqrt{3(\alpha_0+B_2\cos(\theta-\psi))-2S}$$

where S and P_j, $j=1,2,3$, are defined as

$$S\triangleq[A_1+B_1\sin(\theta-\psi)]^2+[A_2+B_2\sin(\theta-\psi)]^2$$

$$P_j\triangleq[A_1+B_1\sin(\theta-\psi)]\sin[\theta+(j-1)120°]+[A_2+B_2\sin(\theta-\psi)]\cos[\theta+(j-1)120°].$$

It can be verified from the above equation that $(P_1+P_2+P_3)=0$. Using this result the common term in the expressions of f_1, f_2, f_3 can be shown to be $(f_1+f_2+f_3)/3$. It immediately follows that the nontrivial solution for f_1, f_2, f_3 is

$$f_1=\frac{2}{3}P_1+\frac{1}{3}\sqrt{3(\alpha_0+B_2\cos(\theta-\psi))-2S}$$

$$f_2=\frac{2}{3}P_2+\frac{1}{3}\sqrt{3(\alpha_0+B_2\cos(\theta-\psi))-2S} \tag{33}$$

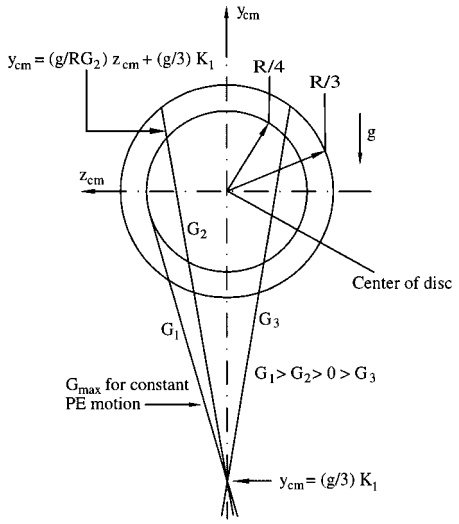


Fig. 7 A geometric interpretation of the motion of the center of mass

straints, $0 \leq f_1, f_2, f_3 \leq 1$, will be met. This is true for the same reason the center-of-mass may remain bounded when individual mass positions become unbounded.

3. Since α_0 is positive, which can be shown from Eq. (32), the intercept of the straight line in Eq. (34) on the y-axis is negative, and outside the circle defined by Eq. (35). This confirms that the range of acceleration of the disk, determined by the slope of the line, is finite.

4. A feasible trajectory is described by sinusoidal variation in both center-of-mass coordinates, y_{cm} and z_{cm}. This, evident from Eqs. (28) and (31), translates to the center-of-mass position oscillating along a straight line, while remaining confined to the circle of radius $R/3$ in Fig. 7. In contrast, when the potential energy is conserved, y_{cm} remains stationary. This implies that the center-of-mass position will remain stationary.

5. From the range of X in Fig. 6, namely $X \leq 0.5R$, and expressions for α_1, α_2, in Eq. (24), we have $-0.75 \leq \alpha_1, \alpha_2 \leq 0.75$. Using Eqs. (19) and (20) we can therefore show that y_{cm}, z_{cm} are individually constrained to lie between $\pm R/4$. This indicates that for constant potential energy maneuvers the straight line in Eq. (34) must intersect, or be tangential to, a smaller circle of radius $R/4$. This indicates that the variable potential energy case holds the promise for higher acceleration.

$$f_3 = \frac{2}{3} P_3 + \frac{1}{3} \sqrt{3(\alpha_0 + B_2 \cos(\theta - \psi)) - 2S}.$$

From Eqs. (22) and (33) it can be seen that trajectories conserving potential energy are identical for the three masses except for the 120 deg phase shift; this is not true for variable potential energy. Also, the trajectories in Eq. (22) are limacons, described by three parameters. The trajectories in Eq. (33), which are not limacons, require five parameters for their description. These parameters are A_1, B_1, G, ψ, and α_0.

5.2 Motion of the Center of Mass. The coordinates of the center-of-mass of the system can be written as

$$z_{cm} = -\frac{R}{3}[f_1 \cos\theta + f_2 \cos(\theta + 120°) + f_3 \cos(\theta - 120°)]$$

$$y_{cm} = \frac{R}{3}[f_1 \sin\theta + f_2 \sin(\theta + 120°) + f_3 \sin(\theta - 120°)].$$

Using these relations, Eq. (29) can be written as

$$\ddot{\theta}\{I_{ds} + mR^2(3 + \alpha_0) + 3mRy_{cm}\} = 3mgz_{cm}.$$

For a constant acceleration $\ddot{\theta} = G$, this reduces to

$$y_{cm} = \frac{g}{RG} z_{cm} + \frac{g}{3} K_1 \tag{34}$$

where K_1 was defined in Eq. (30). Also note that the distance of the center-of-mass from the disk center is constrained by the relation

$$y_{cm}^2 + z_{cm}^2 \leq (R/3)^2. \tag{35}$$

The proof of the above relation is simple and left to the reader. Some observations on the motion of the disk, evident from Eqs. (34) and (35), are now discussed with the help of Fig. 7.

1. Straight lines with different slopes represent different magnitudes (G) of acceleration. A line with a positive slope represents acceleration and a line with a negative slope represents deceleration. A horizontal line in Fig. 7 represents constant velocity motion.

2. A feasible trajectory of the disk requires the straight line in Eq. (34) to pass through the circular region defined by Eq. (35). However, this is not sufficient to guarantee that the physical con-

6 Tracking an Acceleration Profile

6.1 An Optimal Approach to Tracking Acceleration. In this section we present an optimal method for tracking an acceleration profile. We compute discrete changes in acceleration over small intervals of time and seek to determine changes in trajectory parameters that minimize the cost functional

$$J = \int_0^{2\pi} (\Delta f_1)^2 \, d\theta.$$

The basic trajectories can be chosen to be the limacons in Eq. (22) or the more complex forms in Eq. (33). For simplicity, we choose the limacons which are described by fewer parameters. Though the limacons conserve potential energy, we do not expect the potential energy to remain conserved as we track an acceleration profile. This is true since the limacon parameters will continually change during acceleration tracking. While a different cost function could have been chosen, the cost function above promises to minimize the overall change in the shape of the trajectory, which is identical for all three masses. On differentiating f_1 in Eq. (22), we obtain

$$\Delta f_1 = \cos(\theta - \phi)\Delta X + X \sin(\theta - \phi)\Delta\phi + \Delta Y. \tag{36}$$

Substituting Eq. (36) into the expression of J, we get

$$J = \pi \Delta X^2 + \pi (X \Delta\phi)^2 + 2\pi \Delta Y^2.$$

We now rewrite Eq. (23) as

$$\ddot{\theta} = \frac{\beta\alpha_1}{\nu + \alpha_2 + \alpha_0}, \quad \nu \triangleq 3 + \frac{I_{ds}}{mR^2}. \tag{37}$$

A change in $\ddot{\theta}$ can therefore be expressed in terms of changes in path parameters, as follows:

$$\Delta\ddot{\theta} = \left(\frac{\ddot{\theta}}{\alpha_1}\right)\Delta\alpha_1 - \left(\frac{\ddot{\theta}^2}{\beta\alpha_1}\right)\Delta\alpha_2 - \left(\frac{\ddot{\theta}^2}{\beta\alpha_1}\right)\Delta\alpha_0. \tag{38}$$

Substituting Eqs. (24) and (38) into the expression of J, we get

$$J = \frac{4\pi}{9}(\Delta\alpha_1{}^2 + \Delta\alpha_2{}^2) + \frac{\pi}{2}\frac{\{(3\beta\ddot{\theta} - 4\alpha_1\ddot{\theta}^2)\Delta\alpha_1 - (3 + 4\alpha_2)\ddot{\theta}^2\Delta\alpha_2 - 3\beta\Delta\ddot{\theta}\alpha_1\}^2}{[3\alpha_0 - 2(\alpha_1^2 + \alpha_2^2)]}.$$

By computing the partial derivatives of J with respect to α_1, α_2, and equating them to zero, we obtain optimal change in path parameters $\Delta\alpha_1, \Delta\alpha_2$,

$$\Delta\alpha_1 = \frac{9\gamma_1\gamma_2\gamma_4}{8\pi + 9\gamma_1(\gamma_2^2+\gamma_3^2)}, \quad \Delta\alpha_2 = -\frac{9\gamma_1\gamma_3\gamma_4}{8\pi + 9\gamma_1(\gamma_2^2+\gamma_3^2)}, \tag{39}$$

where γ_1, γ_2, γ_3, and γ_4 are defined as

$$\gamma_1 = \frac{\pi}{3\alpha_0 - 2(\alpha_1^2+\alpha_2^2)} \quad \gamma_3 = \left(1 + \frac{4}{3}\alpha_2\right)$$

$$\gamma_2 = \left(\frac{\beta}{\ddot{\theta}} - \frac{4}{3}\alpha_1\right) \quad \gamma_4 = \frac{\beta\Delta\ddot{\theta}\alpha_1}{\ddot{\theta}^2}.$$

Though it may seem that $\Delta\alpha_1$ grows unbounded when $\ddot{\theta}$ is zero, this is not the case. By integrating the expression for $\Delta\alpha_1$, we can show that the following relation is true:

$$\frac{1}{\alpha_1} = \sigma\left[\frac{\beta^2}{\ddot{\theta}^2} + \left(1 - \frac{8}{3}\nu\right)\right]^{1/2} - \frac{4\beta}{(3-8\nu)\ddot{\theta}}$$

where σ is a constant of integration, and β and ν are constants that have been defined in Eqs. (26) and (37), respectively. Clearly, as $\ddot{\theta}$ tends to zero, α_1 tends to zero and $\Delta\alpha_1$ remains finite.

The optimal change in parameter α_0, namely $\Delta\alpha_0$, can now be computed from Eqs. (38) and (39). The changes in parameters α_1, α_2, α_0, can be translated into equivalent changes in X, Y, ϕ, using the following equation:

$$\begin{pmatrix}\Delta X\\ \Delta Y\\ \Delta\phi\end{pmatrix} = \frac{2}{3}\begin{pmatrix}\cos\phi & 0 & X\sin\phi\\ \sin\phi & 0 & X\cos\phi\\ 2X & 4Y & 0\end{pmatrix}^{-1}\begin{pmatrix}\Delta\alpha_1\\ \Delta\alpha_2\\ \Delta\alpha_0\end{pmatrix} \tag{40}$$

which was derived from Eq. (22). The matrix in Eq. (40) becomes singular when either X or Y is zero. This should not be of concern since $X=0$ and $Y=0$ are limiting values for a feasible trajectory, as evident from Fig. 6. The matrix is also singular when $X=1$ and $\phi=\pi/4$. This point clearly lies outside the range of feasible parameter values and should also be of no concern. After computing the changes in X, Y, ϕ, the change in the trajectory of m_1 can be computed from Eq. (36). The changes in trajectories of m_2 and m_3 can be computed similarly.

6.2 Simulation Results. In this section we present simulation results of the disk tracking a sinusoidal acceleration profile. The initial angle of the disk in radians, and position of the three masses in dimensionless variables are given as

$$\theta(0)=0.8, \quad f_1[\theta(0)]=0.7, \quad f_2[\theta(0)]=0.4, \quad f_3[\theta(0)]=0.3.$$

The values of $\alpha_0, \alpha_1, \alpha_2$ are first computed using Eqs. (19), (20), and (21). Subsequently, the initial acceleration of the disk is obtained using Eq. (23) as $\ddot{\theta}(0) = -3.28\ \text{rad/s}^2$. For our simulation, we choose the acceleration profile

$$\ddot{\theta} = -2.0 + 8.5\sin(0.5t - \zeta) \tag{41}$$

with the proper choice of ζ that satisfies the initial condition $\ddot{\theta}(0) = -3.28\ \text{rad/s}^2$. We also choose $\beta \triangleq -(g/R) = -100$ in SI units, and $\nu \triangleq 3 + (I_{ds}/mR^2) = 4.5$. The simulation is carried out over 60 seconds using a time-step of 0.001 second. At each time-step the trajectory parameters X, Y, are verified to lie in the shaded region of Fig. 6. This guarantees that physical constraints are not violated.

The simulation results are shown in Figs. 8 and 9. Figure 8 is a plot of trajectory parameters X, Y, and ϕ for the acceleration profile in Eq. (41). As expected, the trajectory of ϕ has the same frequency as that of the acceleration profile. The trajectory of X is more interesting and has two peaks in every cycle of acceleration.

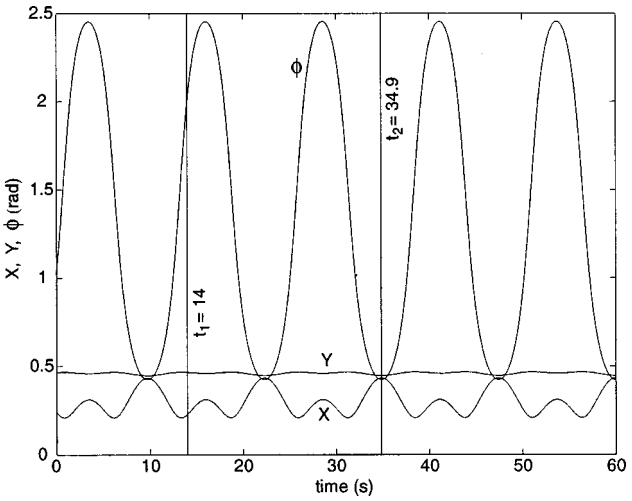


Fig. 8 Variation of trajectory parameters during a sinusoidal variation of acceleration

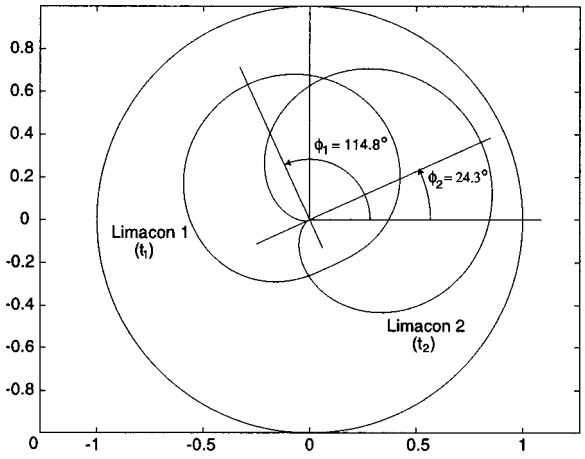


Fig. 9 Normalized trajectories of the unbalance masses at two specific instants of time

Since X represents the radial distance of the center-of-mass, it assumes maximal values for both the maximum and minimum acceleration. However, since magnitudes of the maximum and minimum acceleration are different (6.5 rad/s^2 and -10.5 rad/s^2, respectively), the peaks differ in magnitude. The shape of the limacon in Eq. (22), which changes as a function of the trajectory parameters, is shown in Fig. 9 for two specific instants of time, $t_1 = 14.0$ seconds and $t_2 = 34.9$ seconds. At these instants of time, the limacon is seen to have the following parametric representations:

$$f_1(\theta) = \begin{cases} 0.229\cos(\theta - 114.8°) + 0.467 & \text{for } t = t_1 \\ 0.435\cos(\theta - 24.3°) + 0.448 & \text{for } t = t_2. \end{cases}$$

7 Conclusion

This paper investigates a self-propulsion mechanism comprised of three unbalance masses for a vertically upright rolling disk constrained to move along a straight line path. It is shown that trajectories of the unbalance masses can be designed to propel the disk with a wide range of accelerations. It is also shown that the disk can track an acceleration profile while minimizing an appropriate cost function. In the preliminary analysis, where a static model was used, uniform acceleration maneuvers assumed conser-

vation of potential energy. The potential energy plays an important role throughout the analysis, and hence both constant and variable potential energy cases were considered with the dynamic model. Under conservation of potential energy, the dynamic model results in trajectories similar to those obtained from the static model; the results obtained from the dynamic model are, however, more general. With both models, the trajectories of the unbalance masses are identical limacons with phase shifts of 120 deg. This is not true for the case of variable potential energy. In this most general case, the center-of-mass oscillates along a straight line while the disk undergoes constant acceleration. For all the cases considered, it was observed that a prescribed acceleration could be achieved using multiple trajectories or multiple initial conditions; a set of initial conditions, however, uniquely defines the trajectories and the acceleration. Some simulation results were also presented in this paper. For a prescribed sinusoidal acceleration profile, minimum variation in trajectory parameters were simulated.

Acknowledgment

The authors gratefully acknowledge the support provided by National Science Foundation, NSF Grant No. CMS-9800343, in carrying out this research.

References

[1] Brown, H. B., and Xu, Y., 1997, ''A Single-Wheel Gyroscopically Stabilized Robot,'' IEEE Robotics and Automation Magazine, **4**, No. 3, pp. 39–44.
[2] Koshiyama, A., and Yamafuji, K., 1993, ''Design and Control of All-Direction Steering Type Mobile Robot,'' Int. J. Robot. Res., **12**, No. 5, pp. 411–419.
[3] Halme, A., Schonberg, T., and Wang, Y., 1996, ''Motion Control of a Spherical Mobile Robot,'' *Proc. AMC'96-MIE*.
[4] Bicchi, A., Balluchi, A., Prattichizzo, D., and Gorelli, A., 1997, ''Introducing the Sphericle: An Experimental Testbed for Research and Teaching in Nonholonomy,'' *Proc. IEEE Int. Conference on Robotics and Automation*, pp. 2620–2625.
[5] Mukherjee, R., and Minor, M., 1999, ''A Simple Motion Planner for a Spherical Mobile Robot,'' IEEE/ASME Int. Conference on Advanced Intelligent Mechatronics, Atlanta, GA.
[6] Ehlers, G. W., Yavin, Y., and Frangos, C., 1996, ''On the Motion of a Disk Rolling on a Horizontal Plane: Path Controllability and Feedback Control,'' Comput. Methods Appl. Mech. Eng., **137**, pp. 345–356.
[7] Yavin, Y., 1997, ''Inclination Control of the Motion of a Rolling Disk by Using a Rotor,'' Comput. Methods Appl. Mech. Eng., **146**, pp. 253–263.
[8] Yavin, Y., 1999, ''Stabilization and Motion Control of the Motion of a Rolling Disk by Using Two Rotors Fixed Along Its Axis,'' Comput. Methods Appl. Mech. Eng., **169**, pp. 107–122.
[9] Getz, N., 1995, ''Internal Equilibrium Control of a Bicycle,'' 34th IEEE Conference on Decision and Control, New Orleans, LA.
[10] Rui, C., and McClamroch, N. H., 1995, ''Stabilization and Asymptotic Path Tracking of a Rolling Disk,'' 34th IEEE Conference on Decision and Control, New Orleans, LA.
[11] Kirk, D. E., 1970, *Optimal Control Theory: An Introduction*, Prentic-Hall, Englewood Cliffs, NJ.
[12] Greenwood, D. T., 1988, *Principles of Dynamics*, Prentice-Hall, Englewood Cliffs, NJ.

High-Frequency Low-Loss Ultrasonic Modes in Imbedded Bars

B. N. Pavlakovic

M. J. S. Lowe

P. Cawley

Department of Mechanical Engineering,
Imperial College,
Exhibition Road,
London SW7 2BX, U.K.

The dispersion relationships of a system comprising a circular bar imbedded in a solid medium having a lower acoustic impedance than the bar have been predicted. A generic study of such systems has been undertaken, motivated by a particular interest in the case of a circular steel bar imbedded in cement grout which has application to the inspection of tendons in post-tensioned concrete bridges; measurements to confirm the predictions have been carried out for this case. The attenuation dispersion curves show a series of attenuation minima at roughly equal frequency spacing. The attenuation minima occur at the same frequencies as energy velocity maxima and they correspond to points at which the particle displacements and energy of the particular mode are concentrated towards the center of the bar so leakage of energy into the imbedding medium is minimized. The attenuation at the minima decreases with increasing frequency as the energy becomes more concentrated at the middle of the bar, until the material attenuation in the bar becomes a significant factor and the attenuation at the minima rises again. For the particular case of a steel bar in cement grout, the minimum attenuation is reached at a frequency-radius product of about 23 MHz-mm. The frequency-radius product at which the minimum attenuation is reached and the value of the minimum attenuation both increase as the acoustic impedance of the imbedding medium increases.
[DOI: 10.1115/1.1347995]

1 Introduction

It is frequently necessary to inspect a long length of bar or pipe imbedded in a surrounding fluid or solid medium. Examples include pipes buried in soil, fibers imbedded in a polymer matrix and steel tendons buried in grout or concrete. Guided acoustic waves offer a potentially attractive solution to these nondestructive evaluation (NDE) problems as they can be excited at one point on the system (often at an end) and will then propagate along the bar or pipe, which acts as a waveguide.

Propagation distances of many tens of meters can readily be obtained in steel pipes or bars in air ([1–6]) since in this case the attenuation is predominantly controlled by the material attenuation of the steel, which is relatively low; the leakage into the surrounding air is minimal. When the waveguide is surrounded by a fluid, leakage of energy into the fluid by radiation of longitudinal waves is possible when the phase velocity of the guided mode exceeds the phase velocity of longitudinal waves in the fluid. When the waveguide is imbedded in a solid, leakage by both longitudinal and shear waves can occur which leads to very high attenuation rates, especially when the acoustic impedances of the waveguide and the surrounding solid are similar ([7]).

An enormous amount of work has contributed to our current understanding of cylindrical wave propagation. The following review highlights some key papers, but it is far from comprehensive. Cylindrical wave propagation problems were first studied numerically in the late 19th century. Pochhammer [8] and Chree [9] were the first researchers to investigate the propagation of guided waves in a free bar mathematically, and their names are still associated with the equation that describes the modes of a solid cylinder. However, most of the applications of cylindrical wave propagation have occurred much more recently. In the mid 20th century, a significant amount of research was performed on the modes of solid bars. Much of this work concentrated on the use of rods as acoustic waveguides for use in delay lines that could be used in electronic devices, such as radar. In 1943, Hudson used shell approximations to study the dispersive nature of the fundamental flexural mode in a solid cylinder ([10]). The longitudinal modes of a bar were first examined by Davies in 1948 ([11]). Later work by researchers such as Pao and Mindlin [12,13], Onoe et al. [14], and Meeker and Meitzler [15] fully developed all of the branches of the complete three-dimensional problem of a solid circular cylinder in vacuum. The dispersion curves for a hollow isotropic cylinder were definitively treated by Gazis in 1959 ([16]). Fitch ([17]) matched Gazis' predictions for axially symmetric and nonsymmetric wave propagation with experimental data. Later researchers such as Kumar [18,19] have examined the effect of fluid filling on wave propagation in cylinders. In 1965, Mirsky expanded Morse's work ([20]) on axisymmetric wave propagation in transversely isotropic solid cylinders and Gazis' exact nonaxisymmetric isotropic wave propagation solution ([16]) so that cylindrical wave propagation in transversely isotropic materials could be studied ([21]). Subsequently, several other authors have examined propagation in transversely isotropic rods and cylinders, for example Xu and Datta [22], Dayal [23], Nagy [24], and Berliner and Solecki [25].

Leaky cylindrical systems have been much more difficult to model than their free counterparts. Much of the difficulty comes from the need to calculate complex Bessel functions, which until recently were not readily available but which can now be obtained ([26]). Therefore, early work, such as that conducted by Thurston [27], concentrated on portions of the dispersion curves that could be calculated using only real Bessel functions. However, more recent work by Safaai-Jazi et al. [28], Simmons et al. [29], and Viens et al. [30], has been able to model the entire range of dispersion curve solutions. In addition, recent work by Nagy [24] and Berliner and Solecki [25,31] has looked at wave propagation in transversely isotropic rods that are immersed in a fluid. A recent

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Feb. 23, 2000; final revision, Aug. 15, 2000. Associate Editor: V. K. Kinra. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

Table 1 Material properties used in analysis

Material	Density (kg/m^3)	Longitudinal velocity, c_L (m/s)	Shear velocity, c_s (m/s)	Longitudinal bulk wave attenuation, α_L (nepers/ wavelength)	Shear bulk wave attenuation, α_s (nepers/ wavelength)
Steel	7932	5960	3260	0.003	0.008
Grout	1600	2810	1700	0.043	0.1
Epoxy	1170	2610	1100	···	···
Glass	2600	5570	3520	···	···
Cast iron	7100	4500	2500	···	···

paper by Nayfeh and Nagy [32] considers leaky axisymmetric waves in multilayered transversely isotropic fibers that are imbedded in a solid.

The work described here was motivated by the need to inspect the tendons in post-tensioned concrete bridges. This requirement was highlighted by the collapses of the Ynys-y-Gwas bridge in South Wales in 1985 ([33]) and of a post-tensioned bridge in Palau ([34,35]). In a post-tensioned construction, the bridge gains its strength from the tensioning of internal tendons after the concrete framework of the bridge has already hardened. The tendons can be single wires (usually found in older bridges) or strands of seven wires (usually found in newer bridges). The wires are frequently 5 or 7 mm in diameter (although they can be much larger) and the strands are typically 12–15 mm total diameter. The tendons are located in metal or plastic tubes called ducts, which will often hold several individual tendons. Corrosion protection for the tendons is provided by filling the ducts with grout (cement, water, and possibly additives) once the tendons have been tensioned. However, large air voids can be trapped in the grout. Over time these voids can fill with salt water as de-icing salts leach through small cracks in the concrete or joints between segments of the bridge. Contact with salt water causes the tendons to corrode quickly and can lead to the failure of the bridge. An effective nondestructive test method needs to be able to detect the onset of corrosion in the tendons, or the complete break of one individual tendon of the several that are present in each duct. Some empirical work has been carried out on this and analogous problems [36,37] but the nature of the propagating modes has not been determined.

Since the work originated as a study of the post-tensioned bridge inspection problem, this paper concentrates on the specific case of a steel bar imbedded in grout. However, changes in the behavior of the guided modes as the impedance of the imbedding medium is varied are investigated and the analysis and form of the results is applicable to other systems comprising a wire/bar imbedded in a solid material that has a lower acoustic impedance than that of the bar. The aim of the work was to identify modes which have a minimum of attenuation as they propagate along the waveguide formed by the bar, and so to maximize the length of the system that can be inspected from the end of the bar, or from any access point that can be created along its length.

2 Wave Propagation Solution Method

Each solution for wave propagation in cylindrical systems cited in the previous section applies to a limited range of problems. For example, the solution presented in Berliner and Solecki's work ([25]) only accommodates a single layer and does not consider materials that are imbedded in a solid. The excellent work of Nayfeh and Nagy ([32]) allows an arbitrary number of layers and the possibility of imbedding the structure in a solid; however, it does not model nonaxisymmetric wave propagation, or immersion in a fluid. The authors have developed general purpose software, DISPERSE, for the prediction of the dispersion curves of systems having either flat or cylindrical geometry with an arbitrary number of layers ([38,39]). Each layer can be an elastic isotropic material, an isotropic material with material damping, a transversely isotropic material, or a fluid. Different material types can be easily combined. The solution is valid for both axisymmetric and nonaxisymmetric wave propagation (longitudinal, torsional, and flexural modes) and can model leakage into a solid or liquid medium.

The software is based on the global matrix method for the analysis of multilayered structures that overcomes the problem of instability at high frequency-thickness products commonly associated with the Thomson-Haskell transfer function technique ([38]). The field equations for cylindrical systems are based on those of Gazis [16]. The derivation for a transversely isotropic material closely follows the technique of Mirsky [21], incorporating adaptations similar to those used by Berliner and Solecki [25]. All the predictions presented here were produced using the DISPERSE software. Further details of the derivations and the software implementation can be found in [40].

The material properties for steel and cement grout used in the analysis are given in Table 1. The values for steel were obtained from the literature ([41]) while the longitudinal and shear velocities in grout and the longitudinal wave attenuation were measured on samples made at Imperial College with a water/cement ratio of 0.6. The shear wave attenuation was too high to be measured accurately and the value given in Table 1 was estimated. However, its precise value has little effect on the rate of leakage from the bar and so does not significantly affect the results presented here. Table 1 also gives the properties of epoxy, cast iron, and glass that were used to study the effect of changes in the acoustic impedance of the imbedding medium. Since the attenuation of the imbedding medium has only a secondary effect on the predictions, these materials were modeled as perfectly elastic. Except where stated, all the results are for a steel bar imbedded in an infinite space of grout. This is a reasonable approximation to the practical case of a bar imbedded in a cylinder of grout when the cylinder diameter is much larger than the bar diameter.

The mode names used in this paper follow the format of Silk and Bainton [42]. All the modes which are discussed are propagating along the axis of the imbedded bar; $L(0,n)$ modes are axisymmetric modes having zero circumferential displacement, while $F(m,n)$ modes have displacements varying as $\cos m\theta$ around the circumference of the bar. For both types of modes, the second index, n, is used to sequentially number the modes of a given type. In general, modes of a higher order ''n'' exhibit more complicated displacement profiles through the diameter of the bar. Torsional modes are not considered in this paper.

3 Possibility of Non-Leaky Mode

The ideal solution to the problem of inspecting an imbedded structure would be to find a mode that would propagate along the structure without leaking energy into the surroundings, so reducing the attenuation and making it possible to inspect a long distance along the structure from a single transducer position. This possibility has been investigated for the case of steel imbedded in cement grout. Initial predictions were done on the analytically relatively simple case of a flat steel plate imbedded in grout and showed that a non-leaky mode does exist for this system, as demonstrated in Fig. 1(a) where it can be seen that at low frequencies, the a_0 mode, which is similar to the $F(1,1)$ mode in a bar, has a non-leaky section where its phase velocity dips below the shear

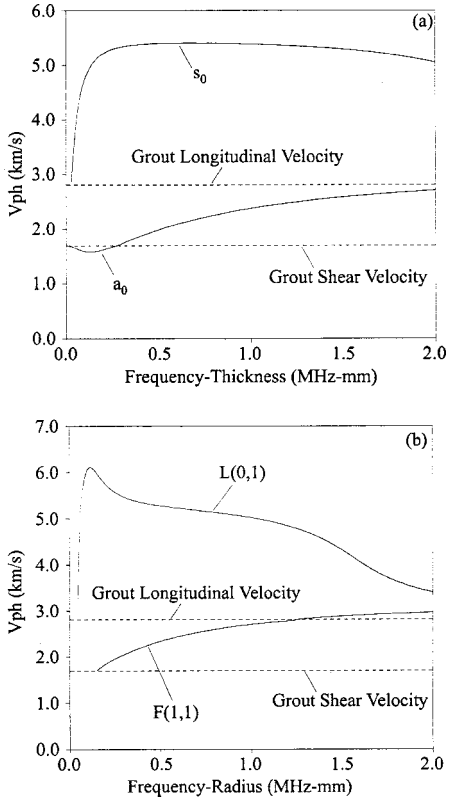


Fig. 1 Phase velocity dispersion curves for (*a*) steel plate imbedded in grout; (*b*) steel bar imbedded in grout

bulk velocity of the surrounding medium. However, as Fig. 1(*b*) shows, this non-leaky section does not exist for the equivalent cylindrical system; the $F(1,1)$ mode could not be found below the bulk shear velocity in the surrounding medium. Mathematically it is probable that the mode does continue from the point at which it reaches the bulk shear velocity, but it does not go into a non-leaky region; instead, it continues at a phase velocity equal to the bulk shear velocity and modes having a phase velocity equal to one of the bulk velocities cannot be traced using the global matrix method ([38]). For interest, Fig. 1(*b*) also shows the $L(0,1)$ mode which is similar to the s_0 mode in a plate.

The existence or nonexistence of non-leaky modes is a topic that warrants more research. The parameters that control the existence of a non-attenuating guided wave that exists at the interface of two semi-infinite solids (the Stoneley wave), have previously been studied and expressed in an explicit form ([43,44]). However, this knowledge has not been expanded to include finite thickness plates and cylinders. The derivation and experimental confirmation of the comparable conditions for an imbedded bar and plate would be very valuable in the design of ultrasonic testing systems.

Although the reason for a non-leaky mode section appearing in the Cartesian system and not in the cylindrical system is not completely understood, examination of related cases provides some insight. Figure 2 shows the calculated phase velocity dispersion curves for the cylindrical system of Fig. 1(*b*), but with the density of the ''grout'' reduced from 1600 kg/m^3 to 500, 600, 700 and 800 kg/m^3, the bulk wave velocities being kept constant at the values given in Table 1. A non-leaky section appears, but as the density increases, the frequency range over which it exists reduces and the minimum phase velocity seen in the non-leaky region approaches the bulk shear velocity in the grout. There is a discontinuity in each curve as it crosses the bulk shear velocity in the ''grout.'' For comparison, Fig. 3 shows the a_0 mode dispersion curves for the Cartesian system of Fig. 1(*a*) as the ''grout'' density is varied. In this case, the phase velocity is reduced to values well below the bulk shear velocity in the ''grout'' and the non-leaky section covers a wider frequency range than in the cylindrical case.

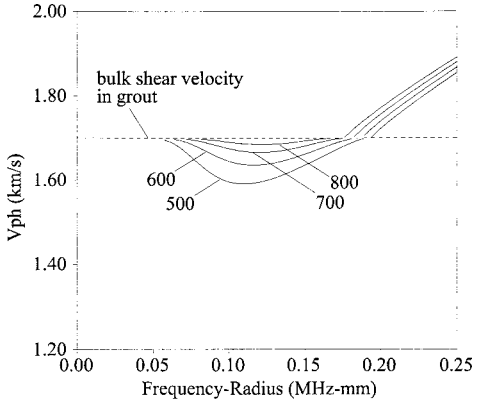


Fig. 2 Phase velocity dispersion curves for *F*(1,1) mode of steel bar imbedded in ''grout'' of varying density (densities shown in kg/m^3)

It should be noted that even if a non-leaky mode had been found, its attenuation would not be zero. This is illustrated in Fig. 4 which shows the attenuation and phase velocity dispersion curves for the a_0 mode of a steel plate imbedded in grout corresponding to the phase velocity relationships of Fig. 1(*a*), but only plotted up to a frequency-thickness product of 0.4 MHz-mm. The attenuation is plotted on a nondimensional Nepers-m/m scale so for a particular plate thickness, the attenuation per meter is given by the value read from the graph divided by the thickness. The attenuation rises sharply once the phase velocity exceeds the bulk shear velocity in the grout, but even in the non-leaky region below this, there is significant attenuation. The finite attenuation in the non-leaky region is primarily due to energy dissipation in the grout; this is possible because although there is no radiation into the grout, an inhomogeneous wave is present in the grout adjacent to the plate and the model includes the damping properties of the imbedding material. The wave fields in the steel and grout are shown in Fig. 4(c). These mode shape plots indicate that the out-of-plane (w) motion is fairly uniform across the thickness of the steel plate, while the in-plane (u) motion varies roughly linearly across the plate. This is as expected for the a_0 mode at low frequencies where it is primarily a bending mode. Both the in-plane and out-of-plane displacements are continuous across the steel-grout boundaries and then reduce exponentially with distance into the

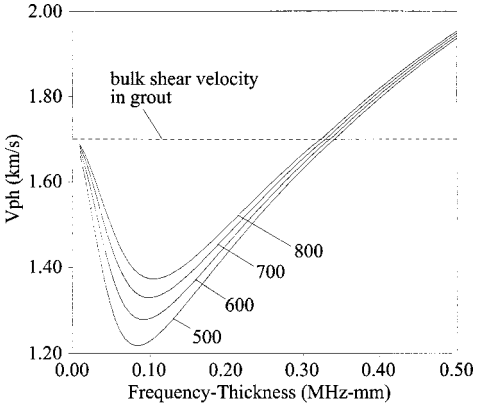


Fig. 3 Phase velocity dispersion curves for a_0 mode of steel plate imbedded in ''grout'' of varying density (densities shown in kg/m^3)

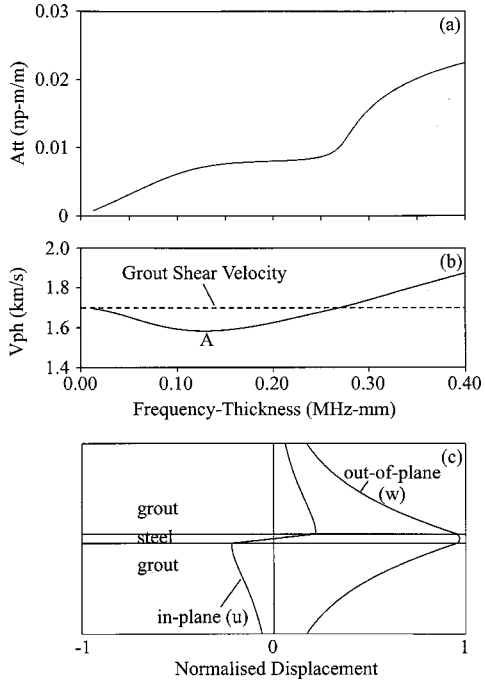


Fig. 4 Dispersion properties for a_0 mode of steel plate imbedded in grout (*a*) attenuation; (*b*) phase velocity; (*c*) mode shape at minimum phase velocity (position *A* in (*b*))

grout; this is characteristic of an inhomogeneous wave. (Figure 4(*c*) only plots the motion in the grout to a distance of 10 plate thicknesses from the plate surface, but the trend continues beyond this point.)

4 High-Frequency Modes

Having demonstrated that there is no non-leaky section of the $F(1,1)$ mode for the steel bar imbedded in grout for the case of the real properties of Table 1, attention was turned to the higher frequency modes. Figures 5(*a*), 5(*b*) and 5(*c*) show the phase, attenuation, and energy velocity dispersion curves of all the $L(0,n)$ modes of a steel bar imbedded in grout up to a frequency-radius product of 50 MHz-mm. The energy velocity rather than the group velocity has been plotted in Fig. 5(*c*) because, while the two are equivalent in a lossless or weakly attenuative system, the group velocity is not well defined in regions of high attenuation ([45]). A definite pattern emerges in the attenuation dispersion curves of Fig. 5(*b*). There is a series of modes that have sharp attenuation minima at higher frequencies. The value of the attenuation at successive minima decreases up to a frequency-radius of about 23 MHz-mm (point B) and then increases slowly.

The surprising finding that the minimum attenuation is seen at a relatively high frequency can be explained by studying the mode shapes. Figures 6(*a*) and 6(*b*) show the axial and radial displacement and strain energy profiles across the bar corresponding to the two attenuation minima marked *A* and *B* in Fig. 5. The motion and strain energy is concentrated at the center of the bar with very little motion at the interface with the grout. This limits the leakage and so explains the relatively low attenuation. An example of the mode shape away from the attenuation minima is shown in Fig. 6(*c*). Here there is significant displacement at the surface of the bar and so more leakage will occur. As the frequency-radius product increases, the strain energy at the attenuation minima is increasingly concentrated in the middle of the bar, and up to about 23 MHz-mm, the minimum value of attenuation decreases. Above 23 MHz-mm, the strain energy at the attenuation minima continues to be increasingly concentrated in the middle of the bar, but the actual value of the attenuation rises as material attenuation in

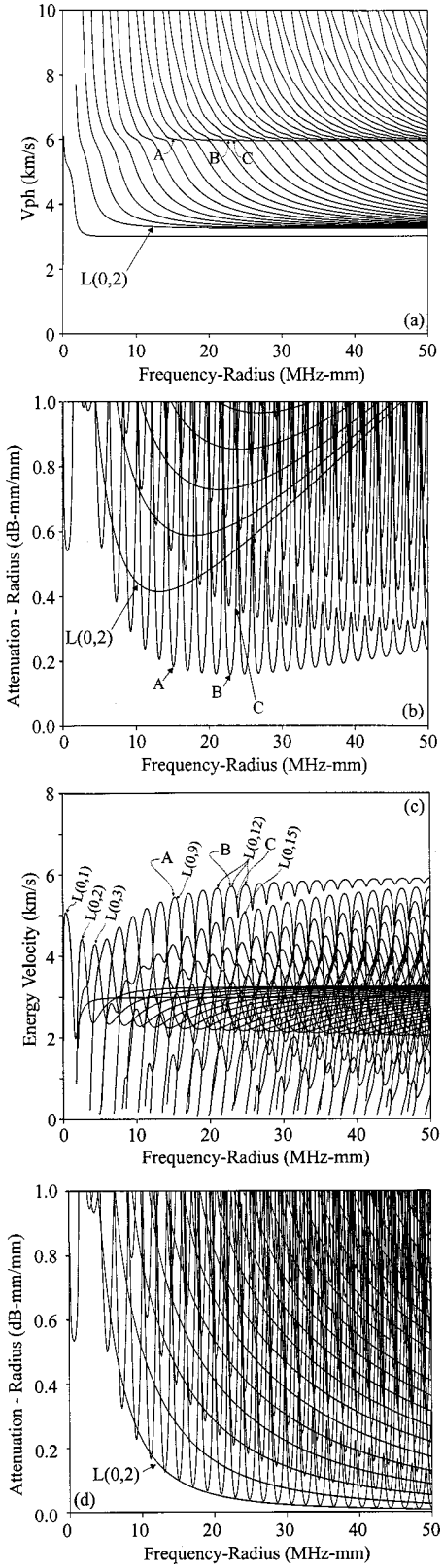


Fig. 5 Dispersion curves of axisymmetric ($L(0,n)$) modes of steel bar imbedded in grout. (*a*) phase velocity; (*b*) attenuation; (*c*) energy velocity; (*d*) as (*b*) but zero steel attenuation.

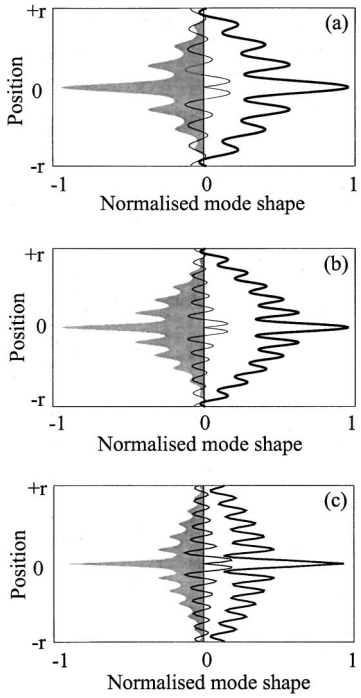


Fig. 6 Mode shapes corresponding to points marked on Fig. 5. (*a*) point *A*; (*b*) point *B*; (*c*) point *C*. (—— axial displacement, —— radial displacement, shaded—strain energy. Strain energy plotted in -ve direction to avoid confusion with axial displacement.)

the steel becomes more significant. Investigations with different assumed values of steel bulk wave attenuation showed that the frequency at which the lowest minimum occurs increases as the steel attenuation decreases. This is as expected since at lower values of steel attenuation, a higher frequency has to be reached before the loss due to material attenuation approaches the loss due to leakage. The attenuation dispersion curves for the case of zero steel attenuation are shown in Fig. 5(*d*); it can be seen that in this case, the minimum attenuation reached by successive minima continues to decrease. The $L(0,2)$ mode has the lowest attenuation when the attenuation of the steel is set to zero, whereas its attenuation was much higher in the real case of Fig. 5(*b*). A numerical study showed that the attenuation of this mode is dominated by the shear wave attenuation in the steel, whereas the longitudinal wave attenuation is more important in the higher order modes. Since the shear wave attenuation of steel is higher than the longitudinal wave attenuation, removing all the attenuation has a larger effect on modes whose attenuation is dominated by the shear wave attenuation.

It is interesting to note that the attenuation minima coincide with the energy velocity maxima. The same maxima in energy velocity exist for modes with zero attenuation in a free, elastic bar, so the energy velocity is not linked to the attenuation. The energy velocity reaches a maximum in this region because the phase velocity is just above the bulk longitudinal wave speed of the fastest material (steel) so the longitudinal partial wave is directed almost parallel to the bar. In addition at this point, the ratio of the amplitude of the longitudinal partial wave to the shear partial wave in the steel is greater than at any other location. The higher frequency flexural ($F(m,n)$) modes also have attenuation minima similar in form to those of the axisymmetric ($L(0,n)$) modes shown in Fig. 5. However, the actual values of attenuation at the minima are higher for the flexural modes than for the axisymmetric modes.

The energy velocity curves of Fig. 5(*c*) show that the first 11 energy velocity maxima, which relate to the attenuation minima of Fig. 5(*b*), correspond to the first 11 $L(0,n)$ modes in sequence. However, all the subsequent maxima correspond to the $L(0,12)$ mode. This is connected to the behavior of the phase velocity dispersion curves of Fig. 5(*a*) which are shown expanded in the region of the steel longitudinal velocity in Fig. 7(*a*). (The curves of Fig. 7(*a*) were calculated for zero material attenuation in the grout, but this makes minimal difference to the results.) The rate of change of phase velocity with frequency of the first 11 modes

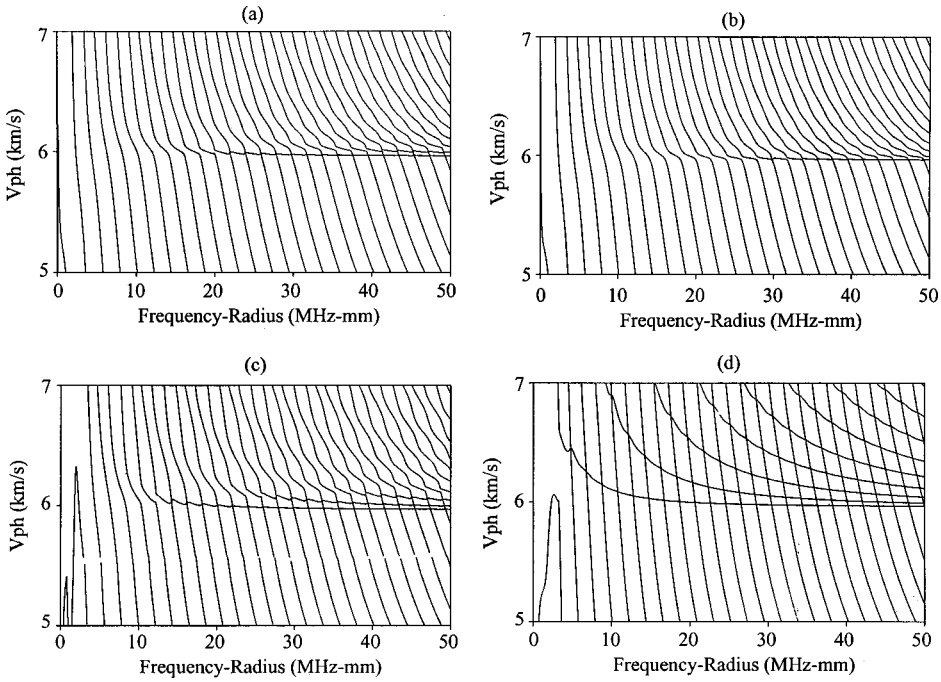


Fig. 7 Phase velocity dispersion curves for steel bar imbedded in (*a*) grout; (*b*) epoxy; (*c*) glass; (*d*) cast iron. Imbedding medium has zero attenuation in these plots.

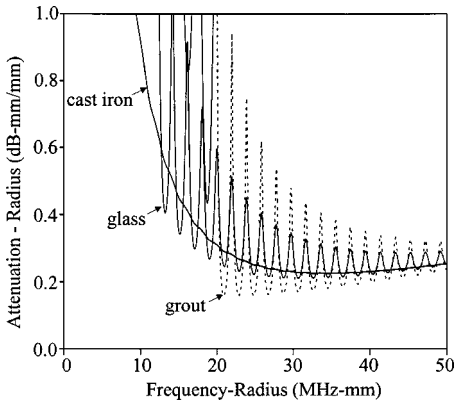


Fig. 8 Attenuation dispersion curves of first "crossing mode" of steel bar imbedded in cast iron, glass, and grout. (Grout case shown dotted for clarity.)

reduces as it approaches the steel bulk longitudinal velocity but the curve then carries on towards the bulk shear velocity. In contrast, the $L(0,12)$ mode curve does not carry on towards the bulk shear velocity but remains above the bulk longitudinal velocity, crossing the curves of the higher order modes. This mode crossing behavior has previously been observed in the case of plastic plates with high material attenuation in vacuum ([46]) where it was shown that as the attenuation increased, the mode order (and hence frequency) at which the crossing was first observed reduced.

Figures 7(*b*), 7(*c*), and 7(*d*) show phase velocity dispersion curves corresponding to those of Fig. 7(*a*) but with the steel bar imbedded in epoxy, glass and cast iron, respectively. In order to confine the study to the effect of the elastic properties of the imbedding medium, all the imbedding media were modeled with zero material attenuation; the steel attenuation was as given in Table 1. In epoxy, the mode crossing first occurs at around 40 MHz-mm, while in glass it occurs at about 15 MHz-mm and in cast iron, the first higher-order mode crosses subsequent modes and a family of modes whose phase velocity is asymptotic to the bulk longitudinal velocity, rather than the bulk shear velocity, can be seen. This behavior is exactly analogous to that seen in plastic plates with different degrees of material attenuation ([46]). However, in the case discussed in this paper, the loss mechanism controlling the mode crossing is leakage into the surrounding medium, rather than material attenuation. Around realistic values for steel such as those given in Table 1, the bulk wave attenuation in the imbedded bar has little effect on the frequency-radius product at which the mode crossing phenomenon first occurs; it would have an effect if it approached the values seen in highly attenuative plastics. In the glass and cast iron cases of Figs. 7(*c*) and 7(*d*), the first mode does not tend to infinite phase velocity at zero frequency, but instead the phase velocity has a maximum before tending to zero at zero frequency. This phenomenon has been discussed by Nayfeh and Nagy [32]. Discontinuities can be seen in the glass case of Fig. 7(*c*) at a phase velocity of around 5.5 km/s. This corresponds to the bulk longitudinal wave velocity in the glass; discontinuities are frequently seen at the bulk wave velocities of the imbedding media as below these velocities, the corresponding leaking waves become inhomogeneous.

Figure 8 shows the attenuation dispersion curves for the first mode displaying the mode crossing behavior for steel imbedded in cast iron, glass, and grout corresponding to the phase velocity plots of Fig. 7. It is clear that the minimum attenuation reached in the grout case is lower than that reached in the other materials and that it occurs at a lower frequency. The low attenuation minima in the grout case shown in Fig. 8 do not extend below about 23 MHz-mm because another mode becomes the lowest attenuation mode at lower frequencies, as shown in Fig. 5. The epoxy case is not shown in Fig. 8 to avoid over-complicating the diagram. Table 2 shows the minimum attenuation reached in the four cases and the frequency at which it occurs. Thus, as the impedance of the imbedding material increases, the value of the minimum attenuation increases and it occurs at a higher frequency.

Table 2 Minimum attenuations with different imbedding media and frequencies at which they occur. In these predictions the imbedding media have zero-bulk wave attenuation.

Imbedding material	Minimum attenuation (dB-m/m)	Frequency at which minimum attenuation reached (MHz-mm)
Cast iron	0.225	34.5
Glass	0.211	30.7
Grout	0.159	22.9
Epoxy	0.148	21.0

5 Experimental Investigation

5.1 Specimens and Setup. The experimental investigation was designed to verify the predictions of the existence of relatively low loss modes in the steel-grout system. The small surface displacements in the mode shapes shown in Fig. 6 indicate that it would be difficult to excite the low attenuation modes by applying surface tractions to the circumference of the bar. However, they should be relatively easy to excite by applying an axial force to the end of the bar.

Two test specimens were constructed, both comprising an 8.1-mm diameter mild steel bar at the center of a plastic pipe filled with grout. The pipe had an internal diameter of approximately 100 mm and so was similar to the ducts used in post-tensioned bridges. The grout had a water to cement ratio of 0.6 (no additives) and was pumped into the plastic pipe using a small version of the pumps typically used in grouted tendon construction. Slightly corroded steel bars were used to improve the bonding between the steel and the grout. The grouted sections covered about two meters of the bars, leaving sections around 50 mm long protruding from the grout at each end. For one of the specimens, the bar was undamaged, while for the second specimen, notches were cut approximately 500 mm from each end of the bar. The notches were created with a saw, one cut being 2 mm deep while the cut at the other end of the bar was 4 mm deep (i.e., about half way through the bar). The notches were not covered when the grout was poured so they were probably filled with grout.

The tests were performed using a LeCroy 9101 arbitrary waveform generator that sent a windowed toneburst to a standard 5 MHz center frequency, unfocussed ultrasonic immersion transducer (Krautkramer Branson 0.5-inch diameter, 5 MHz alpha series) via a custom-built power and receiver amplifier that has pulse-echo capabilities. The output voltage from the power amplifier was approximately 55V peak-peak. Both through transmission tests, with the transmitting transducer at one end of the bar and the receiver at the other end, and pulse echo tests with a single transducer acting as both transmitter and receiver at one end of the bar were carried out. The transducers were gel coupled to the plane ends of the bar and held in place by a spring loaded clamp. The received signals were averaged 250 times on a digital oscilloscope. Although relatively low attenuation minima are predicted in some modes, it is necessary to measure signals which attenuate by over 100 dB over a 2 m path length if the dispersion curves are to be measured over a significant frequency range away from the minima. In cases where the signal-noise ratio was low, the received signal was passed through an analogue band-pass filter before being fed to the oscilloscope.

5.2 Results. Figure 9 shows the results of through transmission tests on the undamaged bar. The excitation signal was a

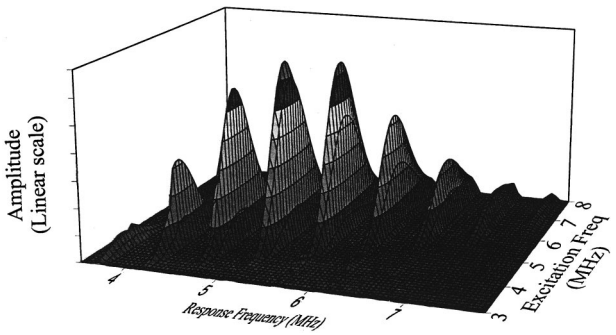


Fig. 9 Spectrum of response of signal transmitted through 2 m imbedded bar as excitation toneburst stepped between 3.0 and 8.0 MHz

10-cycle Gaussian windowed toneburst whose center frequency was varied in steps from 3.0 to 8.0 MHz. The signal received at the other end of the bar was Fourier transformed and Fig. 9 shows the resulting spectrum of the measured response as the center frequency of the excitation was varied. A series of response peaks running parallel to the excitation frequency axis can be seen, indicating that the bar acts like a mechanical filter that passes only selected narrow frequency bands. Hence, the response at any given excitation center frequency is proportional to the amplitude of the spectrum of the excitation in these ''pass bands.'' Figure 10 shows the maximum response obtained at any excitation frequency between 3 and 8 MHz as a function of frequency. (i.e., Fig. 10 is a view of the three dimensional plot of Fig. 9 looking parallel to the excitation frequency axis.) The dotted line shown in Fig. 10 was obtained by carrying out a similar experiment to that of Fig. 9, but instead of transmitting the signal through the imbedded bar, it was transmitted through a 50-mm-thick steel block. It therefore describes the form of the response obtained when the same transducers, amplifiers, and excitation are used on a system with negligible attenuation. The amplitude scales of the dotted and solid lines on Fig. 10 are very different since the minimum attenuation along the imbedded bar was of the order of 75 dB higher than the attenuation through the block. In principle it would be have been possible to compare the absolute amplitudes obtained in the two experiments, but the coupling of the 13-mm-diameter transducers to the large steel block was more satisfactory than that to the 8.1-mm-diameter bar, and the absolute amplitude of the response of the bar also changed in different experiments due to coupling variations. It would have been possible to obtain the curves of Fig. 10 in a single experiment with a relatively broadband input. However, carrying out multiple experiments with narrow band excitation at different center frequencies gave a much better signal/noise ratio, this being particularly significant in the regions between the response peaks.

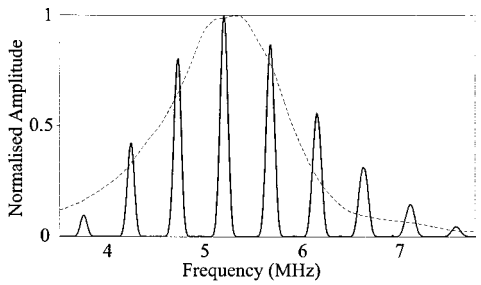


Fig. 10 Maximum response as function of frequency. ——imbedded bar results of Fig. 9; - - - - similar test on 50-mm-thick steel block. Note: The scales on the two plots are independently normalized.

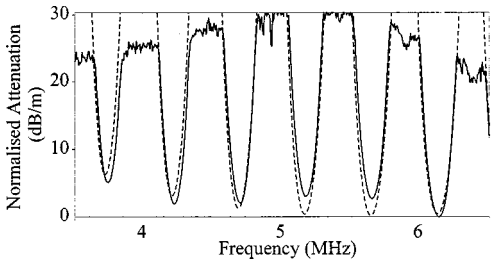


Fig. 11 Predicted (- - - -) and measured (——) normalized attenuation curves for 8.1-mm steel bar imbedded in grout

The form of the frequency response of the imbedded bar can be obtained by dividing the response obtained in the test on the bar (the solid line of Fig. 10) by the response in the test on the steel block (the dotted line of Fig. 10). This frequency response was converted to an attenuation plot by using the relationship

$$\text{Attenuation}(dB) = -20 \log_{10} \frac{a}{a_{\text{ref}}} \tag{1}$$

where a is the amplitude of the normalized frequency response of the imbedded bar and a_{ref} is a reference amplitude. In this case, a_{ref} was taken as the maximum of the frequency response function so the relative attenuation was 0 dB at this frequency. The solid line of Fig. 11 shows the results of this calculation, the result of Eq. (1) being converted to dB/m by dividing by the 2 m length of the imbedded bar. The curve flattens to a series of jagged peaks at higher values of attenuation; this is due to the transmitted signal reducing to the noise floor in these regions. The dotted line of Fig. 11 shows the predicted attenuation of an 8.1-mm-diameter steel bar imbedded in grout, the attenuation being normalized to its minimum value. It should be stressed that the experimental and predicted curves have been independently normalized to their respective minimum values and no fitting of the predictions to the experiments has been done. The agreement between the experimental and predicted curves is very good, both showing a series of sharp attenuation minima that occur at regular frequency spacing. The predicted and measured widths and frequency locations of the minima are very similar and the depth of the attenuation minima follow the same trend for both cases. In general, the higher frequency modes have lower attenuation than the lower frequency

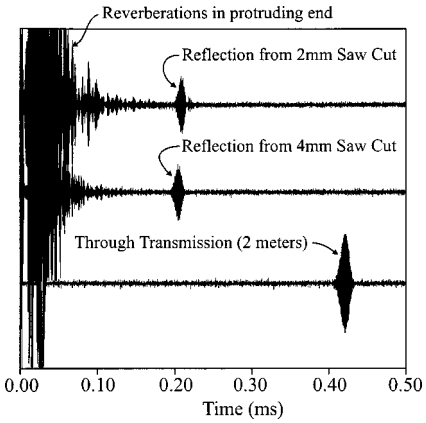


Fig. 12 Time traces from tests on 8.1-mm steel bars imbedded in grout: pulse-echo tests on one bar with a 2-mm saw cut 450 mm into the grout and on another bar with a 4-mm saw cut 450 mm into the grout; through transmission test over 2 m length of undamaged imbedded bar. Excitation was 50 cycle, 3.75 MHz Hanning windowed toneburst.

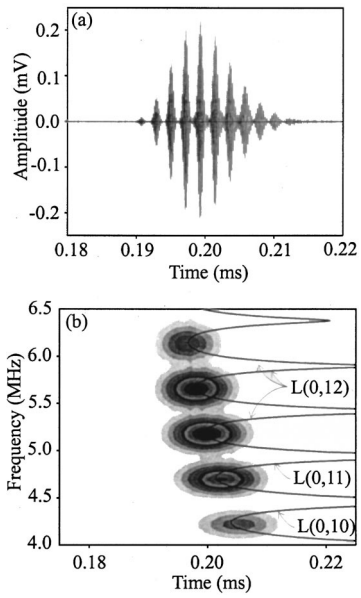


Fig. 13 Reflection of 5-cycle 5.2 MHz center frequency Gaussian windowed toneburst from 4-mm-deep notch in 8.1-mm-diameter imbedded bar. (*a*) Time domain signal; (*b*) wavelet transform of (*a*) showing amplitude as gray scale (black high) in time-frequency plane. Lines show predicted group delays obtained from energy velocity and overall propagation distance.

modes. This good agreement between the experimental and predicted results increases confidence in the accuracy of the modeling.

The bottom trace of Fig. 12 shows the signal received in a through transmission test on the undamaged bar when the excitation was a 3.75-MHz 50-cycle Hanning windowed tone burst. The top two traces show the signal obtained in pulse-echo experiments from the two ends of the damaged bar at the same frequency and on the same amplitude scale. The echoes from the 2-mm and 4-mm deep saw cuts approximately 450 mm from where the bars enter the grout can clearly be seen. The multiple echoes at the beginning of the response correspond to reverberations in the short length of bar that protrudes from the grout. Reflections from the saw cut close to the end of the bar remote from the transducer would appear at a propagation distance of 3.1 m (1.55 m to the defect and back), which is beyond the range plotted in Fig. 12. There was no evidence of these reflections in the received signal since the attenuation produced by the extra propagation distance reduced their amplitude to well below the noise floor. The signals from the two saw cuts shown in Fig. 12 appear to be of similar amplitude. The reflection coefficient from notches in imbedded bars is not simply proportional to the area removed by the notch ([47]) but a larger reflection would be expected from the deeper notch. The similar measured amplitudes may be due to coupling variations at the two ends of the bar or to differences in the extent to which the notches are filled with grout. This test shows that pulse-echo testing can be used to detect defects in imbedded bars. The test range could be improved by optimizing the frequency used to the minimum attenuation and using resonant, rather than broadband, transducers matched to this frequency with carefully tailored excitation and response filters. The number of cycles in the input signal could also be increased.

Careful examination of the reflected and transmitted signals reveals interesting behavior of these guided wave modes. Figure 13(*a*) shows a detailed view of the signal reflected from the 4-mm-deep notch in a pulse echo test. The transducer was placed at the end closer to the notch and the excitation was a 5 cycle, 5.2 MHz Gaussian windowed tone burst. The wavelet transform of this signal obtained using the Morlet wavelet ([48]) is shown as a contour plot in Fig. 13(*b*). This illustrates how the frequency content of the received signal changes with time. Each of the guided wave modes that can exist within the frequency bandwidth of the signal is excited and travels at a slightly different speed. The interference of the various modes causes the complicated time signal shown in Fig. 13(*a*); the different wave packets seen in Fig. 13(*a*) are not the separate modes. For comparison, the solid lines in Fig 13(*b*) represent the energy velocity dispersion curves for an 8.1-mm imbedded bar after they have been converted into a time delay for a signal to travel the distance to the notch and back. The modes excited in this frequency band are $L(0,10)$, $L(0,11)$, and $L(0,12)$. The agreement between the predicted and measured arrival time for each of the modes is very good. Figure 13(*b*) also shows that the response maxima coincide with the time delay minima, which confirms that the energy velocity maxima occur at the same frequencies as the attenuation minima.

6 Conclusions

It has been shown that the dispersion curves of a circular bar imbedded in a lower impedance medium have a series of attenuation minima at roughly equal frequency spacing. The attenuation minima occur at the same frequencies as energy velocity maxima and they correspond to points at which the particle displacements and energy of the particular mode are concentrated towards the center of the bar so leakage of energy into the imbedding medium is minimized. The attenuation at the minima decreases with increasing frequency as the energy becomes more concentrated at the middle of the bar, until the material attenuation in the bar becomes a significant factor and the attenuation at the minima rises again. For a steel bar in cement grout, the minimum attenuation is reached at a frequency-radius product of about 23 MHz-mm. The frequency-radius product at which the minimum attenuation is reached and the value of the minimum attenuation both increase as the acoustic impedance of the imbedding medium increases. Excellent agreement has been obtained between the predictions and experimental measurements on a steel bar in cement grout. The results indicate that with further transducer and instrumentation development, it will be possible to inspect a significant length of imbedded bar using excitation at the free end of the bar. This has application to the inspection of the tendons in post-tensioned bridges at locations close to the anchor points at the ends of the bridge. The knowledge of these mode phenomena also offers potential for maximising the propagation distance of guided modes in other imbedded bar systems. It has also been shown that the non-leaky mode that exists in the case of a flat plate imbedded in cement grout does not exist in the corresponding case of a circular bar imbedded in grout although it may exist for other material combinations.

References

[1] Alleyne, D. N., Lowe, M. J. S., and Cawley, P., 1998, ''The Reflection of Guided Waves From Circumferential Notches in Pipes,'' ASME J. Appl. Mech., **65**, pp. 635–641.
[2] Alleyne, D. N., Cawley, P., Lank, A. M., and Mudge, P. J., 1997, ''The Lamb Wave Inspection of Chemical Plant Pipework,'' *Review of Progress in Quantitative NDE*, Vol. 16, D. O. Thompson and D. E. Chimenti, eds., Plenum Press, New York, pp. 1269–1276.
[3] Mohr, W., and Höller, P., 1976, ''On Inspection of Thin Walled Tubes for Transverse and Longitudinal Flaws by Guided Ultrasonic Waves,'' IEEE Trans. Sonics Ultrason., **SU-23**, pp. 369–374.
[4] Böttger, W., Schneider, H., and Weingarten, W., 1987, ''Prototype EMAT System for Tube Inspection With Guided Ultrasonic Waves,'' Nucl. Eng. Des., **102**, pp. 356–376.
[5] Rose, J. L., Ditri, J. J., Pilarski, A., Rajana, K., and Carr, F. T., 1994, ''A Guided Wave Inspection Technique for Nuclear Steam Generator Tubing,'' NDT & E Int., **27**, pp. 307–330.
[6] Alers, G. A., and Burns, L. R., 1987, ''EMAT Designs for Special Applications,'' Mater. Eval., **45**, pp. 1184–1189.
[7] Lowe, M. J. S., and Cawley, P., 1994, ''The Applicability of Plate Wave Techniques for the Inspection of Adhesive and Diffusion Bonded Joints,'' J. NDE, **13**, pp. 185–200.
[8] Pochhammer, J., 1876, ''Uber die fortpflanzungsgeschwindigkeiten kleiner

schwingungen in einem unbergrenzten isotropen kreiscylinder,'' J. Reine Angew. Math., **81**, pp. 324–336.
[9] Chree, C., 1989, ''The Equations on an Isotropic Elastic Solid in Polar and Cylindrical Coordinates, Their Solutions, and Applications,'' Trans. Cambridge Philos. Soc., **14**, pp. 250–369.
[10] Hudson, G. E., 1943, ''Dispersion of Elastic Waves in Solid Circular Cylinders,'' Phys. Rev., **63**, pp. 46–51.
[11] Davies, R. M., 1948, ''A Critical Study of the Hopkinson Pressure Bar,'' Philos. Trans. R. Soc. London, Ser. A, **240**, pp. 375–457.
[12] Pao, Y.-H., and Mindlin, R., 1960, ''Dispersion of Flexural Waves in an Elastic, Circular Cylinder,'' J. Appl. Mech., **27**, pp. 513–520.
[13] Pao, Y. H., 1962, ''The Dispersion of Flexural Waves in an Elastic, Circular Cylinder—Part 2,'' J. Appl. Mech., **29**, pp. 61–64.
[14] Onoe, M., McNiven, H., and Mindlin, R., 1962, ''Dispersion of Axially Symmetric Waves in Elastic Solids,'' J. Appl. Mech., **29**, pp. 729–734.
[15] Meeker, T. R., and Meitzler, A. H., 1972, ''Guided Wave Propagation in Elongated Cylinders and Plates,'' *Physical Acoustics, Principles and Methods*, Vol. 1A, W. P. Mason and R. N. Thurston, eds., Academic Press, New York, pp. 111–167.
[16] Gazis, D., 1959, ''Three Dimensional Investigation of the Propagation of Waves in Hollow Circular Cylinders,'' J. Acoust. Soc. Am., **31**, No. 5, pp. 568–578.
[17] Fitch, A., 1963, ''Observation of Elastic-Pulse Propagation in Axially Symmetric and Nonaxially Symmetric Longitudinal Modes of Hollow Cylinders,'' J. Acoust. Soc. Am., **35**, No. 5, pp. 706–708.
[18] Kumar, R., 1971, ''Flexural Vibrations of Fluid-Filled Circular Cylindrical Shells,'' Acustica, **24**, pp. 137–146.
[19] Kumar, R., 1972, ''Dispersion of Axially Symmetric Waves in Empty and Fluid-Filled Cylindrical Shells,'' Acustica, **27**, No. 6, pp. 317–329.
[20] Morse, R., 1954, ''Compressional Waves Along an Anisotropic Circular Cylinder Having Hexagonal Symmetry,'' J. Acoust. Soc. Am., **26**, No. 6, pp. 1018–1021.
[21] Mirsky, I., 1965, ''Wave Propagation in Transversely Isotropic Circular Cylinders Part I: Theory,'' J. Acoust. Soc. Am., **37**, No. 6, pp. 1016–1021.
[22] Xu, P.-C., and Datta, S., 1991, ''Characterization of Fibre-Matrix Interface by Guided Waves: Axisymmetric Case,'' J. Acoust. Soc. Am., **89**, No. 6, pp. 2573–2583.
[23] Dayal, V., 1993, ''Longitudinal Waves in Homogeneous Anisotropic Cylindrical Bars Immersed in Fluid,'' J. Acoust. Soc. Am., **93**, No. 3, pp. 1249–1255.
[24] Nagy, P., 1995, ''Longitudinal Guided Wave Propagation in a Transversely Isotropic Rod Immersed in Fluid,'' J. Acoust. Soc. Am., **98**, No. 1, pp. 454–457.
[25] Berliner, M., and Solecki, R., 1996, ''Wave Propagation in Fluid-Loaded, Transversely Isotropic Cylinders. Part I. Analytical Formulation,'' J. Acoust. Soc. Am., **99**, No. 4, pp. 1841–1847.
[26] Amos, D. E., 1998, ''A Remark on Algorithm 644: A Portable Package for Bessel Functions of a Complex Argument and Non-Negative Order,'' ACM Trans. Math. Softw., **21**, No. 4, pp. 388–393.
[27] Thurston, R., 1978, ''Elastic Waves in Rods and Clad Rods,'' J. Acoust. Soc. Am., **64**, No. 1, pp. 1–37.
[28] Safaai-Jazi, A., Jen, C.-K., and Farnell, G., 1966, ''Cutoff Conditions in an Acoustic Fiber With Infinitely Thick Cladding,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control, **UFFC-33**, No. 1, pp. 69–73.
[29] Simmons, J., Drescher-Krasicka, E., and Wadley, H., 1992, ''Leaky Axisymmetric Modes in Infinite Clad Rods. I,'' J. Acoust. Soc. Am., **92**, No. 2, pp. 1061–1090.
[30] Viens, M., Tshukahara, Y., Jen, C., and Cheeke, J., 1994, ''Leaky Torsional Modes in Infinite Clad Rods,'' J. Acoust. Soc. Am., **95**, No. 2, pp. 701–707.
[31] Berliner, M., and Solecki, R., 1996, ''Wave Propagation in Fluid Loaded, Transversely Isotropic Cylinders. Part II. Numerical Results,'' J. Acoust. Soc. Am., **99**, No. 4, pp. 1848–1853.
[32] Nayfeh, A. H., and Nagy, P., 1996, ''General Study of Axisymmetric Waves in Layered Anisotropic Fibers and Their Composites,'' J. Acoust. Soc. Am., **99**, No. 2, pp. 931–941.
[33] Woodward, R., and Williams, F., 1988, ''Collapse of the Ynys-Y-Gwas Bridge, West-Glamorgan,'' Proc. Inst. Civil Eng., **84**, Aug., pp. 635–669.
[34] Parker, D., 1996, ''Tropical Overload,'' New Civil Eng., 12/26 Dec., pp. 18–21.
[35] Parker, D., 1996, ''Pacific Bridge Collapse Throws Doubts on Repair Method,'' New Civil Eng., Oct 17, pp. 3–4.
[36] Niles, G. B., 1996, ''In Situ Method of Inspecting Anchor Rods for Section Loss Using the Cylindrically Guided Wave Technique,'' IEEE Trans. Power Deliv., **11**, No. 3, pp. 1601–1605.
[37] Weight, J. P., 1994, private communication concerning ultrasonic tests on grouted tendons carried out in 1994, City University, London.
[38] Lowe, M. J. S., 1995, ''Matrix Techniques for Modelling Ultrasonic Waves in Multilayered Media,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control, **42**, pp. 525–542.
[39] Pavlakovic, B., Lowe, M. J. S., Alleyne, D. N., and Cawley, P., 1997, ''Disperse: A general Purpose Program for Creating Dispersion Curves,'' *Review of Progress in Quantitative NDE*, Vol. 16, D. O. Thompson and D. E. Chimenti, eds., Plenum Press, New York, pp. 185–192.
[40] Pavlakovic, B. N., 1998, ''Leaky Guided Ultrasonic Waves in NDT,'' Ph.D. thesis, Imperial College, University of London, London, available at http://wallop.me.ic.ac.uk/ndt/theses/brian_thesis.pdf.
[41] Kaye, G., and Laby, T., 1995, *Tables of Physical and Chemical Constants*, 16 Ed., Harlow: Longman's, London.
[42] Silk, M., and Bainton, K., 1979, ''The Propagation in Metal Tubing of Ultrasonic Wave Modes Equivalent to Lamb Waves,'' Ultrasonics, **17**, pp. 11–19.
[43] Stoneley, R., 1924, ''Elastic Waves at the Surface of Separation of Two Solids,'' *Conference of the Royal Society*, Royal Society, London, pp. 416–428.
[44] Scholte, J., 1947, ''The Range of Existence of Rayleigh and Stoneley Waves,'' Geophysics, **5**, pp. 120–126.
[45] Bernard, A., Deschamps, M., and Lowe, M. J. S., 1999, ''Energy Velocity and Group Velocity for Guided Waves Propagating Within an Absorbing or Non-absorbing Plate in Vacuum,'' *Review of Progress in Quantitative NDE*, **18**, D. O. Thompson and D. E. Chimenti, eds., Plenum Press, New York, in press.
[46] Chan, C. W., and Cawley, P., 1998, ''Lamb Waves in Highly Attenuative Plastic Plates,'' J. Acoust. Soc. Am., **104**, pp. 874–881.
[47] Pavlakovic, B., Lowe, M. J. S., and Cawley, P., 1999, ''Prediction of Reflection Coefficients From Defects in Embedded Bars,'' *Review of Progress in Quantitative NDE*, **18**, D. O. Thompson and D. E. Chimenti, eds., Plenum Press, New York, in press.
[48] Önsay, T., and Haddow, A. G., 1994, ''Wavelet Transform Analysis of Transient Wave Propagation in a Dispersive Medium,'' J. Acoust. Soc. Am., **95**, No. 3, pp. 1441–1449.

On Crack Initiation Mechanisms in Fretting Fatigue

B. Yang
Department of Aeronautics and Astronautics
Air Force Research Laboratory,
Wright-Patterson AFB, OH 45433

S. Mall
Fellow ASME,
Materials and Manufacturing Directorate
Air Force Research Laboratory,
Wright-Patterson AFB, OH 45433

By using the crack analogue model of rigid flat-ended contact, crack initiation in fretting fatigue is analyzed. The coefficient of friction at the edge of contact, which characterizes the asymptotic stress field, is considered as the primary controlling parameter in the process. Meanwhile, the maximum tangential stress criterion and the maximum shear stress criterion are used to predict opening-mode and shear-mode crack initiations, respectively. By examining the model prediction and comparing it with experimental observations, it is shown that the observed microcracks at the small angles to a fretting surface were nucleated in shear mode in the early stage of tests with a smooth initial surface, while the microcracks at the large angles were nucleated in opening mode in the later stage with a rough worn surface. This understanding may help to establish the sequential damage mechanisms in the complex process of fretting fatigue.
[DOI: 10.1115/1.1344901]

1 Introduction

Fatigue life and endurance limit of solids are significantly reduced by contact and cyclic fretting when added to plain fatigue conditions; fatigue in this situation is called fretting fatigue. It has been well known that the contact and cyclic fretting loads activate flaws at the contact surface, which are dormant in plain fatigue conditions, to develop cracks. However, a thorough analysis of initiation of fretting fatigue is difficult because many factors are involved in the process ([1]). Thus, a simple model including one, or at most few key factors, becomes important and credible as long as it ably takes into account some of the common observations in fretting fatigue, such as described in the following. In a typical fretting fatigue test, multiple cracks are often found near the edge of contact ([2,3]) or near the slip-stick boundary ([4,5]). These cracks are nucleated at angles less than 90 deg on the fretting boundary beneath the pad. However, the angles vary widely from 25 deg to 80 deg in various materials [3,6–14]. In addition, the coefficient of friction between the specimen and pad (measured as the average over the entire contact surface) evolves from an initially small value of ~0.2–0.4 to a large value of ~0.7–1.2 due to wear and asperity adhesion ([5,7,11]). There are also other characteristic features just beyond the initiation, such as the kinked knee shape of a fretting fatigue crack ([7,10]). On the other hand, the modeling and analysis of fretting fatigue initiation have been based largely on a noncrack or nonfracture mechanics approach in which the stress and strain in critical planes along the contact surface are used to formulate a criterion ([13,15–18]). The noncrack approach is simple in formulation and is easy to apply in the engineering practice. Also, it may work well for the fretting fatigue with a nonconcentrated contact stress field. However, in many cases of fretting fatigue, the stress and strain are highly concentrated near the edge of contact, and consequently are sensitive to slight changes in the fretting conditions, causing difficulties in applying the noncrack approach. In these cases, the initiation of damage could be immediate due to the highly concentrated state of stress and strain compared to a finite strength of materials. However, the nucleation of a crack may or may not occur depending upon the intensity of stress (equivalently, energy release rate) rather than by the strength-type driving forces at critical planes. Recently, Giannakopoulos et al. [19] have proposed the crack analogue approach of fretting contact, and identified some important aspects of the equivalence between contact mechanics and fracture mechanics validating the approach under the condition of small-scale yielding. This approach ably facilitates the analysis of crack initiation in fretting fatigue in the cases of high stress concentration. Giannakopoulos et al. [20] further considered the effects of adhesion in contact fatigue using the crack analogue approach.

The present work is intended to analyze crack initiation mechanisms in fretting fatigue by applying the crack analogue approach. In Section 2, characteristics of the asymptotic stress field at the edge of contact by a rigid flat-ended punch pressing on an elastic substrate are discussed first. For the case where the asymptotic stress field may be characterized by using the stress intensity factors, the crack analogue is invoked. The criteria for crack initiation under mixed-mode loading within the framework of linear elastic fracture mechanics (LEFM) are then summarized, including the maximum tangential stress (MTS) criterion for opening-mode crack extension and the maximum shear stress (MSS) criterion for shear-mode crack extension. These criteria are used to examine initiation angles of a crack at the edge of contact. In Section 3, the predictions by the model and their implications related to experimental observations are discussed. The coefficient of friction at the edge of contact, which characterizes the asymptotic stress field, serves as the primary controlling parameter in this process. It is found that the driving force for shear-mode crack initiation is dominant over the driving force for opening-mode crack initiation with a small value of the coefficient of friction, and the dominance by shear mode crack initiation diminishes with an increasing value of the coefficient of friction. By correlating the model predictions to the experimental observations of crack angles and evolving coefficient of friction, this study shows that the small-angle cracks were nucleated in shear mode in the early stage of tests with smooth surfaces (i.e., a small coefficient of friction), while the large-angle cracks were nucleated in opening mode in the later stage with rough worn surfaces (i.e., a large coefficient of friction). At last, some conclusions are drawn in Section 4.

2 Crack Analogue Approach for Crack Initiation in Fretting Fatigue

Fretting fatigue involves environmental, chemical, and mechanical factors at the contact surfaces. A complete consideration of the process appears to be extremely difficult. The present work focuses on the mechanical part of the process by applying a simple model of fretting contact. Although the pad geometry and material similarity between the pad and substrate in reality may

Contributed by the Applied Mechanics Division of The American Society of Mechanical Engineers for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Sept. 7, 1999; final revision, May 9, 2000. Associate Editor: K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

vary, the configuration with a rigid flat-ended punch pressing on an elastic substrate is considered in this study. The advantage of the simple model is that only one parameter, i.e., the coefficient of friction between the substrate and pad, characterizes the asymptotic stress field. In the following, the asymptotic stress field at the edge of contact is first discussed. For the case where the asymptotic stress field is singular on the order of 0.5, the crack analogue approach is invoked. Then, the MTS and MSS criteria for crack extension are described, within the framework of LEFM.

2.1 Asymptotic Stress Field. Consider a homogeneous, isotropic, linearly elastic body in half plane, indented by a rigid rectangular flat-ended punch of width $2a$, as shown in Fig. 1(a). The Cartesian coordinates (x,y), and the polar coordinates (r,θ), both with the origin at the left edge of contact, are selected. When these two interact, the rigid punch transmits in general a normal force P, a tangential force Q, and a moment M (relative to the point $(x=a,y=0)$) into the substrate. In order to utilize the analytical solution available in the literature, it is further assumed that the condition of gross slip exists between the substrate and the pad, and the punch indents the substrate surface perpendicularly without rotation. However, these conditions may be relaxed if the local crack initiation only is of interest, as discussed later.

Under the conditions assumed, the asymptotic stress field around the edges of contact is known in the literature ([21]). The singularity of the asymptotic stress field is on the order of R^{-m} at the left edge of contact and is on the order of R^{m-1} at the right edge of contact, where R is the distance from the edges and $m=\tan^{-1}(2(1-\nu)/(1-2\nu)/f_{sp})/\pi$, for $Q>0$. If $Q<0$, the two edges switch the stress fields. In the above expression, ν is Poisson's ratio of the substrate, and f_{sp} is the coefficient of friction between the substrate and the pad. In particular, for the case either with $\nu=0.5$ or with $f_{sp}=0$, it turns out that $m=0.5$, showing the same order of stress singularity as for a sharp crack in the LEFM analysis ([22]). Otherwise, the order of stress singularity in the substrate is less than 0.5 at the left edge of contact and is greater than 0.5 at the right edge of contact. Note that the order of stress singularity other than 0.5 in LEFM is not physically meaningful for a successful, stable crack extension ([22]). Some of the restrictions, such as the nonrotational indentation of the rigid punch to the substrate surface, should be relaxed so that the stress singularity at the edges of contact might show the meaningful order of 0.5 leading to a nontrivial but finite energy release rate. Also, for a study involving only local crack initiation at the edge of contact, the condition of gross slip may be relaxed by assuming that if partial slip occurs, it does not affect the order of singularity in the asymptotic stress field. The present work confines itself to the case with $m=0.5$ where the physical meaning of the stress singularity for crack development in isotropic, linearly elastic solids is clear.

For the case with $m=0.5$, the crack analogue to the contact configuration is readily obtained, as shown in Fig. 1(b) ([19]). The

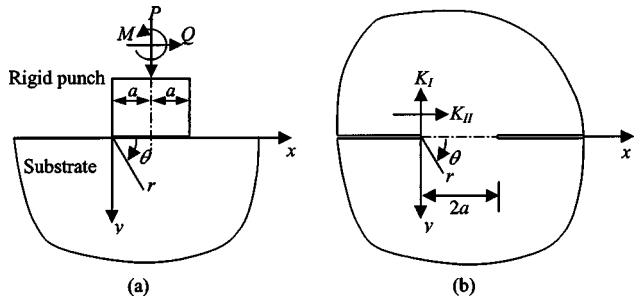


Fig. 1 (*a*) Fretting contact by a rigid flat-ended punch; (*b*) crack analogue of the fretting contact configuration in (*a*). The Cartesian coordinates (x,y) and the polar coordinates (r,θ) are shown for both configurations.

leading terms of the asymptotic stress field at the (left) edge of contact in the substrate, which is modeled as a sharp crack tip, are written as

$$\begin{pmatrix}\sigma_{rr}\\ \sigma_{\theta\theta}\\ \sigma_{r\theta}\end{pmatrix}=\frac{K_I}{\sqrt{2\pi r}}\begin{pmatrix}\frac{5}{4}\cos\left(\frac{\theta}{2}\right)-\frac{1}{4}\cos\left(\frac{3\theta}{2}\right)\\ \frac{3}{4}\cos\left(\frac{\theta}{2}\right)+\frac{1}{4}\cos\left(\frac{3\theta}{2}\right)\\ \frac{1}{4}\sin\left(\frac{\theta}{2}\right)+\frac{1}{4}\sin\left(\frac{3\theta}{2}\right)\end{pmatrix}-\frac{K_{II}}{\sqrt{2\pi r}}\begin{pmatrix}-\frac{5}{4}\sin\left(\frac{\theta}{2}\right)+\frac{3}{4}\sin\left(\frac{3\theta}{2}\right)\\ -\frac{3}{4}\sin\left(\frac{\theta}{2}\right)-\frac{3}{4}\sin\left(\frac{3\theta}{2}\right)\\ \frac{1}{4}\cos\left(\frac{\theta}{2}\right)+\frac{3}{4}\cos\left(\frac{3\theta}{2}\right)\end{pmatrix},\qquad(1)$$

where σ_{rr}, $\sigma_{\theta\theta}$, and $\sigma_{r\theta}$ are the components of the stress tensor in the polar coordinates, and K_I and K_{II}, respectively, are the elastic stress intensity factors in mode I and mode II in the local coordinates. Note that the sign of K_{II} is opposite to the conventional definition. The normal and shear components of traction near the edge of contact (inside the contact zone) are obtained as

$$p_{eoc}=\frac{K_I}{\sqrt{2\pi r}}\quad\text{and}\quad q_{eoc}=-\frac{K_{II}}{\sqrt{2\pi r}},\qquad(2)$$

by setting $\theta=0$ for $\sigma_{\theta\theta}$ and $\sigma_{r\theta}$ in Eq. (1), respectively. As evident by the fretting scars in experiments, slip between the contact surfaces occurs at first at the edge of contact. In addition, it is assumed that the crack initiation takes place at the left edge for a positive Q ([23]) and a Coulomb-type friction law may describe the interaction of the fretting surfaces, i.e., $q_{eoc}/p_{eoc}=f_{sp}$. Thus, the relationship between K_I and K_{II} is established by the coefficient of friction near the edge of contact, f_{sp}, as

$$K_{II}/(-K_I)=f_{sp}.\qquad(3)$$

If gross slip between the substrate and pad occurs, the same relationship as in Eq. (3) can be derived for K_I and K_{II} ([19]). Note that K_I in the present case is negative, which is uncommon in a real crack problem in which the interpenetration of crack surfaces is prohibited. From the above stress field, crack initiation from the edge of contact may be evaluated if a criterion reflecting fracture properties of the substrate is given. The criterion is discussed next.

2.2 Crack Extension Criteria. The problem of crack growth under mixed-mode loading has been under investigation for a few decades. Several criteria for crack growth under combined K_I and K_{II} have been proposed mainly for brittle materials. The most widely applied criteria are the maximum tangential stress (MTS) criterion ([24]), the maximum energy release rate (MERR) criterion ([25,26]), and the minimum strain energy density criterion (S-criterion) ([27]). Within the LEFM framework, the first two criteria are basically the same for prediction of pure opening crack growth as well as in physical meaning ([28,29]), while the third one seems to lack a physical support ([25,30,31]). Otsuka et al. [32] proposed the criterion of maximum shear stress (MSS) based on the experimental observation of crack extension in the maximum shear plane. Shen [33] also observed the shear-mode cracks during the crack coalescence in gypsum. Shen and Stephansson [34] proposed a modified F-criterion for mixed mode crack growth. The MTS and MSS criteria will be used in the following analyses of fretting crack initiation because of their clear physical background. Also these two criteria, respectively, dictate the two bounds of crack-extension mode mixity. They are summarized below.

Consider initiation of a crack into the substrate from the edge of contact at angle θ, subjected to remote K_I and K_{II}, as shown in Fig. 1. The local (effective) driving forces for the crack initiation, k_I and k_{II}, are expressed in terms of θ as

$$\left.\begin{aligned} k_I(\theta) &= \sigma_{\theta\theta}(\theta)\sqrt{2\pi r} \\ k_{II}(\theta) &= -\sigma_{r\theta}(\theta)\sqrt{2\pi r} \end{aligned}\right\}, \tag{4}$$

where $\sigma_{\theta\theta}$ and $\sigma_{r\theta}$ are given in Eq. (1). The MTS and MSS criteria have been proposed by correlating the local stress intensity factors k_I and k_{II} to relevant materials properties under appropriate physical considerations. The MTS criterion states that the crack subjected to K_I and K_{II} tends to extend in the direction of θ_I in which k_I achieves the maximum value and succeeds to do so if the maximum value of k_I is above the threshold. This angle, θ_I, for the maximum value of k_I is obtained as ([24,31])

$$\theta_I = 2\tan^{-1}\left(\frac{\sqrt{K_I^2+8K_{II}^2}-K_I}{4K_{II}}\right), \tag{5}$$

by solving the following equations:

$$\frac{\partial k_I}{\partial\theta}=0 \quad \text{and} \quad \frac{\partial^2 k_I}{\partial\theta^2}<0. \tag{6}$$

It should be noted that $k_{II}=0$ at $\theta=\theta_I$. However, that $k_{II}=0$ itself is insufficient condition to derive the expression of Eq. (5) as considered in the analysis of the crack analogue model by Giannakopoulos et al. [19].

Similarly, the MSS criterion states that a crack subjected to K_I and K_{II} tends to extend in the direction of θ_{II} in which $|k_{II}|$ achieves the maximum value and succeeds to do so if the maximum value of $|k_{II}|$ is above the threshold. This angle, θ_{II}, for the maximum value of $|k_{II}|$ satisfies the following equation:

$$\frac{\partial k_{II}}{\partial\theta}(\theta_{II})=0. \tag{7}$$

By substituting the stress component $\sigma_{r\theta}$ given in Eq. (1) for k_{II}, it is rewritten as

$$2K_{II}\tan^3\left(\frac{\theta_{II}}{2}\right)+2K_I\tan^2\left(\frac{\theta_{II}}{2}\right)-7K_{II}\tan\left(\frac{\theta_{II}}{2}\right)-K_I=0. \tag{8}$$

Unlike applying the inequality of Eq. (6) for θ_I in the MTS case, a careful selection from the real roots of Eq. (8) for θ_{II} is required in this case.

The crack analogue model using the rigid flat-ended contact configuration provides the asymptotic elastic stress field at the edge of contact and the driving forces for crack initiation in fretting fatigue in terms of stress intensity factors. This simple model has distinct advantage that it has only one controlling parameter, i.e., the coefficient of friction at the edge of contact. By examining the model predictions by the crack analogue model in conjunction with the MTS and MSS criteria within the framework of LEFM and by comparing them to experimental observations, the crack initiation mechanisms in fretting fatigue are explored, as discussed next.

3 Analyses and Discussion

In fretting fatigue, stress in the substrate is often highly concentrated at the edge of contact. The crack analogue model, as described above, may be appropriate to handle the situation and characterize the driving forces for fretting fatigue crack initiation. However, there are a few length scales in the analysis of fretting fatigue crack initiation which should be kept in mind. These are width of the pad, size of the partial slip region, size of the stress concentration zone, and size of the fracture process zone ahead of a single crack tip. The stress concentration zone may be considered as equivalent to the zone of plastic deformation or as the zone of distributed damage, since the stress concentration is not so severe as the effect of a crack tip. In order to apply the crack analogue model, the last two length scales (i.e., the stress concentration and the fracture process zone) must be sufficiently small relative to the width of the pad. If the size of the stress concentration zone is comparable to the size of the fracture process zone associated with a single crack, the overall crack initiation toughness is well defined, and the initiation of fretting fatigue crack may be predicted by using the crack analogue model. In the case where initiation of multiple microcracks occurs, the crack analogue model represents the initiation of a dominant crack outrunning all other cracks, for which the fracture driving force attains the maximum value. However, when multiple cracks initiate (with no dominant crack), the total initiation toughness for these cracks is ambiguous. Further, this problem is of the statistical nature and more complex, and the present crack analogue model is not applicable to such cases.

The purpose of the present work is to examine the angle of crack initiation for both modes using the crack analogue model and to analyze the evolving damage mechanisms in fretting fatigue by correlating the predictions to experimental observations. Note that the coefficient of friction at the edge of contact, f_{sp}, which uniquely characterizes the asymptotic stress field under the condition of slip as described in Eqs. (1) and (3), serves as the single controlling parameter in the crack initiation process. Due to the fact that fretting fatigue cracks normally initiate on the ''tensile'' side of contact ([23]), only the case of $Q>0$ is considered below.

3.1 Model Predictions. Angular variations of effective stress intensity factors, k_I and k_{II}, at the edge of contact for a few representative values of f_{sp} were calculated by using Eqs. (1) and (4), and these are plotted in Fig. 2. Angles for the maximum values of k_I and $|k_{II}|$, i.e., θ_I and θ_{II}, were evaluated at different values of f_{sp} by using Eqs. (5) and (8), and these results for f_{sp} in the range from 0 to 3 are plotted in Fig. 3. Effective k_I at $\theta=\theta_I$ and $|k_{II}|$ at $\theta=\theta_{II}$, and their ratio were subsequently calculated. The ratio is plotted as a function of f_{sp} in Fig. 4. Recall that k_I at $\theta=\theta_I$ represents the driving force for opening-mode crack initiation predicted by the MTS criterion, and that $|k_{II}|$ at $\theta=\theta_{II}$ represents the driving force for shear-mode crack initiation predicted by the MSS criterion.

Figure 2 shows the typical angular variations by k_I demonstrating the existence of a peak value of k_I at a certain angle for a given f_{sp}. Furthermore, Fig. 3 demonstrates that the angle for peak k_I, i.e., θ_I, varies monotonically from 180 deg to 70.5 deg as f_{sp} varies from 0 to infinity. Meanwhile, the angle for peak (absolute) shear stress, θ_{II}, varies monotonically from 70.5 deg to 0 deg. The ratio of effective driving forces for opening-mode and shear-mode crack initiations, $k_I(\theta_I)$ to $k_{II}(\theta_{II})$, is equal to zero at $f_{sp}=0$, and increases with increasing f_{sp}. It reaches the maximum value of 1.155 at $f_{sp}=\infty$, as shown in Fig. 4. Since it indicates the competition between the driving forces for crack initiation in these two modes, this figure demonstrates that the driving force is in favor of the shear-mode crack initiation at small values of f_{sp} and alters to favor the opening-mode crack initiation at large values of f_{sp}. The critical value of f_{sp} (which evolves during fretting fatigue) for switching the modes should depend on materials. Note that at $\theta=\theta_{II}$, k_I is always negative, indicating a closed crack if initiated in shear mode.

3.2 Predictions Versus Observations in Fretting Fatigue Tests. As mentioned before, the crack initiation angle observed in fretting fatigue tests ranges widely from 25 deg to 80 deg ([3,6–14]). While these tests can be divided into stages as characterized by f_{sp} between the contact surfaces, at which stage the cracks were initiated is uncertain. In these tests, the initial value of f_{sp} (if reported) was normally ~0.2–0.4, with a smooth initial surface. It evolved with cycling, and stabilized at a value of ~0.7–1.2. The steady-state values, when used as the local f_{sp} at

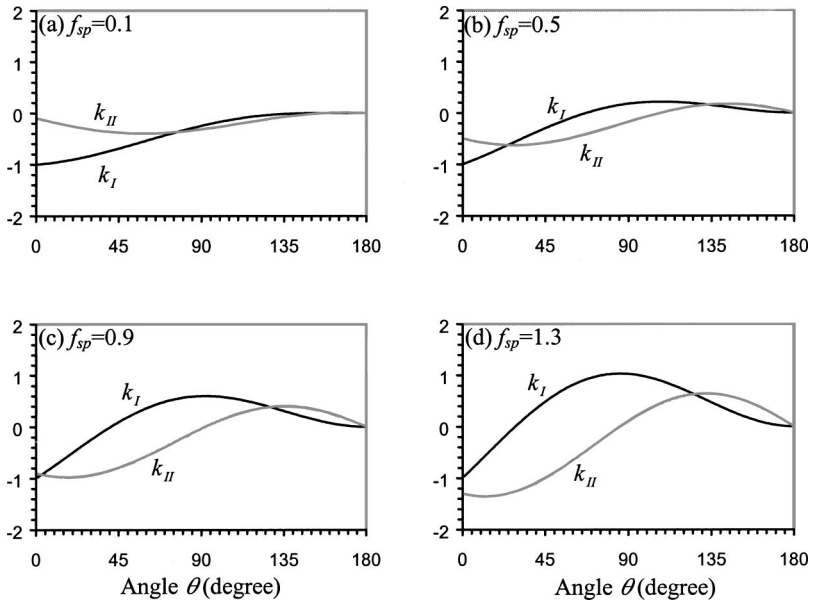


Fig. 2 Variation of effective stress intensity factors k_I and k_{II} with angle θ, for different values of f_{sp}. The values of k_I and k_{II} are normalized by $-K_I$.

the edge of contact, are conservative because they were measured as the average over the entire contact surface normally under a partial slip condition. The cyclic slippery process had led to rougher surface of contact and a larger f_{sp} especially near the edge ([21]).

If the coefficient of friction is given, the crack initiation modes in the early and in the stabilized stages in these tests can be characterized by using the crack analogue model. For example if $f_{sp}=0.2$, the MSS criterion predicts the crack initiation angle in shear mode to be 33.4 deg in the early stage of fretting fatigue, which is in agreement with the lower bound of the angles observed experimentally. Meanwhile, the MTS criterion predicts the crack initiation angle in opening mode to be 139.2 deg, which is much above the range of the observed angles, as shown in Fig. 3. In addition, the driving force appears to favor shear-mode crack initiation rather than opening-mode crack initiation in this case with a small value of f_{sp}. The ratio of the driving forces for opening-mode and shear-mode crack initiations is equal to 0.06 at $f_{sp}=0.2$, as shown in Fig. 4.

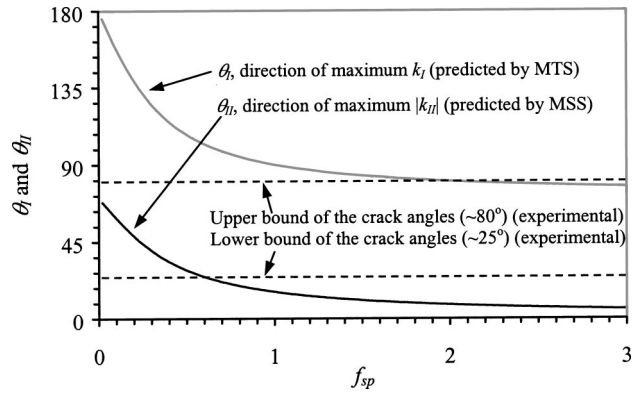


Fig. 3 Variation of θ_I for maximum k_I and of θ_{II} for maximum $|k_{II}|$ with f_{sp}, under the condition of slip at the edge of contact, predicted by the MTS criterion and by the MSS criterion, respectively. Note that θ_I and θ_{II}, respectively, reach their minimum values of 70.5 deg and of 0 deg at $f_{sp}=\infty$. The upper and lower bounds of the crack angles observed in tests are also shown for comparison with the predictions.

In the later stabilized stage of fretting fatigue, for example with $f_{sp}=1.5$, the MTS criterion predicts the crack initiation angle in opening mode to be 83.5 deg, which is in agreement with the upper bound of the angles observed experimentally. Meanwhile, the MSS criterion predicts the angle in shear mode to be 10.5 deg, which is too small compared to the experimental values. However, the small-angle shear-damage mechanism might be responsible for wear detachments forming slivers instead of fretting fatigue crack initiation. In addition, the driving force appears to favor opening-mode crack initiation in this case with a large value of f_{sp}. The ratio of the driving forces for opening-mode and shear-mode crack initiations is equal to 0.81 at $f_{sp}=1.5$, as shown in Fig. 4. Based on these observations, it is suggested that the crack initiation process was shear-mode dominant in the early stage with a smooth contact surface and was opening-mode dominant in the later steady-state stage with a rough worn contact surface in the fretting fatigue tests discussed above.

It should be mentioned that the foregoing analyses are based on the crack analogue model without an actual crack in the substrate. It is equivalent to the first-order perturbation analysis of a kinked crack with traction-free surfaces ([35]). However, it may not al-

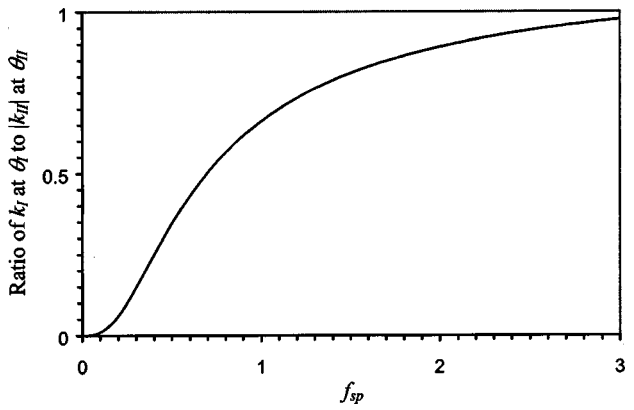


Fig. 4 Ratio of k_I at θ_I to $|k_{II}|$ at θ_{II} as a function of f_{sp}, under the condition of slip at the edge of contact. The ratio indicates the competition between the driving forces for opening-mode and shear-mode initiation of a crack.

ways be the case that the crack surfaces are traction free since the crack may initiate in shear mode and be closed, as mentioned before. If the crack contact happens, the simple perturbation analysis is virtually impossible due to unknown contact interaction, and a numerical perturbation analysis is necessary. Also, effects of the applied bulk stress on the substrate are neglected in the above analysis. This is reasonable since the bulk stress, analogous to the elastic *T*-stress ([36]), exercises no influence on the stress intensity factors K_I and K_{II} in the unperturbed configuration, and hence no influence on the stress intensity factors for the perturbed infinitesimal crack ([19]). For the crack growth beyond initiation, the role of the bulk stress may become important. For example, it may be responsible for the knee-shape kink and reorientation of the crack path as observed in the fretting fatigue tests.

Finally, this study explicitly demonstrates that increasing the amount of friction would cause a significant change in the way the fretting damage would be initiated. This is indirectly evident from the wide variation of the crack initiation angle observed in the fretting literature, as mentioned earlier. Also, the classical critical plane fatigue criteria, based on the noncrack approach, implicitly indicate that changing the level of friction changes the way fretting cracks are initiated. Further, the present crack analogue model has the distinct advantage over the classical critical plane fatigue criteria for scenarios with geometric discontinuities and involving very high stress concentration, since the latter rely heavily on the accurate determination of the local stress and strain field. Due to severe stress and strain gradients in the fretting situation, determination of an accurate stress/strain field is quite a challenging task. On the contrary, the present crack analogue model has an edge over them by capturing the dominant part of the stress field, and thus by providing an efficient and accurate evaluation of the magnitude of driving force for crack initiation as well as its orientation angle.

4 Conclusions

Crack initiation in fretting fatigue is analyzed by using the crack analogue model. This approach provides the asymptotic elastic stress field at the edge of contact and the driving forces in the form of stress intensity factors for crack initiation. Since it uniquely characterizes the asymptotic stress field, the local coefficient of friction between the substrate and pad is considered as the primary parameter to characterize the crack initiation process under the condition of slip near the edge of contact. The MTS and MSS criteria are applied to predict the crack initiation directions in opening mode and in shear mode, respectively. By examining the model prediction and by comparing it with the experimental observation of crack initiation angles (the upper and lower bounds), it is shown that the observed microcracks at the small angles were nucleated in shear mode in the early stage of tests with a smooth contact surface (i.e., a small coefficient of friction), while the observed microcracks at the large angles were nucleated in opening mode in the later stage with a rough worn surface (i.e., a large coefficient of friction). This understanding may help to establish the sequential damage mechanisms in the complex process of fretting fatigue.

Acknowledgments

The authors gratefully acknowledge the support of U.S. Air Force Office of Scientific Research, and U.S. Air Force National Turbine Engine High Cycle Fatigue (HCF) program.

References

[1] Dobromirski, J. M., 1992, ''Variables of Fretting Process: Are There 50 of Them?'' *Standardization of Fretting Fatigue Test Methods and Equipment*, ASTM STP 1159, M. Helmi Attia and R. B. Waterhouse, eds., American Society for Testing and Materials, Philadelphia, pp. 60–66.

[2] Nishioka, K., and Hirakawa, K., 1969, ''Fundamental Investigations of Fretting Fatigue: Part 2,'' Bull. JSME, **12**, pp. 189–287.

[3] Antoniou, R. A., and Radtke, T. C., 1997, ''Mechanisms of Fretting-Fatigue of Titanium Alloys,'' Mater. Sci. Eng., **A237**, pp. 229–240.

[4] Waterhouse, R. B., and Taylor, D. E., 1971, ''The Initiation of Fatigue Cracks in a 0.7% Carbon Steel by Fretting,'' Wear, **17**, pp. 139–147.

[5] Lamacq, V., Dubourg, M.-C., and Vincent, L., 1996, ''Crack Path Prediction Under Fretting Fatigue—A Theoretical and Experimental Approach,'' ASME J. Tribol., **118**, pp. 711–720.

[6] Wharton, M. H., Taylor, D. E., and Waterhouse, R. B., 1973, ''Metallurgical Factors in the Fretting-Fatigue Behavior of 70/30 Brass and 0.7% Carbon Steel,'' Wear, **23**, pp. 251–260.

[7] Endo, K., and Goto, H., 1976, ''Initiation and Propagation of Fretting Fatigue Cracks,'' Wear, **38**, pp. 311–324.

[8] Tanaka, K., Mutoh, Y., Sakoda, S., and Leadbeater, G., 1985, ''Fretting Fatigue in 0.55c Spring Steel and 0.45c Carbon Steel,'' Fatigue Fract. Eng. Mater. Struct., **8**, pp. 129–142.

[9] Nix, K. J., and Lindley, T. C., 1985, ''The Application of Fracture Mechanics to Fretting Fatigue,'' Fatigue Fract. Eng. Mater. Struct., **8**, pp. 143–160.

[10] Sato, K., Fujii, H., and Kodama, S., 1986, ''Crack Propagation Behavior in Fretting Fatigue,'' Wear, **107**, pp. 245–262.

[11] Hills, D. A., Nowell, D., and O'Connor, J. J., 1988, ''On the Mechanics of Fretting Fatigue,'' Wear, **125**, pp. 129–146.

[12] Faanes, S., 1995, ''Inclined Cracks in Fretting Fatigue,'' Eng. Fract. Mech., **52**, pp. 71–82.

[13] Lamacq, V., Dubourg, M.-C., and Vincent, L., 1997, ''A Theoretical Model for the Prediction of Initial Growth Angles and Sites of Fretting Fatigue Cracks,'' Tribol. Int., **30**, pp. 391–400.

[14] Kim, Hyung-Kyu, and Lee, Soo-Bok, 1997, ''Crack Initiation and Growth Behavior of Al 2024-T4 Under Fretting Fatigue,'' Int. J. Fatigue, **19**, pp. 243–251.

[15] Ruiz, C., Boddington, P. H. B., and Chen, K. C., 1984, ''An Investigation of Fatigue and Fretting in a Dovetail Joint,'' Exp. Mech., **24**, pp. 208–217.

[16] Cheng, W., Cheng, H. S., Mura, T., and Keer, L. M., 1994, ''Micromechanics Modeling of Crack Initiation Under Contact Fatigue,'' ASME J. Tribol., **116**, pp. 2–8.

[17] Szolwinski, Matthew P., and Farris, Thomas N., 1996, ''Mechanics of Fretting Fatigue Crack Formulation,'' Wear, **198**, pp. 93–107.

[18] Fellows, L. J., Nowell, D., and Hills, D. A., 1997, ''On the Initiation of Fretting Fatigue Cracks,'' Wear, **205**, pp. 120–129.

[19] Giannakopoulos, A. E., Lindley, T. C., and Suresh, S., 1998, ''Aspects of Equivalence Between Contact Mechanics and Fracture Mechanics: Theoretical Connections and Life-Prediction Methodology for Fretting-Fatigue,'' Acta Mater., **46**, pp. 2955–2968.

[20] Giannakopoulos, A. E., Venkatesh, T. A., Lindley, T. C., and Suresh, S., 1999, ''The Role of Adhesion in Contact Fatigue,'' Acta Mater., **47**, pp. 4653–4664.

[21] Hills, D. A., Nowell, D., and Sackfield, A., 1993, *Mechanics of Elastic Contacts*, Butterworth-Heinemann Ltd., Oxford, UK.

[22] Williams, M. L., 1957, ''On the Stress Distribution at the Base of a Stationary Crack,'' ASME J. Appl. Mech., **24**, pp. 109–114.

[23] Nowell, D., and Hills, D. A., 1987, ''Mechanics of Fretting Fatigue Tests,'' Int. J. Mech. Sci., **29**, pp. 355–365.

[24] Erdogan, F., and Sih, G. C., 1963, ''On the Crack Extension in Plates Under Plane Loading and Transverse Shear,'' J. Basic Eng., **91**, pp. 764–769.

[25] Wu, C. H., 1978, ''Fracture Under Combined Loads by Maximum Energy Release Rate Criterion,'' ASME J. Appl. Mech., **45**, pp. 553–558.

[26] Palaniswamy, K., and Knauss, W. G., 1978, ''On the Problem of Crack Extension in Brittle Solids Under General Loading,'' *Mechanics Today*, S. Nemat-Nasser, ed., Pergamon, New York, pp. 87–148.

[27] Sih, G. C., 1974, ''Strain-Energy-Density Factor Applied to Mixed Mode Crack Problems,'' Int. J. Fract., **10**, pp. 305–321.

[28] Nuismer, R. J., 1975, ''An Energy Release Rate Criterion for Mixed Mode Fracture,'' Int. J. Fract., **11**, pp. 245–250.

[29] Shen, Minsheng, and Shen, M.-H. Herman, 1995, ''Direction of Crack Extension Under General Plane Loading,'' Int. J. Fract., **70**, pp. 51–58.

[30] Theocaris, P. S., and Andrianopoulos, N. P., 1984, author's closure on the discussion by G. C. Sih and E. E. Gdoutos of ''The Mises Elastic-Plastic Boundary as the Core Region in Fracture Criteria,'' Eng. Fract. Mech., **20**, pp. 691–694.

[31] Qian, J., and Fatemi, A., 1996, ''Mixed Mode Fatigue Crack Growth: A Literature Survey,'' Eng. Fract. Mech., **55**, pp. 969–990.

[32] Otsuka, A., Mori, K., and Miyata, T., 1975, ''The Condition of Fatigue Crack Growth in Mixed Mode Condition,'' Eng. Fract. Mech., **7**, pp. 429–439.

[33] Shen, B., 1995, ''The Mechanism of Fracture Coalescence in Compression—Experimental Study and Numerical Simulation,'' Eng. Fract. Mech., **51**, pp. 73–85.

[34] Shen, B., and Stephansson, O., 1994, ''Modification of the G-Criterion for Crack Propagation Subjected to Compression,'' Eng. Fract. Mech., **47**, pp. 177–189.

[35] Cotterell, B., and Rice, J. R., 1980, ''On a Slightly Curved or Kinked Crack,'' Int. J. Fract., **16**, pp. 155–169.

[36] Rice, J. R., 1974, ''Limitations to the Small-Scale Yielding Approximation for Crack-Tip Plasticity,'' J. Mech. Phys. Solids, **22**, pp. 17–26.

An Intersonic Slip Pulse at a Frictional Interface Between Dissimilar Materials

G. G. Adams
Professor, Fellow ASME,
Department of Mechanical Engineering,
Northeastern University,
Boston, MA 02115
e-mail: adams@neu.edu

Two homogeneous and isotropic elastic half-spaces are acted upon by remote normal and shear tractions. The applied shear stress is less than that which is required to produce overall sliding of the two bodies. The possible existence of a slip pulse is investigated, i.e., a finite-width region, on the interface, of altered normal and shear stress which satisfies the Amontons-Coulomb law of friction. Pulses which travel at a speed which is greater than the minimum shear wave speed and less than the maximum dilatational wave of the two bodies, are of interest in this investigation. Such pulses are shown to exist for sufficient friction and for modest mismatches in material combinations. The pulse is weakly singular at the leading edge and bounded at the trailing edge. Furthermore it travels at speeds just below the lesser dilatational wave speed and in the opposite direction of sliding of the lower wave-speed material. In addition, a pair of equations are given which relate the interfacial normal and shear stress to the interfacial slip velocity. These relations are analogous to the subsonic results of Weertman, but are valid for an arbitrary speed range. [DOI: 10.1115/1.1349119]

1 Introduction

The interaction of elastic waves with friction has been the subject of many recent investigations. These problems have relevance in the areas of tribology and seismology and thus span a range in scale from nanometers to hundreds of kilometers.

It is well known that a Rayleigh wave can propagate along the free surface of a semi-infinite elastic body and has an amplitude which decays exponentially with distance from the free surface. Similar waves can travel along the interface of two contacting elastic bodies. Such waves were investigated by Stoneley [1] for bonded contact and are known as Stoneley waves. Stoneley waves exist only if the material properties of the two bodies do not differ greatly. Achenbach and Epstein [2] investigated interface waves in unbonded frictionless contact in which separation does not occur. These ''smooth contact Stoneley waves'' (also known as slip waves or generalized Rayleigh waves) are qualitatively similar to those for bonded contact and occur for a somewhat wider range of material combinations. Comninou and Dundurs [3] investigated slip waves with periodic regions of separation along a frictionless interface. The possibility of two identical half-spaces sliding with friction due to the presence of separation waves and/or stick-slip waves was studied by Comninou and Dundurs [4]. Both of these analyses showed that such waves could exist only with square-root singularities at the tips of the slip zones. Freund [5] pointed out that the singularities encountered by Comninou and Dundurs would require energy sources and sinks.

The *frictional* sliding of an elastic half-space against a rigid surface (Martins, Guimarães, and Faria, [6]) and of two elastic half-spaces (Adams, [7]) have also been investigated. Friction can cause surface waves (similar to slip waves) to grow with time; the rate of growth is inversely proportional to the wavelength. In simulations, this phenomenon can lead to numerical problems which are related to mesh size. Recent work by Ranjith and Rice [8] used the Prakash-Clifton [9,10] friction model. In that model there is no instantaneous dependence of shear stress on normal stress, but rather the shear stress depends on a simple fading memory of the prior history of normal stress. The use of this friction law removes the short wavelength ill-posedness of frictional sliding.

The notion that certain observed friction behavior is not a property of the interface, but rather a consequence of system dynamics, was suggested by Martins, Oden, and Simões [11]. Adams [12] investigated the sliding of two dissimilar elastic bodies due to periodic regions of slip and stick propagating along the interface. It was found that such motion allows for the interface sliding conditions to differ from the observed sliding conditions. In particular the *interface* coefficient of friction (defined as the ratio of shear stress to contact pressure at the interface) can be constant or an increasing/decreasing function of slip velocity. However, the *apparent* coefficient of friction (defined as the ratio of the *applied* shear stress to the *applied* normal stress) will be less than the interface friction coefficient. Furthermore the apparent coefficient of friction can decrease with sliding speed even though the interface friction coefficient is constant. Thus the measured coefficient of friction does not necessarily represent the behavior of the sliding interface.

In the limit as the slip region becomes very small compared to the stick region, the results of Adams [12] become that of a slip pulse traveling through a region which otherwise sticks. Rice [13] derived that result by using the moving dislocation formulation of Weertman [14]. The existence of such an isolated slip pulse was postulated by Weertman [15]. Andrews and Ben-Zion [16] obtained a numerical solution for a slip pulse, the amplitude of which increases and the width of which decreases as the pulse continues to propagate. This self-sharpening effect is consistent with the analytical solution of Adams [7] for sliding. Recently Caroli [17] investigated the interface between a viscoelastic material and a rigid surface. It was shown that a periodic set of slip-pulses is impossible for this viscoelastic-rigid interface which satisfies Coulomb's law of friction. However, experimental evidence, described in ([17]), indicates the existence of slip-pulses and so ([17]) discusses possible improvements which could be made to the friction model.

Adams [18] and Nosonovsky and Adams [19] investigated the sliding of elastic half-spaces. They showed that steady sliding is compatible with the formation of pairs of body waves (a plane

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, March 28, 2000; final revision, August 15, 2000. Associate Editor: R. C. Benson. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

dilatational wave and a plane shear wave) radiated from the sliding interface. Each wave moves at a different angle with respect to the interface such that the trace velocities along the interface are equal and supersonic with respect to both elastic media. This supersonicity does not violate causality as it is only the trace velocity which is supersonic; the waves move at the dilatational and shear wave speeds in their respective bodies. It was also shown that a rectangular wave train, or a rectangular pulse, can allow for motion of the two bodies with a ratio of remote shear to normal stress which is less than the ratio of shear to normal stress required to produce sliding at the interface. Thus the apparent coefficient of friction is less than the interface coefficient of friction. Furthermore the apparent friction coefficient decreases with increasing speed even if the interface friction coefficient is speed-independent. This result, as well as ([12]), supports the interpretation of certain friction behavior as a consequence of the dynamics of the system, rather than as strictly an interface property ([11]).

A slip-pulse, at the interface between two elastic half-spaces, traveling at an *intersonic* speed (between the lower shear wave speed and the higher dilatational wave speed) was found by Cochard and Rice [20] using a numerical technique and the Prakash-Clifton friction law. For the particular material combination studied, the pulse traveled at a speed just below the slower dilatational wave-speed and in the opposite direction of sliding of the slower wave-speed material. In the present investigation, an analytical solution is obtained for an intersonic slip pulse with Coulomb friction. The problem is first formulated in terms of a periodic set of slip-pulses. Then, through an appropriate limiting process, a singular integral equation is obtained for a single pulse. This procedure allows the interface normal and shear stresses to be related to the slip velocity (a generalization of the subsonic Weertman solution). An intersonic slip pulse is shown to exist for moderate values of friction and for modest differences in the material properties. The behavior of these pulses is then studied in detail.

2 Problem Description

Consider two perfectly flat homogeneous and isotropic elastic half-spaces pressed against each other and sheared with remote tractions p^* and q^*, respectively (Fig 1). The ratio

$$\mu^* \equiv q^*/p^* \tag{1}$$

is less than that required to cause the two bodies to slide with respect to each other, i.e., $\mu^* < \mu$ where μ is the interface coefficient of friction. It is emphasized that the *interface* coefficient of friction μ is the ratio of shear to normal contact pressure at the *interface* which would cause local slipping to occur. Clearly, under these loading conditions, it is possible for the two bodies to remain in static equilibrium. However, we investigate here the possibility of relative motion of the two bodies with $\mu^* < \mu$, due to the existence of periodic stick and slip regions which propagate

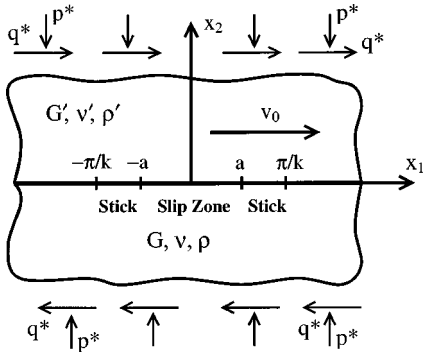


Fig. 1 A periodic system of slip-pulses at the frictional interface between two elastic bodies

along the interface with some intersonic wave speed c. Thus under these conditions the quantity μ^* may be interpreted as the *apparent* coefficient of friction, since sliding occurs with that ratio of applied shear to normal tractions. Sufficiently far from the interface the bodies move with relative velocity v_0 in the x_1-direction. By taking an appropriate limit it will be shown that a solution for a single slip-pulse can be found for which $v_0 \to 0$.

We consider perturbation displacements $(\hat{u}_1, \hat{u}_2, \hat{u}_1', \hat{u}_2')$ in the form of real nondispersive traveling waves with wave number k as given by, e.g., Comninou and Dundurs [3]

$$\begin{aligned}
\hat{u}_1 &= \mathrm{Re}\left\{ \sum_{m=1}^{\infty} [D_{1m} \exp(mk\zeta_1 x_2) + D_{2m} \exp(mk\zeta_2 x_2)] \exp[imk(x_1 - ct)] \right\} \\
\hat{u}_2 &= \mathrm{Re}\left\{ \sum_{m=1}^{\infty} [-i\zeta_1 D_{1m} \exp(mk\zeta_1 x_2) - i\zeta_2^{-1} D_{2m} \exp(mk\zeta_2 x_2)] \exp[imk(x_1 - ct)] \right\} \\
\hat{u}_1' &= \mathrm{Re}\left\{ \sum_{m=1}^{\infty} [D_{3m} \exp(-mk\zeta_3 x_2) + D_{4m} \exp(-mk\zeta_4 x_2)] \exp[imk(x_1 - ct)] \right\} \\
\hat{u}_2' &= \mathrm{Re}\left\{ \sum_{m=1}^{\infty} [i\zeta_3 D_{3m} \exp(-mk\zeta_3 x_2) + i\zeta_4^{-1} D_{4m} \exp(-mk\zeta_4 x_2)] \exp[imk(x_1 - ct)] \right\}.
\end{aligned} \tag{2}$$

This perturbation is taken with respect to the homogeneous reference state of stress given by $\tau_{22} = -p^*$ and $\tau_{12} = q^*$. It is noted that this reference state does not satisfy the frictional sliding condition at the interface. For a given wave speed c, for which ζ_k is real

$$\begin{aligned}
&\zeta_1 = \sqrt{1-(c/c_1)^2}, \quad \zeta_2 = \sqrt{1-(c/c_2)^2}, \\
&\zeta_3 = \sqrt{1-(c/c_1')^2}, \quad \zeta_4 = \sqrt{1-(c/c_2')^2}
\end{aligned} \tag{3}$$

which requires that the magnitudes of the wave components decay as $|x_2| \to \infty$. Similarly for ζ_k imaginary

$$\begin{aligned}
&\zeta_1 = -\mathrm{sgn}(c)\, i\sqrt{(c/c_1)^2 - 1}, \quad \zeta_2 = -\mathrm{sgn}(c)\, i\sqrt{(c/c_2)^2 - 1}, \\
&\zeta_3 = -\mathrm{sgn}(c)\, i\sqrt{(c/c_1')^2 - 1}, \quad \zeta_4 = -\mathrm{sgn}(c)\, i\sqrt{(c/c_2')^2 - 1}
\end{aligned} \tag{4}$$

which constitute the radiation condition, i.e., waves cannot be generated at $|x_2| = \infty$. For intersonic speeds, c will be in a range for which some ζ_k are real and others are imaginary. Furthermore in (2)–(4)

$$c_1 = \sqrt{\frac{\lambda + 2G}{\rho}}, \quad c_2 = \sqrt{\frac{G}{\rho}}, \quad c_1' = \sqrt{\frac{\lambda' + 2G'}{\rho'}}, \quad c_2' = \sqrt{\frac{G'}{\rho'}} \tag{5}$$

$$\beta = \frac{c_1}{c_2} = \sqrt{\frac{2(1-\nu)}{1-2\nu}}, \quad \beta' = \frac{c_1'}{c_2'} = \sqrt{\frac{2(1-\nu')}{1-2\nu'}}, \quad \kappa = \frac{c_2'}{c_2}$$

where c_1, c_2 are the dilatational and shear wave speeds, respectively, G is the shear modulus, λ is the Lamé's constant, ν is the

Poisson's ratio, and ρ is the mass density. Quantities with a prime ($'$) refer to the upper half-space (Fig. 1). In addition to the wave number k, which defines the periodicity of the solution, there is the index m. Thus the desired solutions are periodic in space with wavelength $2\pi/k$ and consists of an infinite number of components, each with wave number mk.

The continuity conditions pertain to the complete solution, i.e., the sum of the uniform and perturbation solutions. Written in terms of a moving coordinate η, the shear and normal stresses and the normal displacements are continuous

$$\tau_{12}(\eta,0)=\tau'_{12}(\eta,0),\quad \tau_{22}(\eta,0)=\tau'_{22}(\eta,0),\quad u_2(\eta,0)=u'_2(\eta,0)$$
$$\eta=x_1-ct,\quad -\pi/k\leqslant\eta\leqslant\pi/k \tag{6}$$

where

$$\tau_{12}=G\left(\frac{\partial\hat{u}_1}{\partial x_2}+\frac{\partial\hat{u}_2}{\partial x_1}\right)+q^*,\quad \tau_{22}=(\lambda+2G)\frac{\partial\hat{u}_2}{\partial x_2}+\lambda\frac{\partial\hat{u}_1}{\partial x_1}-p^*. \tag{7}$$

Satisfaction of the continuity conditions yields

$$D_{1m}=e_1D_{4m},\quad D_{2m}=e_2D_{4m},\quad D_{3m}=e_3D_{4m},$$
$$m=1,2,3,\ldots \tag{8}$$

where e_1,e_2,e_3 are independent of m and are given by

$$e_1=(G'/G)[-((\beta')^2(1+\zeta_2^2)(-1+\zeta_3^2))+2(-1-\zeta_2^2-\zeta_2\zeta_3+\zeta_3\zeta_4+\zeta_2^2\zeta_3\zeta_4+\zeta_2\zeta_3\zeta_4^2)$$
$$+(G'/G)(2-4\zeta_3\zeta_4+2\zeta_4^2+(\beta')^2(-1+\zeta_3^2)(1+\zeta_4^2))]/\{[(2-4\zeta_1\zeta_2+2\zeta_2^2+\beta^2(-1+\zeta_1^2)(1+\zeta_2^2))\zeta_3$$
$$+(G'/G)(2(-2+\beta^2)\zeta_3-2\beta^2\zeta_1^2\zeta_3+\zeta_1(-2+(\beta')^2+2\zeta_2^2-(\beta')^2\zeta_2^2+4\zeta_2\zeta_3-(\beta')^2\zeta_3^2+(\beta')^2\zeta_2^2\zeta_3^2))]\zeta_4\} \tag{9}$$

$$e_2=-\{(G'/G)\zeta_2[\zeta_1(-4+2(\beta')^2-2(\beta')^2\zeta_3^2+4\zeta_3\zeta_4)-(-2+\beta^2)\zeta_3(-1+\zeta_4^2)+\beta^2\zeta_1^2\zeta_3(-1+\zeta_4^2)$$
$$+(G'/G)\zeta_1(2-4\zeta_3\zeta_4+2\zeta_4^2+(\beta')^2(-1+\zeta_3^2)(1+\zeta_4^2))]\}/\{[(2-4\zeta_1\zeta_2+2\zeta_2^2+\beta^2(-1+\zeta_1^2)(1+\zeta_2^2))\zeta_3$$
$$+(G'/G)(2(-2+\beta^2)\zeta_3-2\beta^2\zeta_1^2\zeta_3+\zeta_1(-2+(\beta')^2+2\zeta_2^2-(\beta')^2\zeta_2^2+4\zeta_2\zeta_3-(\beta')^2\zeta_3^2+(\beta')^2\zeta_2^2\zeta_3^2))]\zeta_4\} \tag{10}$$

$$e_3=\{-2+4\zeta_1\zeta_2-2\zeta_2^2-\beta^2(-1+\zeta_1^2)(1+\zeta_2^2)+(G'/G)[\beta^2(-1+\zeta_1^2)(1+\zeta_4^2)-2(-1+\zeta_1\zeta_2-\zeta_1\zeta_4+\zeta_1\zeta_2^2\zeta_4-\zeta_4^2+\zeta_1\zeta_2\zeta_4^2)]\}/$$
$$\{[(2-4\zeta_1\zeta_2+2\zeta_2^2+\beta^2(-1+\zeta_1^2)(1+\zeta_2^2))\zeta_3+(G'/G)(2(-2+\beta^2)\zeta_3-2\beta^2\zeta_1^2\zeta_3+\zeta_1(-2+(\beta')^2$$
$$+2\zeta_2^2-(\beta')^2\zeta_2^2+4\zeta_2\zeta_3-(\beta')^2\zeta_3^2+(\beta')^2\zeta_2^2\zeta_3^2))]\zeta_4\}. \tag{11}$$

The computations leading to (9)–(11) are quite complicated and were performed using the Mathematica symbolic manipulation software (Wolfram [21]). Note that if c is subsonic then e_1,e_2,e_3 are real. However, if c is intersonic or supersonic ($|c|>\mathrm{Min}(c_2,c_2')$) then e_1,e_2,e_3 are complex.

Now the mixed conditions which pertain to the periodic regions of stick and slip, i.e.,

$$v_S=\dot{\hat{u}}'_1(\eta,0)-\dot{\hat{u}}_1(\eta,0)+v_0=0,\quad a<|\eta|<\pi/k,\quad \text{Stick Region} \tag{12}$$

$$\tau_S=\tau_{12}(\eta,0)+\mu\tau_{22}(\eta,0)=0,\quad |\eta|<a,\quad \text{Slip Region} \tag{13}$$

are applied. The quantity v_S, defined in (12), is called the *slip velocity*. By substituting (2) into (12)

$$v_S=v_0+ck\,\mathrm{Re}\left\{e_0 i\sum_{m=1}^{\infty}mD_{4m}\exp(imk\eta)\right\},$$
$$e_0=e_1+e_2-e_3-1 \tag{14}$$

is obtained. By using the integral transformation given by

$$D_{4m}=\frac{1}{ce_0mi\pi}\int_{-a}^{a}\phi(\xi)\exp(-imk\xi)d\xi,\quad m=1,2,3,\ldots,\infty \tag{15}$$

the slip velocity becomes

$$v_S=v_0+\frac{k}{\pi}\mathrm{Re}\left\{\int_{-a}^{a}\sum_{m=1}^{\infty}\exp[imk(\eta-\xi)]\phi(\xi)d\xi\right\}. \tag{16}$$

Using the identities ([22])

$$\sum_{n=1}^{\infty}\cos nx=-\frac{1}{2}+\pi\sum_{n=-\infty}^{\infty}\delta(x-2\pi n),\quad \sum_{n=1}^{\infty}\sin nx=\frac{1}{2}\cot\frac{x}{2} \tag{17}$$

(16) results in

$$v_S=v_0-\frac{k}{2\pi}\int_{-a}^{a}\phi(\xi)d\xi+\phi(\eta)H(a-|\eta|)$$
$$+\mathrm{Re}\left\{\frac{ik}{2\pi}\int_{-a}^{a}\cot\frac{k(\eta-\xi)}{2}\phi(\xi)d\xi\right\}. \tag{18}$$

It can now be seen that (18) will satisfy the stick condition (12) provided that $\phi(\xi)$ is real and that the resultant condition

$$\frac{k}{2\pi}\int_{-a}^{a}\phi(\xi)d\xi=v_0 \tag{19}$$

is satisfied. Thus the slip velocity becomes

$$v_S=\phi(\eta)H(a-|\eta|) \tag{20}$$

and the slip velocity automatically vanishes in the stick regions. Furthermore $\phi(\eta)$ is seen to be equal to v_S in the slip region.

It remains to satisfy the slip stress condition (13). By substituting (2) into (13)

$$\tau_S=Gk\,\mathrm{Re}\left\{\sum_{m=1}^{\infty}(\delta_1-i\mu\delta_2)mD_{4m}\exp(imk\eta)\right\}+q^*-\mu p^* \tag{21}$$

where

$$\delta_1=2e_1\zeta_1+e_2(\zeta_1+\zeta_2^{-1}),\quad \delta_2=2e_1+\beta^2\zeta_1^2e_1+2e_2-\beta^2e_1 \tag{22}$$

is obtained, in which δ_1 and δ_2 are real for subsonic c, but are otherwise complex. Use of the identities (17) and the resultant condition (19) yields the singular integral equation

$$\mathrm{Re}\left(\frac{\delta_1 - i\mu\delta_2}{e_0 c}\right)\frac{k}{2\pi}\int_{-a}^{a}\cot k\left(\frac{\eta-\xi}{2}\right)\phi(\xi)d\xi + \mathrm{Im}\left(\frac{\delta_1 - i\mu\delta_2}{e_0 c}\right)\phi(\eta) = (\mu-\mu^*)\frac{p^*}{G} + v_0\,\mathrm{Im}\left(\frac{\delta_1 - i\mu\delta_2}{e_0 c}\right), \quad -a<\eta<a. \tag{23}$$

Thus (23) subject to (19) represents a singular integral equation for determining the slip velocity for prescribed material properties and given p^*, μ^*, and μ. The resultant condition allows for the corresponding sliding velocity v_0 to be determined. The width of the pulse $2a$ is arbitrary.

For an isolated slip-pulse, as opposed to the periodic slip and stick zones considered thus far, $k\to 0 \Rightarrow v_0 \to 0$ (from (19)), and (23) simplifies to

$$\mathrm{Re}\left(\frac{\delta_1 - i\mu\delta_2}{e_0 c}\right)\frac{1}{\pi}\int_{-a}^{a}\frac{\phi(\xi)d\xi}{\eta-\xi} + \mathrm{Im}\left(\frac{\delta_1 - i\mu\delta_2}{e_0 c}\right)\phi(\eta) = (\mu-\mu^*)\frac{p^*}{G}, \quad -a<\eta<a. \tag{24}$$

Furthermore the shear and normal stresses become

$$\tau_{12}/G = q^*/G + \mathrm{Im}(\delta_1/e_0 c)\,\phi(\eta) + \mathrm{Re}(\delta_1/e_0 c)\,\frac{1}{\pi}\int_{-a}^{a}\frac{\phi(\xi)d\xi}{\eta-\xi} \tag{25}$$

$$\tau_{22}/G = -p^*/G - \mathrm{Re}(\delta_2/e_0 c)\,\phi(\eta) + \mathrm{Im}(\delta_2/e_0 c)\,\frac{1}{\pi}\int_{-a}^{a}\frac{\phi(\xi)d\xi}{\eta-\xi}. \tag{26}$$

It is interesting to observe that these results (25)–(26) generalize the analogous subsonic results of Weertman [15] to the intersonic and supersonic speed regimes. For subsonic speeds δ_1, δ_2, e_0 are all real and hence the second term in (25) and the third term in (26) vanish. Furthermore the subsonic slip-pulse obtained by Adams [12] occurs at the speed for which $\delta_1 = 0$ which corresponds to the slip wave speed or, as it is sometimes called, the generalized Rayleigh wave speed. In such a pulse the interface shear stress is equal to the remotely applied shear stress (25) and the normal stress is linearly related to the slip velocity (26). Thus a constant slip velocity in the slip region gives rise to a constant change in the contact pressure such that (13) is satisfied in the slip region. For supersonic speeds, $\delta_1/e_0 c$ is pure imaginary and $\delta_2/e_0 c$ is pure real so that the third terms in (25) and (26) vanish and thus the shear and normal stresses are linearly related to the slip velocity. This behavior allows for the slip pulse of Nosonovsky and Adams [19] to occur at any supersonic speed. However, in the intersonic speed regime, all three terms in each of (25) and (26) are present.

It may appear that an algebraic solution of (24) is possible at any speed for which the quantity $\mathrm{Re}\{(\delta_1 - i\mu\delta_2)/e_0 c\}$ vanishes. While such speeds exist and a solution given by

$$\phi(\eta) = \phi_0 \tag{27}$$

does satisfy the singular integral Eq. (24), the corresponding interfacial normal stresses are given by

$$\tau_{22}/G = -p^*/G - \mathrm{Re}(\delta_2/e_0 c)\,\phi_0 + \mathrm{Im}(\delta_2/e_0 c)\,\frac{\phi_0}{\pi}\log\left|\frac{a-\eta}{a+\eta}\right|, \quad -a<\eta<a, \tag{28}$$

which has logarithmic singularities of opposite signs at the two ends. Thus the requirement that the contact stresses be compressive is violated, no matter how large the remotely applied normal pressure.

Solutions of the form

$$\phi(\eta) = \Phi(\eta)(a-\eta)^{\alpha}(a+\eta)^{\beta} \tag{29}$$

will now be considered. For the solution to be bounded at both ends ($\alpha>0, \beta>0$) the consistency condition is required, i.e.,

$$\int_{-a}^{a}(\mu-\mu^*)(p^*/G)(a-\eta)^{-\alpha}(a+\eta)^{-\beta}d\eta = 0. \tag{30}$$

It can readily be seen that this condition is violated, except for the case of global sliding with $\mu=\mu^*$, and thus solutions of (24) which are bounded at both ends do not exist.

Now consider a solution singular at one end, and bounded at the other ($\alpha\beta<0, -1/2<\alpha<1, -1/2<\beta<1$). A solution may be obtained by following the procedure of Muskhelishvilli [23]. Equivalently the results which are tabulated by Erdélyi [24] may be used which gives

$$\gamma \equiv \beta = -\alpha, \quad \phi(\eta) = \phi_0\left(\frac{a+\eta}{a-\eta}\right)^{\gamma} \tag{31}$$

$$\tan\pi\gamma = -\mathrm{Re}\{(\delta_1 - i\mu\delta_2)/e_0\}/\mathrm{Im}\{(\delta_1 - i\mu\delta_2)/e_0\} \tag{32}$$

$$\phi_0 = -(\mu-\mu^*)(p^*/G)(\sin\pi\gamma)/\mathrm{Re}\{(\delta_1 - i\mu\delta_2)/e_0 c\}. \tag{33}$$

Finally the magnitude of the slip distance U_{Slip} may be determined by

$$U_{\mathrm{Slip}} = -\frac{1}{c}\int_{-a}^{a}\phi(\xi)d\xi = (\mu-\mu^*)(p^*/G)2\pi a\gamma/\mathrm{Re}\{(\delta_1 - i\mu\delta_2)/e_0\}. \tag{34}$$

The speed of the slip-pulse is arbitrary, except that any candidate solution must satisfy the following inequality constraints:

$$\begin{aligned}
&\textbf{A:}\ v_S \geqslant 0, \quad |\eta|<a \Rightarrow \mathrm{Im}\{(\delta_1 - i\mu\delta_2)/e_0 c\} \geqslant 0\\
&\textbf{B:}\ \tau_{22} \leqslant 0, \quad |\eta|<a \Rightarrow \mathrm{Re}\{\delta_2/e_0 c\} - \mathrm{Im}\{\delta_2/e_0 c\}/\tan\pi\gamma \geqslant 0,\\
&\qquad \text{and} \quad \mu^* \geqslant -\mathrm{Re}\{\delta_1/e_0 c\}/\mathrm{Im}\{\mu\delta_2/e_0 c\}\\
&\textbf{C:}\ \tau_{22} \leqslant 0, \quad |\eta|>a \Rightarrow -\mathrm{Im}\{\delta_2/e_0 c\}/\sin\pi\gamma \geqslant 0\\
&\qquad \text{and} \quad \mu^* \geqslant -\mathrm{Re}\{\delta_1/e_0 c\}/\mathrm{Im}\{\mu\delta_2/e_0 c\}
\end{aligned} \tag{35}$$

$$\textbf{D:}\ \tau_{12} + \mu\tau_{22} < 0, \quad |\eta|>a \Rightarrow -\mathrm{Re}\{(\delta_1 - i\mu\delta_2)/e_0 c\}/\sin\pi\gamma \geqslant 0$$

$$\textbf{E:}\ -\tau_{12} + \mu\tau_{22} < 0, \quad |\eta|>a \Rightarrow \mathrm{Re}\{(\delta_1 + i\mu\delta_2)/e_0 c\}/\sin\pi y \geqslant 0 \quad \text{and} \quad \mu^* \geqslant -\mathrm{Re}\{\delta_1/e_0 c\}/\mathrm{Im}\{\mu\delta_2/e_0 c\}$$

It can be seen that the second equations in each of **B**, **C**, and **E** are redundant. Furthermore, by using (32), it can be shown that **A** and **D** are also equivalent. Thus there are a total of five independent inequality constraints. The effect of the inequality constraints is to severely limit the range of speeds of the slip-pulse.

3 Results and Discussion

Results for a slip-pulse have been obtained and are shown in Figs. 2–8. In Figs. 2–4 are results corresponding to the material properties used in [8] and [16], i.e., $\kappa = 5/6$, $\rho'/\rho = 5/6$, and

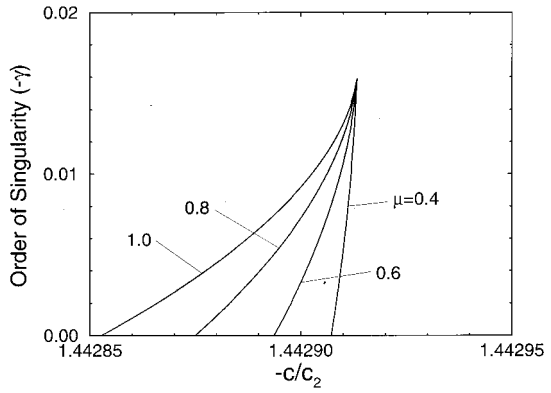


Fig. 2 The order of the singularity $(-\gamma)$ versus the negative of the normalized wave speed $(-c/c_2)$ for $\kappa=\rho'/\rho=5/6$, $\nu=\nu'=1/4$ and for various values of the friction coefficient

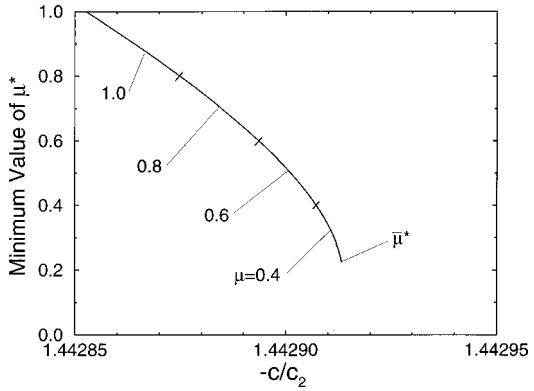


Fig. 3 The minimum value of μ^* versus the negative of the normalized wave speed $(-c/c_2)$ for $\kappa=\rho'/\rho=5/6$, $\nu=\nu'=1/4$ and for various values of the friction coefficient

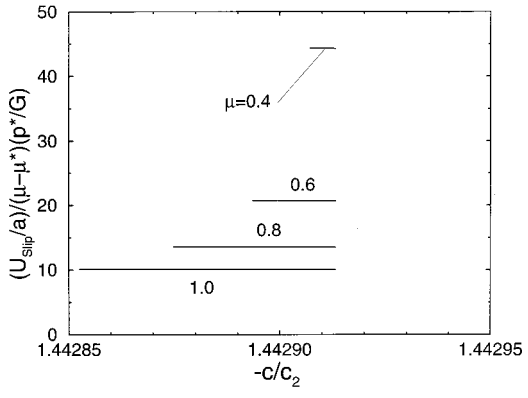


Fig. 4 The normalized slip distance $(U_{Slip}/a)/(\mu-\mu^*)(p^*/G)$ versus the negative of the normalized wave speed $(-c/c_2)$ for $\kappa=\rho'/\rho=5/6$, $\nu=\nu'=1/4$ and for various values of the friction coefficient

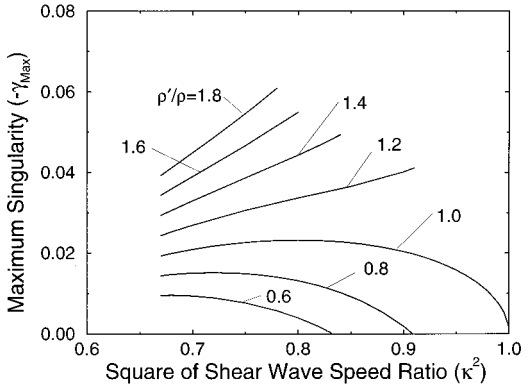


Fig. 5 The maximum value of the order of the singularity $(-\gamma_{Max})$ versus the square of the shear wave speed ratio (κ^2) for various values of ρ'/ρ, with $\mu=1$ and $\nu=\nu'=1/4$

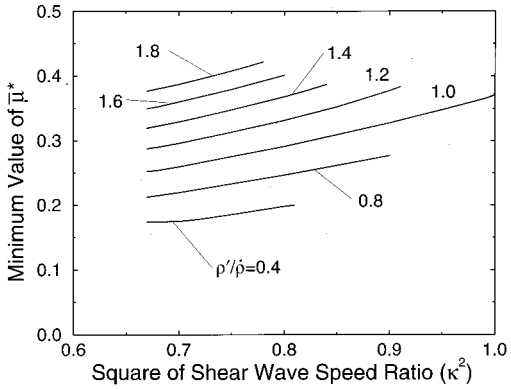


Fig. 6 The minimum value of the friction coefficient for which a slip-pulse exists $(\bar{\mu}^*)$ versus the square of the shear wave speed ratio (κ^2) for various values of ρ'/ρ, with $\mu=1$ and $\nu=\nu'=1/4$

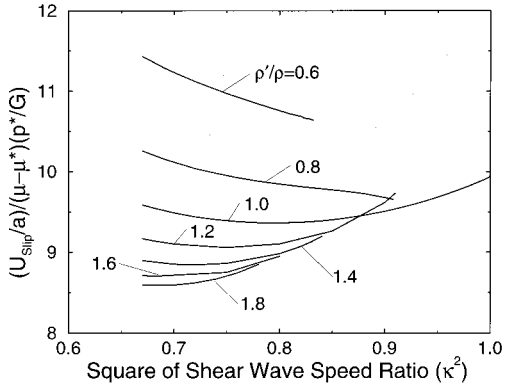


Fig. 7 The normalized slip distance $(U_{Slip}/a)/(\mu-\mu^*)(p^*/G)$ versus the square of the shear wave speed ratio (κ^2) for various values of ρ'/ρ, with $\mu=1$ and $\nu=\nu'=1/4$

$\nu=\nu'=1/4$. Figure 2 gives the order of the singularity $(-\gamma)$ at the leading edge versus the negative of the normalized wave speed $(-c/c_2)$ for four different values of the friction coefficient. Note that the effect of the inequalities is to severely restrict the range of possible wave speeds to an extremely narrow band which is just below the dilatational wave speed of the lower wave-speed material. The smaller the friction coefficient, the narrower is the wave-speed range. The direction of wave propagation is opposite to that of sliding of the lower wave-speed material and its maximum magnitude is slightly less than c_1'. Also note that the order of the singularity, which always occurs at the leading edge of the slip zone, is especially small.

Another effect of the inequality constraints is to restrict the minimum value of the ratio of remotely applied shear to normal

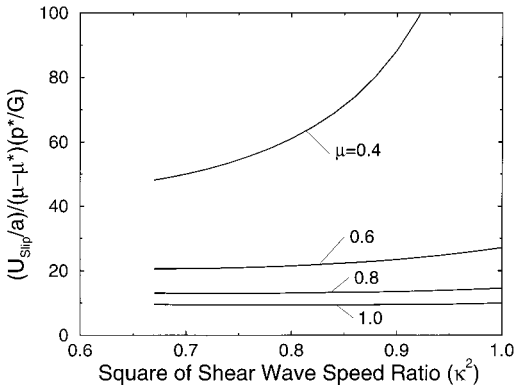


Fig. 8 The normalized slip distance $(U_{Slip}/a)/(\mu-\mu^*)(p^*/G)$ versus the square of the shear wave speed ratio (κ^2) for various values of μ with $\rho'/\rho=1$ and $\nu=\nu'=1/4$

pressure (μ^*) for which a slip-pulse can exist. Figure 3 shows the minimum value of μ^* versus $-c/c_2$ for four values of the friction coefficient. The four curves are identical within the regions in which they overlap. Each curve begins at the lowest acceptable wave speed with a minimum value μ^* equal to the friction coefficient μ. As the pulse-speed increases toward c_1', the minimum value of μ^* reaches a value which we denote $\overline{\mu}^*$. Not only is $\overline{\mu}^*$ independent of μ, but the minimum value of μ for which a slip-pulse exists also corresponds to $\overline{\mu}^*$. In Fig. 4 is shown the normalized slip distance $(U_{\text{Slip}}/a)/(\mu-\mu^*)(p^*/G)$ versus normalized speed $-c/c_2$ for various values of the friction coefficient. For fixed μ the slip distance does not vary significantly with speed, but is proportional to the pulse width $2a$, the remotely applied pressure p^*/G, and the quantity $(\mu-\mu^*)$. Thus the slip distance is greatest when the remotely applied shear stress is smallest, subject to the lower bound on μ^* discussed in connection with Fig. 3.

Figures 5–8 give results for a range of material combinations, with $\mu=1$ (Figs. 5–7) and with $\nu=\nu'=1/4$ (Figs. 5–8). Without loss of generality the lower shear wave speed material is taken to be the upper half-space. In Fig. 5 is shown the maximum value of the order of the leading edge singularity $(-\gamma_{\text{Max}})$ versus the square of the shear wave speed ratio (κ^2) for various values of the density ratio (ρ'/ρ). Note that each curve starts at $\kappa^2=2/3$, which corresponds to $c_1'=\sqrt{2}c_2$. This speed is known as the Eshelby singular dislocation velocity ([25]) and corresponds to the speed at which an edge dislocation in an isotropic medium can move without a shock wave appearing in its displacement field. The value of $-\gamma_{\text{Max}}$ increases as the density ratio increases for fixed shear wave speed ratio. In Fig. 6 is shown the parameter $\overline{\mu}^*$ (which represents the minimum value of μ and of μ^* for which a slip-pulse occurs with a given material combination) versus κ^2 for various values of ρ'/ρ. The parameter $\overline{\mu}^*$ decreases as the density ratio decreases for fixed shear wave-speed ratio. The variation of the normalized slip distance with the square of the shear wave-speed ratio for different values of the density ratio is shown in Fig. 7 and for different values of the friction coefficient in Fig. 8 (with $\rho'/\rho=1$). Note that as the density ratio and/or the friction coefficient decreases, the magnitude of the slip distance increases. This trend is, however, more pronounced with changes in the friction coefficient than it is with changes in ρ'/ρ.

4 Conclusions

The possible existence of an intersonic pulse at the frictional interface between two different elastic half-spaces has been investigated. The interface normal and shear stresses have been related to the slip velocity, which yields a generalization of the subsonic Weertman solution. An intersonic slip pulse is shown to exist for sufficient friction and for modest mismatches in the material combinations. These slip-pulses travel at a speed just below the slower dilatational wave speed; travel in the opposite direction of sliding of the lower wave-speed material; and are weakly singular at the leading edge and bounded at the trailing edge.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CMS-9622196 of the Surface Engineering and Tribology Program. The author is grateful to Prof. J. R. Rice for many helpful discussions.

References

[1] Stoneley, R., 1924, ''Elastic Waves at the Surface of Separation of Two Solids,'' Proc. R. Soc. London, **A106**, pp. 416–428.
[2] Achenbach, J. D., and Epstein, H. I., 1967, ''Dynamic Interaction of a Layer and a Half-Space,'' J. Eng. Mech., **EM5**, pp. 27–42.
[3] Comninou, M., and Dundurs, J., 1977, ''Elastic Interface Waves Involving Separation,'' ASME J. Appl. Mech., **44**, pp. 222–226.
[4] Comninou, M., and Dundurs, J., 1978, ''Elastic Interface Waves and Sliding Between Two Solids,'' ASME J. Appl. Mech., **45**, pp. 325–330.
[5] Freund, L. B., 1978, discussion, ''Elastic Waves Involving Separation,'' ASME J. Appl. Mech., **45**, pp. 226–228.
[6] Martins, J. A. C., Guimarães, J., and Faria, L. O., 1995, ''Dynamic Surface Solutions in Linear Elasticity and Viscoelasticity With Frictional Boundary Conditions,'' ASME J. Vibr. Acoust., **117**, pp. 445–451.
[7] Adams, G. G., 1995, ''Self-Excited Oscillations of Two Elastic Half-Spaces Sliding With a Constant Coefficient of Friction,'' ASME J. Appl. Mech., **62**, pp. 867–872.
[8] Ranjith, K., and Rice, J. R., 2001, ''Slip Dynamics at an Interface Between Dissimilar Materials,'' J. Mech. Phys. Solids, **49**, pp. 341–361.
[9] Prakash, V., and Clifton, R. J., 1993, ''Time Resolved Dynamic Friction Measurements in Pressure-Shear,'' *Experimental Techniques in the Dynamics of Deformable Solids*, Vol. AMD-165, ASME, New York, pp. 33–48.
[10] Prakash, V., 1998, ''Frictional Response of Sliding Interfaces Subjected to Time Varying Normal Pressure,'' ASME J. Tribol., **120**, pp. 97–102.
[11] Martins, J. A. C., Oden, J. T., and Simões, F. M. F., 1990, ''A Study of Static and Kinetic Friction,'' Int. J. Eng. Sci., **28**, pp. 29–92.
[12] Adams, G. G., 1998, ''Steady Sliding of Two Elastic Half-Spaces With Friction Reduction due to Interface Stick-Slip,'' ASME J. Appl. Mech., **65**, pp. 470–475.
[13] Rice, J. R., 1997, ''Slip Pulse at Low Driving Stress Along a Frictional Fault Between Dissimilar Media,'' EOS Trans. Am. Geophys. Union, **78**, No. 46, Fall Meeting Supplement, p. F464.
[14] Weertman, J. J., 1980, ''Unstable Slippage Across a Fault That Separates Elastic Media of Different Elastic Constants,'' J. Geophys. Res., **85**, pp. 1455–1461.
[15] Weertman, J. J., 1963, ''Dislocations Moving Uniformly on the Interface Between Isotropic Media of Different Elastic Properties,'' J. Mech. Phys. Solids, **11**, pp. 197–204.
[16] Andrews, D. J., and Ben-Zion, Y., 1997, ''Wrinkle-Like Slip Pulse on a Fault Between Different Materials,'' J. Geophys. Res., **102**, pp. 553–571.
[17] Caroli, C., 2000, ''On Slip Pulses at a Sheared Frictional Viscoelastic/Non-Deformable Interface,'' manuscript.
[18] Adams, G. G., 2000, ''Radiation of Body Waves Induced by the Sliding of an Elastic Half-Space Against a Rigid Surface,'' ASME J. Appl. Mech., **67**, pp. 1–5.
[19] Nosonovsky, M., and Adams, G. G., ''Dilatational and Shear Waves Induced by the Frictional Sliding of Two Elastic Half-Spaces,'' Int. J. Eng. Sci., in press.
[20] Cochard, A., and Rice, J. R., 2000, ''Fault Rupture Between Dissimilar Materials: Ill-Posedness, Regularization, and Slip-Pulse Response,'' Journal of Geophysical Research, **105**, pp. 25,891–25,907.
[21] Wolfram, S., 1991, ''Mathematica, A System for Doing Mathematics by Computer,'' 2nd Ed., Addison-Wesley, Reading, MA.
[22] Gel'fand, I. M., and Shilov, G. E., 1964, *Generalized Functions*, Vol. 1, Academic Press, New York.
[23] Muskhelishvili, N. I., 1958, *Singular Integral Equations*, P. Noordhoff, Groningen.
[24] Erdélyi, A., 1954, *Tables of Integral Transforms* (Bateman Manuscript Project), McGraw-Hill, New York.
[25] Eshelby, J. D., 1956, ''Supersonic Dislocations and Dislocations in Dispersive Media,'' Proc. Phys. Soc., **B69**, pp. 1013–1019.

Shear Coefficients for Timoshenko Beam Theory

J. R. Hutchinson
Professor Emeritus,
Department of Civil and
Environmental Engineering,
University of California,
Davis, CA 95616
Life Mem. ASME

The Timoshenko beam theory includes the effects of shear deformation and rotary inertia on the vibrations of slender beams. The theory contains a shear coefficient which has been the subject of much previous research. In this paper a new formula for the shear coefficient is derived. For a circular cross section, the resulting shear coefficient that is derived is in full agreement with the value most authors have considered ''best.'' Shear coefficients for a number of different cross sections are found.
[DOI: 10.1115/1.1349417]

Introduction

Timoshenko [1] was the first to introduce shear deformation, as well as rotary inertia, into the derivation of vibrating beam theory. He introduced a shear coefficient to account for the variation of the shear stress across the cross section. Timoshenko, in that first paper, used a value of 2/3 for a rectangular cross section. Many authors have found and used different values. Kaneko [2] did an excellent review of all the various shear coefficients that have been tried. His conclusion was that the values implied in Timoshenko's [3] come the closest to experimental results. Those values are $k=(6+12\nu+6\nu^2)/(7+12\nu+4\nu^2)$ for the circle and $k=(5+5\nu)/(6+5\nu)$ for the rectangle. Those coefficients were found, for the circle by matching with the Pochhammer-Chree solution for long wavelengths, and for the rectangular by matching with the plane stress solution. I will refer to those two values as Timoshenko's values. Cowper [4] derived shear coefficients for various cross sections for the static problem. His values agree with Timoshenko's values only for the case when Poisson's ratio is zero.

In Hutchinson [5] a highly accurate series solution for a completely free beam of circular cross section was compared with Timoshenko beam theory and it was concluded that Timoshenko's value was best for long wave lengths. Leissa and So [6] developed a highly accurate Rayleigh-Ritz solution for the circular cross section and compared their solution to Timoshenko beam theory using Cowper's shear coefficient. In a comment of that paper, Hutchinson [7], it was shown that for long wavelengths, use of Timoshenko's values of shear coefficient gave better results than use of Cowper's. Kaneko [2] also made comparisons with the experimental work of Spence and Seldin [8], Spinner, Reichard, and Tefft [9], and his own experimental results for both the circular and rectangular cross sections.

In Hutchinson and Zilmer [10] a three-dimensional series solution and a plane stress solution for the completely free beam was compared to the Timoshenko beam theory for rectangular cross sections. The plane stress solution compared very well with the Timoshenko beam theory using Timoshenko's shear coefficient. We were not able to use enough terms in the series solution to do a close comparison. For the solid cylinder problem the order of the characteristic matrix is the number of terms in the axial direction, whereas, for the rectangular cross section the order of the matrix is $n_x n_y + n_y n_z + n_z n_x$ where n_x, n_y, and n_z are the number of terms in the x, y, and z-directions, respectively. Kaneko [2] concluded on the basis of the experimental results that Timoshenko's value of the shear coefficient was best for this problem as well.

In Hutchinson and El-Azhari [11] a series solution for the completely free hollow cylinder was compared with the Timoshenko beam theory. Armenàkas, Gazis, and Herrmann [12] presented extensive tabulated results for infinitely long hollow circular cylinders. Leissa and So [13] gave accurate results for a free ended hollow circular cylinder. These three references will be used to check the new shear coefficient derived for this problem.

The approach used in this paper, to get around the discrepancies inherent in beam theory, is to choose a ''best'' guess for the stress field and a ''best'' guess for the displacement field. A variational form is then used in which these two fields can be incompatible. The variational form used is the Hellinger-Reissner principle, see Reissner [14]. The results of this approach are then compared to the Timoshenko Beam solution for long wavelengths, and an expression for the new shear coefficient is found. To set up the basis of comparison, the elementary Timoshenko beam formulation is carried out first.

Elementary Timoshenko Beam Formulation

The sign convention for beam geometry and shear and moment used throughout this paper is shown in Fig. 1. The assumption is made that the beam is symmetric (i.e., $I_{yz}=0$). The rotation of the cross section is denoted as ψ. The slope of the displacement v is made up of two effects. The rotation of the cross section plus the additional slope caused by the shear. If the shear were constant over the cross section the additional slope would be V/GA. The fact that it is not constant leads to the definition of a shear coefficient k, such that the additional slope is V/kGA. Thus,

$$v' = \psi + \frac{V}{kGA} \tag{1}$$

where primes denote differentiation with respect to x. The moment curvature relation is

$$M = EI_z\psi'. \tag{2}$$

Summation of forces in the vertical direction on a differential element gives

$$V' = \rho A\ddot{v} \tag{3}$$

where dots represent derivatives with respect to time. Summation of moments on a differential element gives

$$M' + V = \rho I_z\ddot{\psi}. \tag{4}$$

Eliminating V and M from the above four equations gives the following two equations:

$$EI_z\psi'' + (v' - \psi)kGA = \rho A\ddot{v} \tag{5}$$

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, June 7, 2000; final revision, August 15, 2000. Associate Editor: R. C. Benson. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

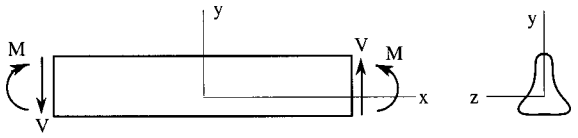


Fig. 1 Coordinates and positive moment and shear sign convention for the uniform beam

$$(v''-\psi')kGA=\rho A\ddot{v}. \tag{6}$$

In order to do the comparisons necessary in the next step it is necessary to solve these equations for a wavelength L and then let the wavelength increase. To this end the displacement v and the rotation ψ are expressed as follows:

$$v=B\sin(\alpha x)\sin(\omega t) \tag{7}$$

$$\psi=C\cos(\alpha x)\sin(\omega t). \tag{8}$$

The wavelength L is

$$L=\pi/\alpha. \tag{9}$$

Substituting Eqs. (7) and (8) into (5) and (6) gives

$$\begin{bmatrix} \alpha kGA & -EI_z\alpha^2-kGA+\rho I_z\omega^2 \\ -\alpha^2 kGA+\rho A\omega^2 & kGA\alpha \end{bmatrix}\begin{Bmatrix} B \\ C \end{Bmatrix}=\begin{Bmatrix} 0 \\ 0 \end{Bmatrix}. \tag{10}$$

Setting the determinant of the coefficients in Eq. (10) to zero gives

$$\lambda^4-\alpha^4+\frac{I_z}{A}\alpha^2\lambda^4\left(1+\frac{E}{kG}\right)-\lambda^8\frac{I_z^2}{A^2}\frac{E}{kG}=0 \tag{11}$$

where the frequency parameter λ is defined as

$$\lambda^4=\frac{\rho A\omega^2}{EI_z}. \tag{12}$$

The solution for a simple beam is $\alpha=\lambda$, and as the wavelength increases both α and λ approach zero. The first two terms in Eq. (11) are fourth order, the third term is sixth order, and the last term is eight order. For comparisons in the next section the eighth-order term will be neglected.

New Timoshenko Beam Formulation

The displacement and stress fields are described as follows:

$$u=-y\psi(x,t) \tag{13}$$

$$v=\varphi(x,t)+\frac{\nu}{2}y^2\psi'-\frac{\nu}{2}z^2\psi' \tag{14}$$

$$w=\nu yz\psi' \tag{15}$$

$$\sigma_x=-Ey\psi' \tag{16}$$

$$\sigma_y=0 \tag{17}$$

$$\sigma_z=0 \tag{18}$$

$$\tau_{yz}=0 \tag{19}$$

$$\tau_{xy}=\frac{V}{I_z}f_1(y,z) \tag{20}$$

$$\tau_{xz}=\frac{V}{I_z}f_2(y,z). \tag{21}$$

The displacement field is chosen consistent with the assumption that plane cross sections remain plane after deformation. The normal stresses and the shear stress τ_{yz} are also consistent with this assumption. The shearing stresses τ_{xy} and τ_{xz} are chosen consistent with the distribution in a tip-loaded cantilever. The tip-loaded cantilever is the only beam loading case for which an exact solution is available in the literature. The functions f_1 and f_2, defined above, are the in-plane distribution of the shearing stresses τ_{xy} and τ_{xz}. The tip-loaded cantilever solution can be found in Love [15]. That solution yields

$$f_1=-\frac{1}{2(1+\nu)}\left(\frac{\partial\chi}{\partial y}+\frac{\nu y^2}{2}+\frac{(2-\nu)z^2}{2}\right) \tag{22}$$

$$f_2=-\frac{1}{2(1+\nu)}\left(\frac{\partial\chi}{\partial z}+(2+\nu)yz\right) \tag{23}$$

where $\chi(y,z)$ is a harmonic function which satisfies the boundary condition

$$\frac{\partial\chi}{\partial n}=-n_y\left(\frac{\nu y^2}{2}+\frac{(2-\nu)z^2}{2}\right)-n_z(2+\nu)yz \tag{24}$$

on the boundary of the cross section. Solutions for χ are available in Love [15] and Cowper [4] as well as in a number of textbooks. Some interesting properties of f_1 and f_2 are as follows:

$$\int_A f_1(y,z)dA=I_z \tag{25}$$

$$\int_A f_2(y,z)dA=0. \tag{26}$$

The shear V in Eqs. (20) and (21) will be expressed as in the previous sections by combining Eqs. (2) and (4) thus,

$$V=I_z(\rho\ddot{\psi}-E\psi''). \tag{27}$$

The dynamic form of the Hellinger-Reissner principle for this problem can be written as

$$\delta\int_{t_1}^{t_2}\int_{\text{Vol}}\{\sigma_x u_{,x}-\sigma_x^2/2E+\tau_{xy}(u_{,y}+v_{,x})-\tau_{xy}^2/2G + \tau_{xz}(u_{,z}+w_{,x})-\tau_{xz}^2/2G-\tfrac{1}{2}\rho u_{,t}^2-\tfrac{1}{2}\rho v_{,t}^2 - \tfrac{1}{2}\rho w_{,t}^2\}d\text{Vol}\,dt=0. \tag{28}$$

Introduction of the definitions in Eqs. (13) to (21) and integration over the cross-sectional area yields

$$\delta\int_{t_1}^{t_2}\int_L\left\{\frac{1}{2}EI_z\psi'^2+I_z(E\psi''-\rho\ddot{\psi})(\psi-\varphi') -\frac{\nu}{2}(E\psi''-\rho\ddot{\psi})\psi''C_1-\frac{1}{2G}(-E\psi''+\rho\ddot{\psi})^2C_2-\frac{1}{2}\rho I_z\dot{\psi}^2 -\frac{1}{2}\rho\left[A\dot{\varphi}^2+\nu\dot{\varphi}\dot{\psi}'(I_z-I_y)+\frac{\nu^2}{4}\dot{\psi}'^2C_3\right]\right\}dLdt=0 \tag{29}$$

where the area integrals are defined as follows:

$$A=\int_A dA \quad I_z=\int_A y^2dA \quad I_y=\int_A z^2dA \tag{30}$$

$$C_1=\int_A(f_1y^2-f_1z^2+2f_2yz)dA \tag{31}$$

$$C_2=\int_A(f_1^2+f_2^2)dA \tag{32}$$

$$C_3=\int_A(y^4+z^4+2y^2z^2)dA. \tag{33}$$

Eliminating higher order terms and expanding the expressions in Eq. (29) yields

$$\delta\int_{t_1}^{t_2}\int_L\left\{\frac{1}{2}EI_z\psi'^2+EI_z\psi''\psi-EI_z\psi''\varphi'-\rho I_z\ddot{\psi}\psi+\rho I_z\ddot{\psi}\varphi'-\frac{\nu E}{2}C_1\psi''^2-\frac{E^2}{2G}C_2\psi''^2-\frac{1}{2}\rho I_z\dot{\psi}^2-\frac{1}{2}\rho A\dot{\varphi}^2-\frac{\nu\rho}{2}(I_z-I_y)\dot{\varphi}\dot{\psi}'\right\}dLdt=0. \quad (34)$$

Applying the Calculus of Variations to Eq. (34) yields the following two equations:

$$\psi''-\varphi'''+\frac{\rho}{E}\ddot{\varphi}'+\frac{C_4}{I_z}\psi'''-\frac{\rho}{E}\ddot{\psi}-\frac{\nu\rho}{2E}\left(1-\frac{I_y}{I_z}\right)\ddot{\varphi}'=0 \quad (35)$$

$$\psi'''-\frac{\rho}{E}\ddot{\psi}'+\frac{\rho A}{EI_z}\ddot{\varphi}+\frac{\nu\rho}{2E}\left(1-\frac{I_y}{I_z}\right)\ddot{\psi}'=0 \quad (36)$$

where

$$C_4=-\nu C_1-\frac{E}{G}C_2. \quad (37)$$

Treating φ and ψ in the same way v and ψ were treated in Eqs. (7) and (8) yields the following:

$$\begin{bmatrix}\alpha^3-C_5\alpha\lambda^4 & -\alpha^2+\frac{C_4}{I_z}\alpha^4+\frac{I_z}{A}\lambda^4\\ -\lambda^4 & \alpha^3-C_5\alpha\lambda^4\end{bmatrix}\begin{Bmatrix}B\\C\end{Bmatrix}=\begin{Bmatrix}0\\0\end{Bmatrix} \quad (38)$$

where

$$C_5=\frac{I_z}{A}\left[1-\frac{\nu}{2}\left(1-\frac{I_y}{I_z}\right)\right]. \quad (39)$$

Setting the determinant of the coefficients in Eq. (38) to zero, noting that λ^8 can be expressed as $\lambda^6\alpha^2$ and dividing by α^2 gives

$$\lambda^4-\alpha^4+2C_5\alpha^2\lambda^4-\frac{C_4}{I_z}\alpha^2\lambda^4-\frac{I_z}{A}\lambda^6-C_5^2\lambda^8=0. \quad (40)$$

The fourth-order terms in Eqs. (11) and (40) are identical. Equating the sixth-order terms in Eqs. (11) and (40) and solving for the shear coefficient k yields

$$k=-\frac{2(1+\nu)}{\left[\frac{A}{I_z^2}C_4+\nu\left(1-\frac{I_y}{I_z}\right)\right]} \quad (41)$$

where C_4 follows from Eqs. (31), (32), and (37) as

$$C_4=-\int_A[\nu(f_1y^2-f_1z^2+2f_2yz)+2(1+\nu)(f_1^2+f_2^2)]dA. \quad (42)$$

Shear Coefficients for Various Cross Sections

In this section shear coefficients for a variety of cross section are derived from Eq. (41). With the exception of the thin-walled cross sections, all of the following results were calculated using the value of χ taken from Love [15] (pp. 335–337). It should be noted that Love used the coordinates x and y in the plane of the cross section, whereas, I am using y and z in the plane. To get from Love's notation to mine change x to y and y to z. The functions f_1 and f_2 are calculated from Eqs. (22) and (23), and the coefficient C_4 is calculated from Eq. (42). All calculations in this section were carried out using Mathematica.

Circular Cross Section.

$$k=\frac{6(1+\nu)^2}{7+12\nu+4\nu^2} \quad (43)$$

Hollow Circular Cross Section.

$$k=\frac{6(a^2+b^2)^2(1+\nu)^2}{7a^4+34a^2b^2+7b^4+\nu(12a^4+48a^2b^2+12b^4)+\nu^2(4a^4+16a^2b^2+4b^4)} \quad (44)$$

where b is the outer radius and a is the inner radius.

Elliptical Cross Section.

$$k=\frac{6a^2(3a^2+b^2)(1+\nu)^2}{20a^4+8a^2b^2+\nu(37a^4+10a^2b^2+b^4)+\nu^2(17a^4+2a^2b^2-3b^4)} \quad (45)$$

where the bounding curve is defined as $y^2/a^2+z^2/b^2=1$.

Rectangular Cross Section.

$$C_4=\tfrac{4}{45}a^3b(-12a^2-15\nu a^2+5\nu b^2)+\sum_{n=1}^{\infty}\frac{16\nu^2b^5\left(n\pi a-b\tanh\left(\frac{n\pi a}{b}\right)\right)}{(n\pi)^5(1+\nu)} \quad (46)$$

$$k=-\frac{2(1+\nu)}{\left[\frac{9}{4a^5b}C_4+\nu\left(1-\frac{b^2}{a^2}\right)\right]} \quad (47)$$

where the depth of the beam (y-direction) is $2a$ and the width of the beam (z-direction) is $2b$.

Thin-Walled Cross Sections. For thin-walled cross sections the shear stress distribution can be found from the elementary shear formula. That value of shear stress can be used to find the f_1 and f_2 required in Eq. (42). As an example of this consider the thin-walled circular cylinder. The shear stress from the elementary formula is found as

$$\tau=\frac{VO}{Ib}=-\tau_{\theta x}=\frac{V}{I}a^2\sin\theta. \quad (48)$$

The shear stress can then be expressed as

$$\tau_{yx}=-\tau_{\theta x}\sin\theta=\frac{V}{I}a^2\sin^2\theta \quad (49)$$

$$\tau_{zx}=\tau_{\theta x}\cos\theta=-\frac{V}{I}a^2\sin\theta\cos\theta. \quad (50)$$

From the definitions of f_1 and f_2 it follows that

$$f_1=a^2\sin^2\theta \quad (51)$$

$$f_2=-a^2\sin\theta\cos\theta. \quad (52)$$

The shear coefficient which comes from this is

Table 1 Comparison of the frequencies from Table 9 in Leissa and So [13] with the frequencies from Timoshenko beam theory using the new shear coefficient and Cowper's shear coefficient. D_i/D_o is the ratio of the inner diameter to the outer diameter. *S* and *A* refer to the symmetric and antisymmetric modes. The "num" in the second column refers to the mode number. The tabulated frequencies are $\omega R_0\sqrt{\rho/G}$ where R_0 is the outer radius. $L/D_0=5$ and $\nu=0.3$.

	D_i/D_o	0.1	0.1	0.5	0.5	0.9	0.9
	num	S	A	S	A	S	A
L & S	1	0.1651	0.3990	0.1776	0.4096	0.2018	0.4344
New		0.1651	0.3989	0.1776	0.4098	0.2022	0.4386
Cowper		0.1649	0.3972	0.1771	0.4064	0.2012	0.4331
L & S	2	0.6811	0.9807	0.6726	0.9344	0.6740	0.8756
New		0.6807	0.9802	0.6740	0.9396	0.6929	0.9237
Cowper		0.6760	0.9711	0.6655	0.9244	0.6810	0.9032
L & S	3	1.2835	1.5718	1.1842	1.3469	1.0469	1.0874
New		1.2838	1.5773	1.1975	1.4012	1.1382	1.1817
Cowper		1.2694	1.5562	1.1749	1.3672	1.1110	1.1476

$$k=\frac{1+\nu}{2+\nu}. \tag{53}$$

Evaluation and Discussion of Results

The shear coefficient for the circular cross section in Eq. (43) corresponds exactly to the shear coefficient implied in Timoshenko's 1922 paper ([3]). As pointed out in the Introduction this coefficient has been widely accepted as ''correct'' and verified by both experiment and by accurate three-dimensional solutions.

The shear coefficient for a hollow circular cross section in Eq. (44) equals the one for the circular cross section in Eq. (43) when *a* goes to zero. Comparisons of the new coefficient with Cowper's coefficient are made in Table 1 using the accurate frequency values from Leissa and So [13] for the completely free beam. It can be seen from this table that for long wavelengths the frequencies are closer using this new coefficient than Cowper's coefficient. Comparisons were also made to the infinitely long hollow cylinder. Over 200 numerical values given in Armenàkas et al. [12] were compared to the values computed using the new coefficient and Cowper's coefficient. A sample of these results is given in Table 2. Again, it can be seen that the new coefficient gives better frequencies than the Cowper coefficient. The hollow cylinder solution was also used to determine the range of applicability of the Timoshenko beam theory for hollow cylinders. Figure 2 shows a plot of the shear coefficient which would be required to match the infinitely long three-dimensional theory with the Timoshenko beam theory. In this plot the new shear coefficient values are the straight horizontal lines. It can be seen that the required shear coefficient approaches the new shear coefficient value as the diameter-to-wave-length ratio D/L approaches zero. For the thin-walled cylinder ($a=0.99$ and $a=0.9$) the D/L would have to be less than about 0.5 to produce reasonable frequencies using Timoshenko theory. For $a=0.5$ the D/L would have to be less than about 3.5, whereas for a solid cylinder ($a=0$) the D/L could be as large as about 5.

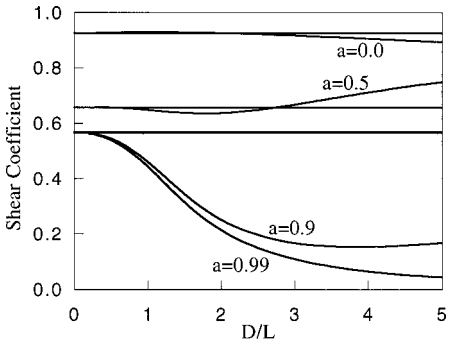


Fig. 2 Shear coefficient versus outer diameter to wavelength ratio for an infinitely long hollow cylinder. *a* is the ratio of the inner to outer radius. Straight horizontal lines are the new shear coefficient and the curved lines are the coefficient which is required to match the true solution.

The shear coefficient for an elliptical cross section in Eq. (45) equals the one for the circular cross section for $a=b$. Plots of the reciprocal of the shear coefficient as a function of the width-to-depth ratio of the elliptical cross section are given in Fig. 3. The plots show both the new shear coefficient (solid line) and the Cowper coefficient (dashed line) for values of Poisson's Ratio of 0.0, 0.25, and 0.5. For a Poisson's ratio of 0.0 the new shear coefficient and Cowper's shear coefficient coincide.

The new shear coefficient for the rectangular cross section corresponds to the Cowper coefficient for a Poisson's ratio of zero. It corresponds to Timoshenko's value when the width dimension is much less than the depth dimension, but it is a function of the width-to-depth ratio. None of the many values listed in Kaneko [2] have the shear coefficient as a function of the width-to-depth ratio, although Cowper states, ''It is remarkable that *K* is independent of the aspect ratio of the rectangle.'' Figure 4 shows a plot of the reciprocal of the shear coefficient as a function of the width-to-depth ratio for values of Poisson's ratio of 0.0, 0.25, and 0.5. As mentioned in the Introduction, the way that Timoshenko's value was found was to match the plane stress solution. This new coefficient does match the plane stress solution when the width-to-depth ratio is small. A comparison to the three-dimensional series solution is shown in Table 3. The series solution is for a completely free beam. The number of terms used in the series were chosen to make the number roughly correspond to the dimensions, and to keep the order of the characteristic matrix less than 2000. The example was chosen so that the depth and length of the beam remain constant so that if the shear coefficient were

Table 2 Comparison of the frequencies from Armenàkas et al. [12] with the frequencies from Timoshenko beam theory using the new shear coefficient and Cowper's shear coefficient. *H* is the thickness of the cylinder wall *R* is the mean radius of the wall, *L* is the wavelength and $\Omega=\omega H\sqrt{\rho/G}/\pi$.

H/R	H/L	New k	Cowper k	Ω New	Ω Cowper	Ω Armenàkas
0.25	0.001	0.5789	0.5438	0.0000144	0.0000144	0.0000144
0.25	0.005	0.5789	0.5438	0.000359	0.000359	0.000359
0.25	0.01	0.5789	0.5438	0.00141	0.00141	0.00141
0.25	0.02	0.5789	0.5438	0.00533	0.00531	0.00533
0.25	0.03	0.5789	0.5438	0.01106	0.01099	0.01106
0.25	0.04	0.5789	0.5438	0.01793	0.01774	0.01789
0.6	0.01	0.6408	0.6041	0.000621	0.000621	0.000621
0.6	0.05	0.6408	0.6041	0.01432	0.01427	0.01432
0.6	0.10	0.6408	0.6041	0.04777	0.04729	0.04776
0.6	0.12	0.6408	0.6041	0.06356	0.06278	0.06351
0.6	0.14	0.6408	0.6041	0.08000	0.07886	0.07987

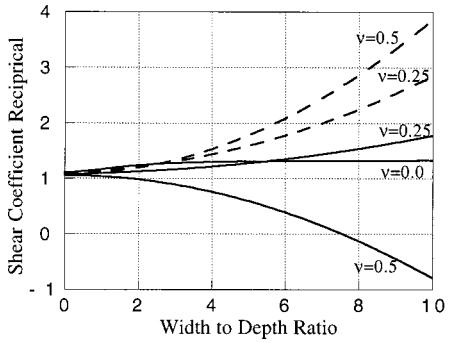


Fig. 3 Shear coefficient reciprocal versus width-to-depth ratio for an elliptical cross section for different values of Poisson's ratio. (——) new coefficient; (— — —) Cowper's coefficient.

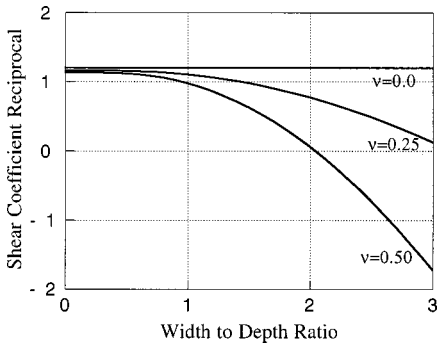


Fig. 4 Shear coefficient reciprocal versus width-to-depth ratio for a rectangular cross section for different values of Poisson's ratio

not changing the dimensionless frequency would remain constant. If one were to use Cowper's coefficient the shear coefficient would be 0.8571 and the frequency would be 0.10785. If Timoshenko's coefficient were used the shear coefficient would be 0.8824 and the frequency would be 0.10790. It can be seen in the table that the frequency values for the three-dimensional solution and the Timoshenko beam solution using the new coefficient follow the same trend.

Experimental results for the rectangular cross section have been limited to the square cross section with the exception of the work of Spinner et al. [9]. The work of Spinner et al. [9] did not contain enough data to prove or disprove the dependence of the shear coefficient on the aspect ratio. Kaneko [2] reduced the Spinner et al. [9] data on the assumption that the shear coefficient was not a function of the aspect ratio and came out with a shear coefficient which was a little less than the Timoshenko's value, whereas the new coefficient is greater. Kaneko's [2] own experimental results for a square cross section agree completely with Timoshenko's value. Spence and Seldin [8] found the shear coefficients which would match their experimental results for three different square cross sections. Those results are shown in Table 4 along with the values of Timoshenko's coefficient and the new coefficient. The best that can be said of the experimental comparisons is that the new shear coefficient is not out of line with experimental results, but the experimental results neither confirm nor negate the dependence of the new shear coefficient on the aspect ratio of the rectangular beam.

Table 3 Comparison of frequencies found using a three-dimensional series solution (3-D ω) with frequencies found using the new shear coefficient (New ω). The n_x, n_y, and n_z are the number of terms in the series in the x, y, and z-directions, respectively. $L/2$ is the half-length, a is the half-depth, and b is the half-width of the beam. The new shear coefficient is in the last column. Poisson's ratio was chosen as 1/2 in order to produce the greatest change with aspect ratio. The tabulated frequencies are $\omega 2a\sqrt{\rho/G}$.

n_x	n_y	n_z	$L/2$	a	b	3-D ω	New ω	k
130	13	2	5	0.5	0.05	0.1081	0.1079	0.8824
94	10	10	5	0.5	0.5	0.1082	0.1079	0.8946
77	8	16	5	0.5	1	0.1086	0.1081	1.0232
66	7	21	5	0.5	1.5	0.1091	0.1087	1.5864
60	6	24	5	0.5	2	0.1096	0.1095	15.9841

Table 4 Spence and Seldin [8] experimental determination of shear coefficient compared to the new coefficient, Timoshenko's coefficient, and Cowper's coefficient

Material	L/D	L/T	ν	S & S k	New k	Timo k	Cowper k
Aluminum	9.919	9.919	0.3665	0.874+-.005	0.880	0.872	0.852
Tool Steel	16.05	16.02	0.2870	0.873+-.002	0.871	0.866	0.849
Fused silica	45.55	42.06	0.1395	0.832+-.01	0.852	0.851	0.842

The use of the thin-wall approximation for calculation of the thin-walled circular cylinder led to exactly the same result that could be found by letting a approach b in Eq. (44) for the thick-walled cylinder. This approach is applicable to any thin-walled beams such as box beams or wide flange beams.

Static Problems

The main thrust of this paper has been concerned with the dynamic problem but since the work of Cowper [4] was for the static problem a brief look at the static problem is in order. The governing equations for the static problem corresponding to Eqs. (36) and (37) can be found to be

$$\psi - \varphi' + \frac{C_4}{I_z}\psi'' = 0 \tag{54}$$

$$\psi''' = 0. \tag{55}$$

Comparing the solution of these equations for the end-loaded cantilever to the solution of the beam including the shear deformation lead to the following equation for the shear coefficient:

$$k_s = -\frac{2(1+\nu)I_z^2}{AC_4}. \tag{56}$$

This coefficient was found for the deflection of the original centroidal axis. If the deflection is for the mean deflection of the cross section, as was done by Cowper [4], then the expression becomes

$$k_{sc} = -\frac{2(1+\nu)}{\left[\dfrac{A}{I_z^2}C_4 + \dfrac{\nu}{2}\left(1 - \dfrac{I_y}{I_z}\right)\right]}. \tag{57}$$

Note, that the half in the denominator makes this coefficient different from the dynamic coefficient defined in Eq. (41). This coefficient corresponds to Cowper's coefficient only when Poisson's ratio equals zero. For the cases of the circular cross section and the hollow circular cross section this new static coefficient is the same as the dynamic coefficients given in Eqs. (43) and (44).

Conclusions

The new shear coefficient is in complete agreement with the values that have been found from three-dimensional elasticity theory for the circular cross section and the plane stress solution. For the hollow circular cross section it is also shown by comparison to three-dimensional elasticity to be correct. For rectangular cross sections the new coefficient was found to be a function of the aspect ratio. Previous researchers have all either assumed it to not be a function or have derived it in such a way that it was not a function of the aspect ratio. Comparison to a three-dimensional series solution indicates that the new coefficient is probably correct, but experimental evidence is inconclusive.

Acknowledgment

I wish to thank Arthur Leissa and Charles Bert for inspiring this research at the Second International Symposium on Vibrations of Continuous Systems.

References

[1] Timoshenko, S. P., 1921, ''On the Correction for Shear of the Differential Equation for Transverse Vibrations of Bars of Prismatic Bars,'' Philos. Mag., **41**, pp. 744–746.
[2] Kaneko, T., 1975, ''On Timoshenko's Correction for Shear in Vibrating Beams,'' J. Phys. D **8**, pp. 1927–1936.
[3] Timoshenko, S. P., 1922, ''On the Transverse Vibrations of Bars of Uniform Cross Section,'' Philos. Mag., **43**, pp. 125–131.
[4] Cowper, G. R., 1966, ''The Shear Coefficient in Timoshenko's Beam Theory,'' ASME J. Appl. Mech., **33**, pp. 335–340.
[5] Hutchinson, J. R., 1981, ''Transverse Vibrations of Beams, Exact Versus Approximate Solutions,'' ASME J. Appl. Mech., **48**, pp. 923–928.
[6] Leissa, A. W., and So, J., 1995, ''Comparisons of Vibration Frequencies for

Rods and Beams From One-Dimensional and Three-Dimensional Analyses,'' J. Acoust. Soc. Am., **98**, pp. 2122–2135.
[7] Hutchinson, J. R., 1996, comments on ''Comparisons of Vibration Frequencies for Rods and Beams From One-Dimensional and Three-Dimensional Analyses,'' J. Acoust. Soc. Am., **98**, pp. 2122–2135 (1995); **100**, pp. 1890–1893.
[8] Spence, G. B., and Seldin, E. J., 1970, ''Sonic Resonances of a Bar and Compound Torsional Oscillator,'' J. Appl. Phys., **41**, pp. 3383–3389.
[9] Spinner, S., Reichard, T. W., and Tefft, W. E., 1960, ''A Comparison of Experimental and Theoretical Relations Between Young's Modulus and the Flexural and Longitudinal Resonance Frequencies of Uniform Bars,'' J. Res. Natl. Bur. Stand., Sect. A, **64A**, pp. 147–155.
[10] Hutchinson, J. R., and Zillmer, S. D., 1986, ''On the Transverse Vibration of Beams of Rectangular Cross-Section,'' ASME J. Appl. Mech., **53**, pp. 39–44.
[11] Hutchinson, J. R., and El-Azhari, S. A., 1986, ''Vibrations of Free Hollow Circular Cylinders,'' ASME J. Appl. Mech., **53**, pp. 641–646.
[12] Armenàkas, A. E., Gazis, D. C., and Herrmann G., 1969, *Free Vibrations of Circular Cylindrical Shells*, Pergamon Press, Oxford, UK.
[13] Leissa, A. W., and So, J., 1997, ''Free Vibrations of Thick Hollow Circular Cylinders From Three-Dimensional Analysis,'' ASME J. Vibr. Acoust., **119**, pp. 89–95.
[14] Reissner, E., 1950, ''On a Variational Theorem in Elasticity,'' J. Math. Phys., **29**, pp. 90–95.
[15] Love, A. E. H., 1944, *A Treatise on the Mathematical Theory of Elasticity*, Dover, New York.

Modeling the Fracture of a Sandwich Structure due to Cavitation in a Ductile Adhesive Layer

S. Zhang

K. J. Hsia[1]
Mem. ASME,

Department of Theoretical
and Applied Mechanics,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801

The strength and durability of adhesively bonded sandwich structures often depend on the mechanisms of fracture, which in turn depend on the properties of the adhesive and the microstructures of the interface. When the thin adhesive layer is ductile, cavitation either within the layer or along the interface is often the dominant failure mechanism. In the present paper, fracture due to cavity growth in a thin ductile layer is analyzed. A new method utilizing fluid mechanics solutions is developed. Solutions of fluid flow field are used to approximate the plastic deformation field in the corresponding solid body with a cavity. The equilibrium condition is satisfied by using the principle of virtual work rate. Stress-separation curves due to cavitation in the thin layer can thus be obtained. The method is validated by reevaluating the one-dimensional problem of cavity growth in a sphere—a problem for which an exact, analytical solution exists. A two-dimensional plane strain cavitation problem is analyzed using the new method. The stress-separation curves and the fracture resistance due to this mechanism are obtained. The results show that both the stress-separation curves and the fracture resistance are sensitive to the strain-hardening exponent and the initial void size, but not the yield strength of the material. The new method has clear advantages over numerical methods, such as the finite element method, when parametric studies are performed.
[DOI: 10.1115/1.1346678]

1 Introduction

The strength and durability of sandwich structures consisting of two ceramic or metal pieces bonded by a thin adhesive layer are determined by various failure mechanisms. These fracture mechanisms include interfacial debonding and other processes such as cavitation or microcracking within the adhesive layer or at the interface. Identifying and understanding the failure mechanisms in these structures will greatly enhance our ability to design better, more durable structures.

The failure mechanisms, however, are ultimately determined by the properties of the adhesive and by the microstructures of the interface in sandwich structures. The current research stems from the need to tailor the surface microstructures of aluminum panels by surface treatments in preparation for adhesive bonding (see, e.g., [1,2]). Within the constraints of surface treatment technology, a guideline to achieve an optimized microstructure is highly desirable.

In many such structures, the adhesive is often a soft or ductile phase. It may be a polymer-based material for joints in aircraft structures, or ductile metal in metal/ceramic composites. There have been many studies on failure mechanisms within a ductile layer bonding two substrate pieces together ([3–7]). When the adhesive layer is sufficiently soft, the failure process is crack propagation by void growth and coalescence within the ductile layer or along the interface. In this case, a large hydrostatic stress is developed in the ductile layer due to the constraint on plastic flow by the substrate, leading to void nucleation ahead of the crack tip. The location and density of void nucleation, however, is often related to the microstructures of the interface, such as initial pore density and interface roughness. For given microstructures, the initial void density can be considered as given. Failure of such sandwich structures is then directly related to the microstructures of the interface.

Fracture process of a material can be characterized by the stress-separation curve ahead of the crack tip. For purely brittle fracture of crystals, such stress-separation curves can be derived from the interatomic potentials. When nonlinear processes are involved, however, derivations of such stress-separation curves must invoke micromechanisms during fracture. For example, plastic dissipation must be taken into account in the case of elastoplastic fracture. Analyses of crack growth resistance due to plastic dissipation were carried out by Tvergaard and Hutchinson [8,9], who identified several dimensionless groups of material parameters characterizing the fracture process. An equivalent stress-separation curve for fracture due to cavity growth and coalescence may be derived from the detailed study of the cavitation process. In the present paper, failure due to cavity growth and coalescence will be studied by analyzing the stress-separation curves during cavity growth.

Cavitation has been studied by many researchers since the 1960s. The pioneer work by McClintock [10] revealed that the volume expansion rate of a long cylindrical cavity in a nonhardening material subjected to transverse tensile stress increases exponentially with the transverse stress. Rice and Tracey [11] analyzed the growth of a single spherical void embedded in an infinite body subjected to remote uniform tensile stresses. They found that the ratio of void growth rate to remote strain rate increases exponentially as the ratio of mean normal stress to yield stress increases. Their analysis also predicted that void growth is mainly due to volume change rather than shape change of the void when the remote normal stress is large. Both the above analyses were carried out on an infinite body, which is inappropriate for cavities in a confined ductile layer. Needleman [12] and Andersson [13] studied void growth numerically in a finite body using

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Oct. 1, 1999; final revision, July 19, 2000. Associate Editor: B. Moran. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

the finite element method. In these studies the interaction between voids was taken into account, but the amount of cavity growth was limited. More recently Tvergaard [14] analyzed void growth in a thin, ductile layer between ceramics using finite element method, and employed a remeshing technique for the final stage of growth. These numerical studies require tremendous computing power, and are usually rather time-consuming.

If the material of the ductile layer obeys an elastoplastic constitutive law, the nonlinearity of the governing equations seems to exclude the possibility of obtaining exact solutions for all but the one-dimensional case studied by McClintock [10] and Huang et al. [15]. In the present paper, to derive the stress-separation curves for a material undergoing cavitation in a thin ductile layer, we develop a novel approach to finding an admissible deformation field around the void. The approach utilizes fluid mechanics solutions of a point source in a finite unit cell, and approximates the plastic deformation field with a fluid flow field. The appropriateness of the approach is verified by reevaluating the spherically symmetric cavitation problem for which an analytical solution exists ([10,15]). A two-dimensional plane strain problem is then analyzed to obtain the stress-separation relation of a unit cell with a center cavity. The results show that the stress-separation relations depend not only on the material properties but also on the geometrical parameters (microstructures), such as initial void size and void spacing.

2 Spherically Symmetric Cavitation

We begin our discussion by considering the spherically symmetric cavitation problem. Consider a spherical void centered in an isotropic, rigid-plastic sphere (either perfectly plastic or strain hardening without elastic response) subjected to hydrostatic tension σ_s (see Fig. 1). A uniaxial relation between the true stress, σ, and the logarithmic strain, ε, of the solid is given by

$$\sigma/\sigma_Y = f(\varepsilon) = \left|\frac{E}{\sigma_Y}\varepsilon\right|^N \mathrm{sgn}(\varepsilon) \tag{1}$$

where σ_Y is the tensile yield strength of the solid, E is the Young's Modulus, N is the hardening exponent ($0 \leqslant N \leqslant 1$), $\mathrm{sgn}(\varepsilon)$ represents the sign of ε. The limit $N=0$ corresponds to a rigid-perfectly plastic material.

Two different methods to obtain the relation between the hydrostatic stress and the void expansion are presented. One method is based on classical plasticity theory in solid mechanics. The other employs a fluid mechanics approach, and treats the void as a point source of material flowing outward under the applied stress. With the latter the plastic deformation is to be represented by a potential flow generated from a point source. Both methods are capable of solving this one-dimensional cavitation problem analytically, as shown below.

2.1 Solid Mechanics Method. In what follows, capital letters stand for variables in the initial configuration and lowercase ones stand for variables in the current, deformed configuration. Let R_0 and r_0 be the radii of the cavity in the initial and current configuration, and R_1 and r_1 be the radii of the outer boundary in the initial and current state, respectively (Fig. 1). By symmetry, the true strain components in spherical coordinates (r,θ,φ) are

$$\varepsilon_\theta = \varepsilon_\varphi = -\frac{1}{2}\varepsilon_r = -\ln\frac{r}{R} \tag{2a}$$

$$\varepsilon_e = 2\ln\left(\frac{r}{R}\right) = \varepsilon_r \tag{2b}$$

where ε_e is the von Mises equivalent strain, defined as

$$\varepsilon_e = \sqrt{\frac{2}{3}\varepsilon_{ij}\varepsilon_{ij}} \tag{3}$$

where ε_{ij} $(i,j=r,\theta,\varphi)$ are the logarithmic strain components, and the summation convention applies in Eq. (3). The equilibrium condition in terms of radial stress σ_r and hoop stress σ_θ in the current configuration is

$$\frac{d\sigma_r}{dr} + \frac{2}{r}(\sigma_r - \sigma_\theta) = 0. \tag{4}$$

The von Mises equivalent stress in the spherically symmetric case can be expressed as

$$\sigma_e = \sigma_r - \sigma_\theta. \tag{5}$$

Assuming that the material obeys J_2-deformation theory, i.e., the relation between equivalent stress and equivalent strain follows that of uniaxial relation in Eq. (1), one has

$$\frac{\sigma_r - \sigma_\theta}{\sigma_Y} = f(\varepsilon_r) \tag{6}$$

where the function $f(\varepsilon_r)$ is given in Eq. (1). Using the boundary conditions $(\sigma_r|_{r=r_0}=0, \sigma_r|_{r=r_1}=\sigma_s)$ and the incompressibility requirement $(r^3 - r_0^3 = R^3 - R_0^3)$, and substituting Eqs. (2) and (6) into Eq. (4), one finds

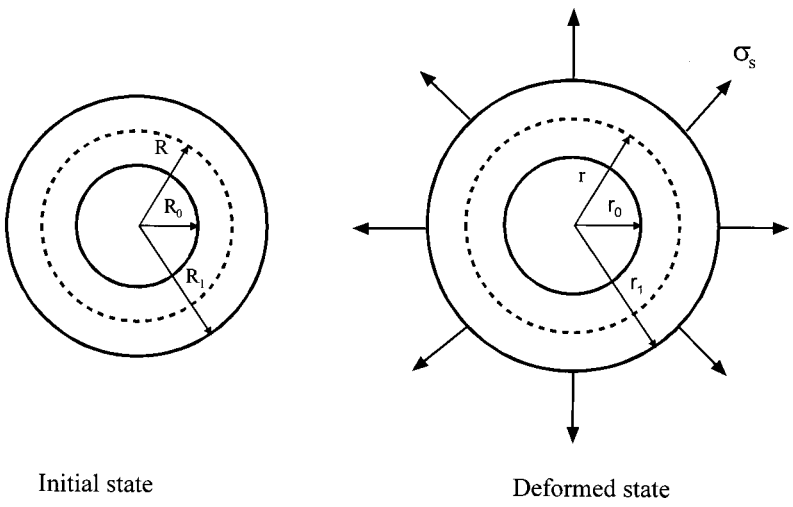


Fig. 1 Geometry of the spherically symmetric void in the initial and deformed states

$$\frac{\sigma_s(r_0)}{\sigma_Y} = -2\int_{R_0}^{R_1} \frac{R^2}{R^2+r_0^3-R_0^3} f\left\{-\frac{2}{3}\ln\left(1+\frac{r_0^3-R_0^3}{R^3}\right)\right\} dR. \tag{7}$$

Equation (7) gives the relationship between the applied hydrostatic stress and the cavity radius, $\sigma_s(r_0)$. It should be pointed out that the solution in Eq. (7) is identical to that obtained by Huang et al. [15].

2.2 Fluid Mechanics Method. The problem shown in Fig. 1 can also be solved using a fluid mechanics approach. Consider a potential flow generated by a point source of strength Q at the center of the sphere. The velocity components of the potential flow in spherical coordinates (r,θ,φ) are given by

$$\dot{u}_r = \frac{Q}{4\pi r^2}, \quad \dot{u}_\theta = \dot{u}_\varphi = 0. \tag{8}$$

Integrating the radial velocity with respect to time with the initial condition $r|_{t=0} = R$ gives

$$Qt = \frac{4\pi}{3}(r^3 - R^3). \tag{9}$$

This equation shows that $r^3 - R^3$ is an invariant throughout the solid body at any specific time t. This invariance requirement implies incompressibility of the material.

The strain rate components can be obtained from the velocity components by taking the derivative of the radial velocity with respect to the radius r, or by dividing the radial velocity by r, as

$$\dot{\varepsilon}_r = -\dot{\varepsilon}_e = -2\dot{\varepsilon}_\theta = -2\dot{\varepsilon}_\varphi = -\frac{Q}{2\pi r^3} \tag{10}$$

where $\dot{\varepsilon}_e$ is the equivalent strain rate defined as

$$\dot{\varepsilon}_e = \sqrt{\frac{2}{3}\dot{\varepsilon}_{ij}\dot{\varepsilon}_{ij}} \tag{11}$$

where $\dot{\varepsilon}_{ij}$ $(i,j = r,\theta,\varphi)$ are the true strain rate components. By integrating the strain rates with respect to time and using the incompressibility condition, one finds exactly the same expression as in Eq. (2). This result shows that the potential flow generated by a point source gives the same plastic deformation field around the void as that by the solid mechanics method given in Section 2.1.

To obtain the stress-cavity growth relations, we use here the principle of virtual work rate instead of the equilibrium equation in solid mechanics. Such an approach is entirely based on the estimates of the velocity field and strain rate field, and makes use of the constitutive law in an integral sense. The principle of virtual work rate in the current state is

$$\int_S \sigma_s \dot{u}_r dS = \int_V \sigma_{ij}\dot{\varepsilon}_{ij} dV \tag{12}$$

where S is the surface (including the outer surface and the inner surface although the work done on the inner surface is zero since it is traction-free) and V is the volume of the solid. Assuming again that the material follows the J_2 flow rule and it is a von Mises material, one has

$$s_{ij} = \frac{2}{3}\frac{\sigma_e}{\dot{\varepsilon}_e}\dot{\varepsilon}_{ij} \tag{13}$$

where s_{ij} $(i,j = r,\theta,\varphi)$ represent the deviatoric stress components defined as

$$s_{ij} = \sigma_{ij} - \frac{1}{3}\sigma_{kk}\delta_{ij} \tag{14}$$

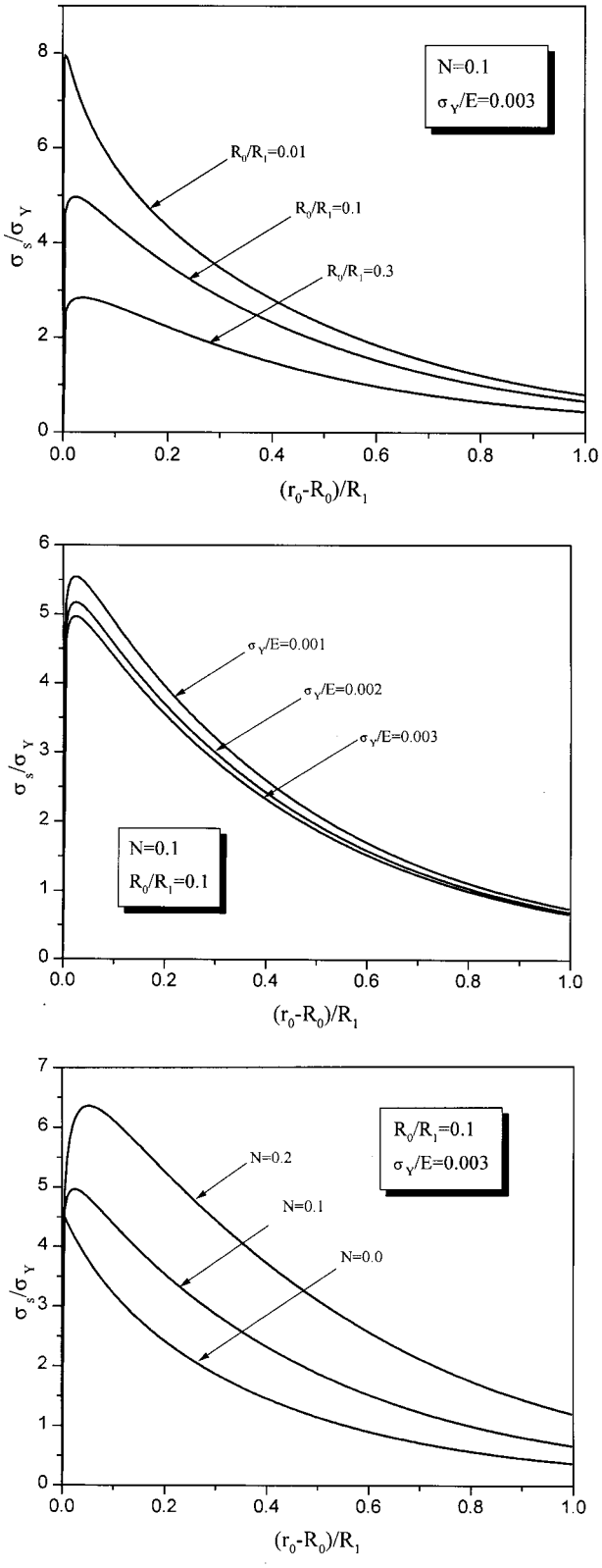


Fig. 2 Stress versus void radius for the growth of a spherically symmetrical void; (*a*) effects of the initial radius of cavity, (*b*) effects of the material constant σ_Y/E, (*c*) effects of the strain hardening exponent N

where $\sigma_{kk}=\sigma_{rr}+\sigma_{\theta\theta}+\sigma_{\varphi\varphi}$, and δ_{ij} is the Kronecker delta. Assuming that the constitutive law between the equivalent true stress and true strain follows the uniaxial relation expressed in Eq. (1), and using the flow field given in Eqs. (8), (10), and (11), one can rewrite Eq. (12) as

$$\frac{\sigma_s}{\sigma_Y}=\frac{1}{2\pi}\int_V f[2\ln(r/R)]r^{-3}dV. \tag{15}$$

Changing the variables to those for the initial state by using the incompressibility condition, one finally obtains the same stress-separation relationship as that in Eq. (7).

Using either Eq. (15) or Eq. (7), curves of the hydrostatic stress versus cavity radius are plotted in Figs. 2(*a*)–(*c*). Figure 2(*a*) shows the normalized hydrostatic stress versus the normalized current radius of the void for prescribed strain-hardening exponent and material constants σ_Y/E. Curves for three different values of initial void radius are plotted. The figure shows that the normalized stress reaches a maximum value rapidly, then decays monotonically as the cavity grows. The solution also shows that a smaller initial void size gives rise to a higher hydrostatic stress for a given amount of void growth and a higher maximum stress. When the loading process is stress-controlled, reaching the maximum load results in instability of void growth ([3,16–18]). When the loading is displacement-controlled, a softening stage (load drop) is experienced.

The effects of strain hardening exponent and yield strength on the stress versus void growth behavior are shown in Figs. 2(*b*)–(*c*). Figure 2(*b*) shows the normalized hydrostatic stress versus normalized void radius for different values of the material constant σ_Y/E for given initial radius of the cavity and strain-hardening exponent. Figure 2(*c*) shows the normalized stress versus normalized void radius for different values of strain-hardening exponent for given initial radius of the void and material constants. The figures show that the stress versus cavity growth curves are insensitive to the value of the normalized yield strength σ_Y/E, but rather sensitive to the strain-hardening exponent and the initial radius of the void. It is not unexpected to see strong dependence of the curves on hardening exponent and weak dependence on initial yield strength since the deformation becomes very large around the void as it grows. But the prediction of strong dependence on the initial cavity size is interesting.

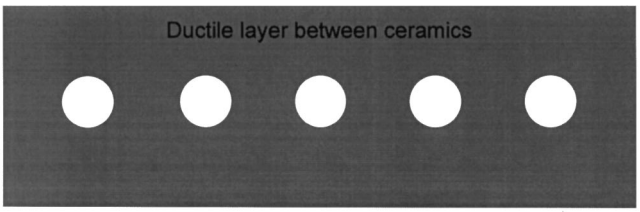


Fig. 3 Schematics of a cavitated ductile interface layer with periodic cavity distribution

3 Fully Confined Two-Dimensional Void Growth in a Thin Ductile Layer

In the previous section we showed that the method of using a fluid mechanics solution to approximate the plastic flow field is indeed a viable one. For the one-dimensional problem considered in the previous section, the solution turns out to be exact. But for more complicated cases, the solution can only be considered approximate. In the present section, we consider a two-dimensional problem.

A periodic array of cavities in a ductile layer fully confined by the interfaces, shown schematically in Fig. 3, is considered. Plane strain deformation is assumed. The cavities can be either completely within the ductile layer generated at, e.g., second phase particles, or at the interface generated from the interfacial pores. Due to symmetry of the problem, the solution should be identical for these two cases. The stress-separation curves, $\sigma(\delta)$, due to cavity growth and coalescence can be evaluated by considering a representative unit cell containing a single cavity. It is expected that, because of the confinement, plastic flow-induced cavity growth may start before the strength of the interface is reached. Due to the constraint by the rigid interface, high triaxial stresses will develop in the thin layer and will be the main driving force for cavity growth.

The geometry of the unit cell is presented in Fig. 4. The void spacing is $2w_0$, the layer thickness is $2h_0$, and the initial radius of the void is R_0. A Cartesian coordinate system with origin at the center of the cavity is established as in Fig. 4. Uniform tensile stress, σ_s, is applied normal to the thin layer. The periodicity condition requires that the width of the unit cell, $2w_0$, remains constant during deformation. The separation displacement, δ, is evaluated at the interface, $x=\pm h_0$.

To obtain the approximate plastic deformation field in the unit cell, we now consider a potential flow generated by a source of strength Q located at $z=0$ in an infinite channel in the domain $-w_0<y<w_0$ and $-\infty<x<\infty$; here $z=x+yi$ is the complex variable. The complex potential ψ of the flow field is found by conformal mapping, as

$$\psi(z)=\frac{Q}{2\pi}\ln\left\{\sinh\left(\frac{\pi z}{2w_0}\right)\right\}. \tag{16}$$

The velocity field corresponding to the potential flow is

$$\dot{u}_x-\dot{u}_y i=\frac{Q}{4w_0}\coth\left(\frac{\pi z}{2w_0}\right). \tag{17}$$

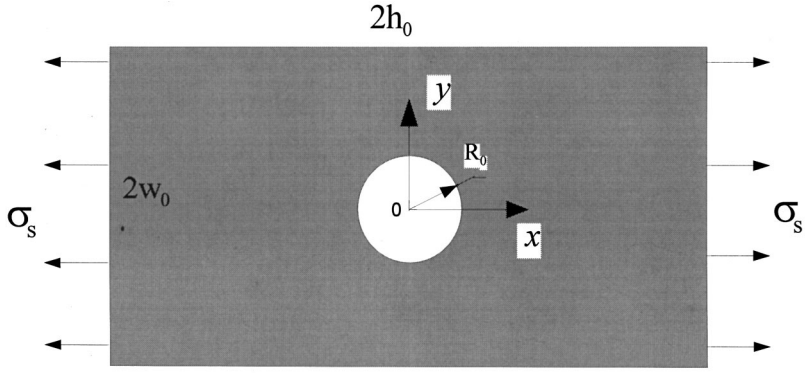


Fig. 4 Unit cell model used in the two-dimensional void growth analysis

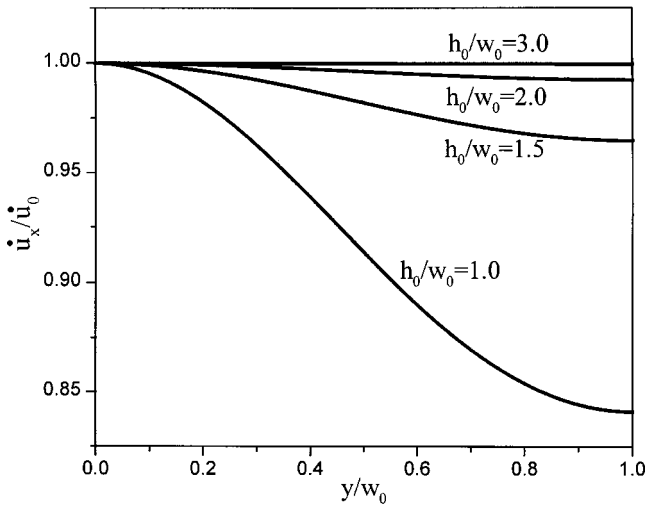


Fig. 5 Uniformity of the velocity field at the interface $x=h_0$

The Cauchy strain rates corresponding to the velocity field can be obtained by differentiating Eq. (17), as

$$\dot{\varepsilon}_{xx}-\dot{\varepsilon}_{xy}i=-\frac{\pi Q}{8w_0^2}\mathrm{csch}^2\left(\frac{\pi z}{2w_0}\right)$$
$$\dot{\varepsilon}_{xy}-\dot{\varepsilon}_{yy}i=-i\frac{\pi Q}{8w_0^2}\mathrm{csch}^2\left(\frac{\pi z}{2w_0}\right). \tag{18}$$

It is easily shown that the flow becomes uniform as $x\to\pm\infty$ ($\dot{u}_s-\dot{u}_y i=Q/4w_0$ as $x\to\pm\infty$), leading to vanishing strain rates. In fact, the flow becomes nearly uniform when the location under consideration, $x=h_0$, is several times larger than the width w_0. Figure 5 shows the variation of the normal velocity $\dot{u}_x$ at $x=\pm h_0$ across the width of the channel for different values of the ratio h_0/w_0. In this figure, $\dot{u}_x$ is normalized by $\dot{u}_0$, which is the normal velocity at $x=h_0$, $y=0$. It is seen that when the ratio of h_0/w_0 is unity, the maximum difference in $\dot{u}_x$ across the width is about 15 percent. When the ratio is 2.0, the maximum difference in $\dot{u}_x$ is only two percent, and $\dot{u}_x$ is approximately uniform across the width of the channel. In the following, we use the solution of the infinite channel to approximate the plastic flow field in the finite sized unit cell in Fig. 4.

The equilibrium condition can be satisfied in a weak form by using the principle of virtual work rate, as

$$2\int_{-w_0}^{w_0}\sigma_s\dot{u}_x^I dy=\int_V s_{ij}\dot{\varepsilon}_{ij}dV-2\int_{-w_0}^{w_0}\Delta T_i^I\dot{u}_i^I dy \tag{19}$$

where ΔT_i^I is the difference in traction at $x=\pm h_0$ between the average normal stress σ_s and the stress corresponding to the plastic flow, $\dot{u}_x^I$ and $\dot{u}_i^I$ are the displacement rates at $x=\pm h_0$ given in Eq. (17), and s_{ij}, $\dot{\varepsilon}_{ij}$ are the true deviatoric stress and true strain rate in the current configuration. Generally, when the ratio h_0/w_0 is sufficiently large (say, $\geqslant 1$), the contribution of the last term in Eq. (19) is negligibly small. In our numerical results, the contribution from this term is ignored.

Assuming that the material obeys the plastic stress-strain relation given in Eq. (1), and applying J_2-flow theory, we can obtain the relation of separation stress versus the displacement at $x=\pm h_0$ following the same steps as in Eq. (13) through Eq. (15), as

$$\frac{\sigma_s}{\sigma_Y}=\frac{1}{Q}\int_V f(\varepsilon_e)\dot{\varepsilon}_e dV. \tag{20}$$

The integration in Eq. (20) can be evaluated by a change of variables from Eulerian to Lagrangian coordinates, i.e., to the initial (undeformed) coordinates. The corresponding relations between the two sets of variables can be obtained by solving the differential equations in Eq. (17).

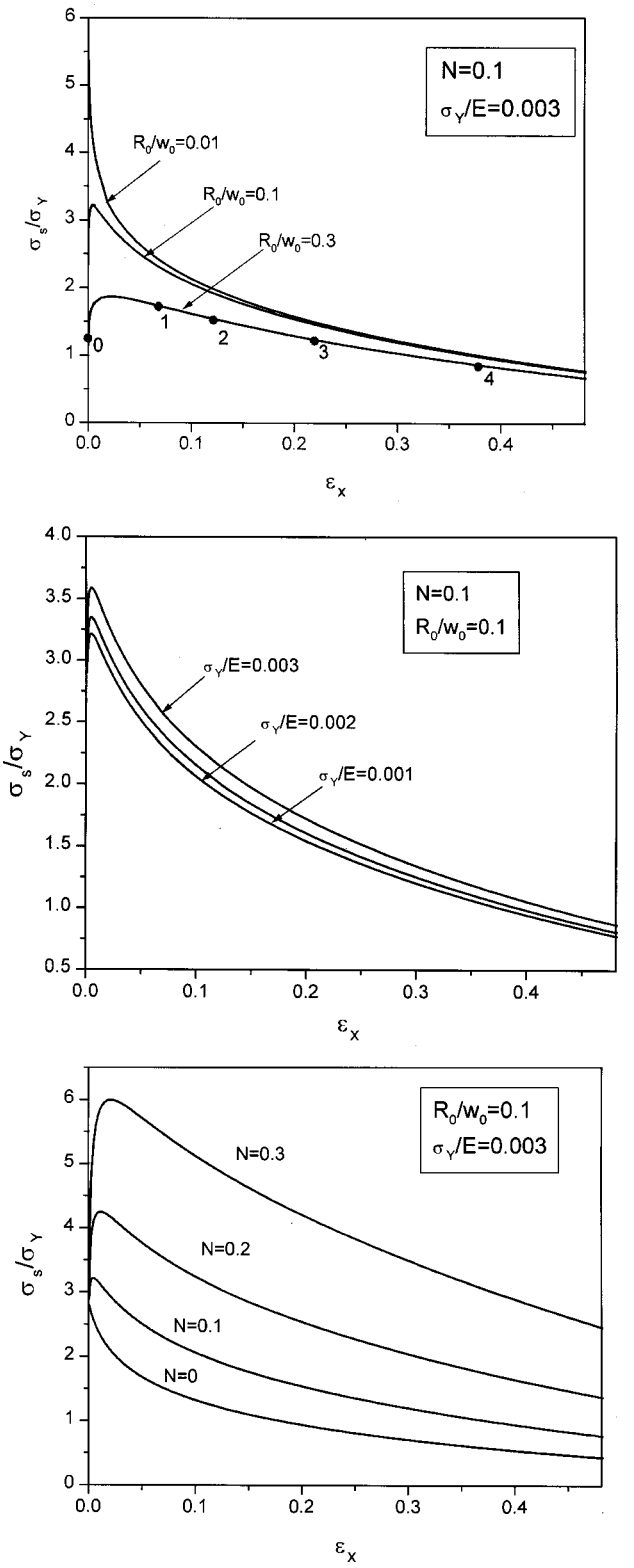


Fig. 6 Stress-separation relations for two-dimensional void growth; (*a*) effects of the initial radius of cavity, (*b*) effects of the material constant σ_Y/E, (*c*) effects of the strain-hardening exponent *N*

Results of the stress-separation curves are presented in Figs. 6(*a*)–(*c*). In these calculations, void spacing is taken to be the

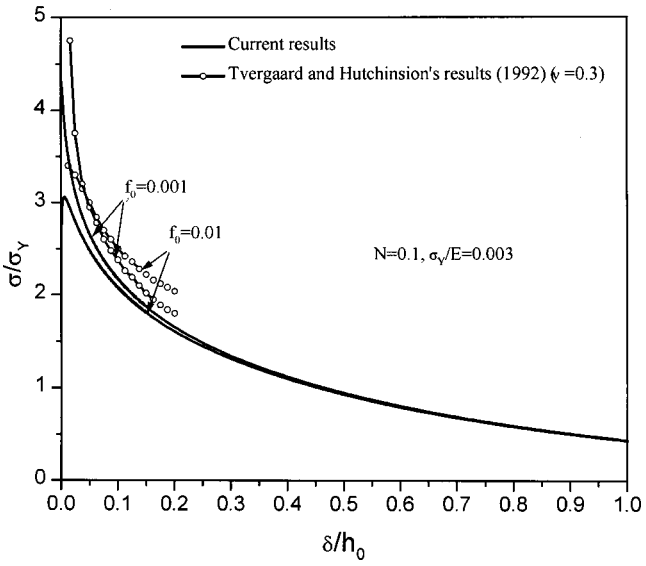


Fig. 7 Comparison of the stress-separation curves with Tvergaard and Hutchinson's calculations

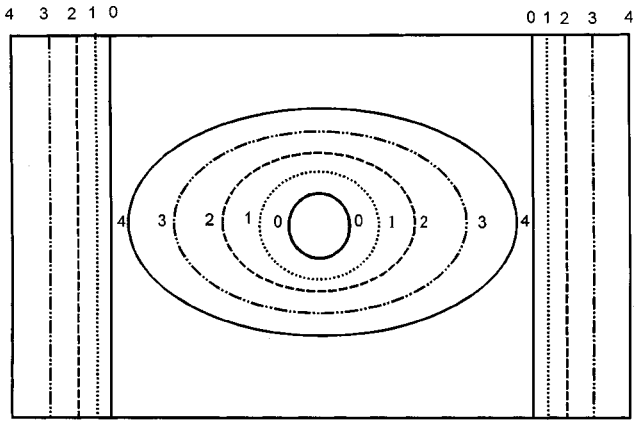


Fig. 8 Evolution of void shape for the two-dimensional case

same as the initial thickness of the layer, i.e., $h_0/w_0=1.0$. Figure 6(*a*) shows the separation stress normalized by the yield stress versus the average logarithmic strain of the cell, $\varepsilon_x=\ln(1+\delta/h_0)$, for different initial cavity size, a prescribed strain-hardening exponent N and normalized yield strength σ_Y/E. Figures 6(*b*) and (*c*) depict the effects of the material yield strength σ_Y/E and the strain-hardening exponent N on the stress-separation curves, respectively. The same trends as those in Fig. 2 are obtained. Figure 6(*a*) shows that, for small initial void size, peak separation stress (i.e., cavitation instability under the load-controlled loading condition) is reached rapidly at a very small strain level with a high maximum separation stress. For example, when $R_0/w_0=0.01$, $\sigma_s|_{\max}=5.45\sigma_Y$. On the other hand, larger initial void sizes result in lower peak stresses reached at much larger strains. The stress-separation curves decay monotonically beyond its peak stress as the void expands, leaving a narrow neck region between adjacent voids. Figures 6(*b*) and 6(*c*) again show that the stress separation curves are insensitive to σ_Y/E but rather sensitive to hardening exponent N.

Figure 7 shows the comparison of the stress-separation curves between the current result and that obtained by Tvergaard and Hutchinson [8] based on the Gurson model for elastoplastic material. Here f_0 is the area fraction of voids in the initial, undeformed state ($f_0=\pi R_0^2/4w_0h_0$), ν is Possion's ratio, δ is the separation displacement at the interface. All the parameters in our calculations are identical to those used by Tvergaard and Hutchinson [8] except for the Possion's ratio since there is no elastic deformation in our model. The Possion's ratio in Tvergaard and Hutchinson's calculation is 0.3 while in the current model it is 0.5 (rigid-plastic material). The absence of elastic deformation is likely the reason why the current model predicts a lower peak stress at a smaller displacement level and a somewhat lower separation stress than theirs, as shown in Fig. 7. Nevertheless, the two sets of curves in Fig. 7 exhibit general agreement. Due to the limitations of their finite element method, Tvergaard and Hutchinson terminated their calculations at a much lower separation displacement level than what we did using our model.

It is of interest to examine the shape evolution of cavities as they grow. Generally, an initially circular cavity becomes elliptical as it grows. Although the exact shape can be obtained by following the displacement of each material point on the boundary of the cavity, here we schematically depict the shape evolution by drawing an ellipse based on the long and short axes. Figure 8 shows an example of the evolution of the void shape under the prescribed parameters $R_0/w_0=0.3$, $N=0.1$, and $\sigma_Y/E=0.003$, where the numbers 0,1, . . . ,4 represent different instants of time during cavity growth. The corresponding instants are shown in Fig. 6(*a*). It is clear from Fig. 8 that, based on this model, the cavity mainly grows in the layer-thickness direction. This feature may be an artifact of the model since the mechanism of necking of ligaments is not accounted for.

4 Fracture Resistance

For purely brittle fracture of crystalline materials, the fracture toughness can be obtained by integrating the stress-separation curve at the atomic level. For a ductile material undergoing fracture due to plasticity induced cavitation, the fracture toughness can be evaluated by integrating an equivalent stress-separation curve, such as those obtained in the previous section, at a much larger length scale—the microscopic level. In this section we evaluate the fracture resistance of a sandwich structure exhibiting plasticity-induced cavitation in the thin ductile layer. Only the two-dimensional case will be considered in this section since the thickness of ductile layer and cavity spacing are ambiguously defined for the spherically symmetric cases.

The separation curves obtained in the previous sections are sensitive to the initial cavity size and to the hardening exponent, but they are nearly independent of the ratio of the yield strength to Young's modulus. Therefore, the effect of that ratio may be neglected. A universal expression of the stress-separation curve can be written as

$$\sigma_s(\delta)/\sigma_Y=F(N,V_f,G,\delta) \tag{21}$$

where N is the hardening exponent, δ is the separation displacement, V_f denotes the density of the voids along the interface, and G is a geometrical parameter related to the thickness of the ductile layer and the spacing between cavities. In the two-dimensional plane strain case, the initial density of cavities $V_f=R_0/w_0$, and the geometrical parameter G is the ratio of ductile layer thickness to the spacing between cavities h_0/w_0.

As pointed out by Tvergaard and Hutchinson [8,9], in evaluating the amplification of the fracture resistance due to plastic deformation, the work of separation per unit area (the initial separation resistance), Γ_0, and the ratio of peak separation stress to yield stress are two important parameters. The latter can be determined readily from the separation curves. The work of separation for the unit cell is

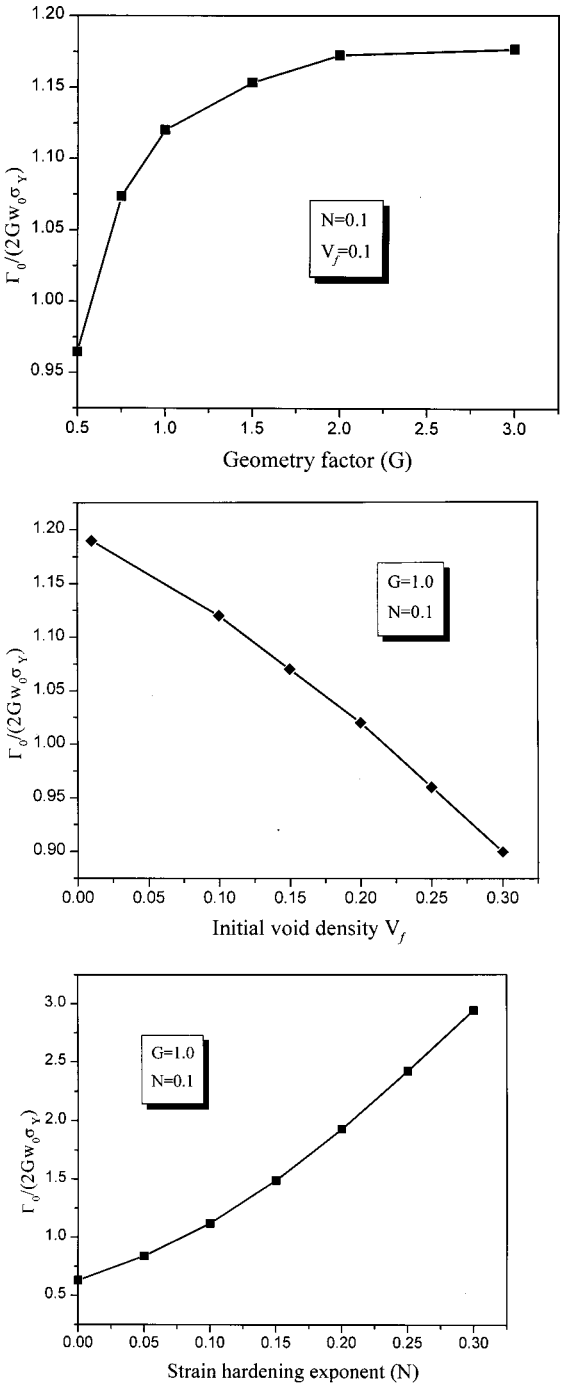


Fig. 9 (*a*) Normalized initial fracture resistance versus the geometrical parameter *G*, (*b*) normalized initial fracture resistance versus the void volume fraction V_f, (*c*) normalized initial fracture resistance versus the strain-hardening exponent *N*

$$W = 2\int_0^{U_1^*} T_s(\delta)\,d\delta \tag{22}$$

where U_1^* is the separation displacement level at which the total separation occurs ($\sigma_s \cong 0$), and T_s is the resultant force at the boundary of the cell. The initial separation resistance Γ_0 can then be expressed as

$$\Gamma_0 = W/A = 2\sigma_Y \int_0^{U_1^*} [\sigma_s(\delta)/\sigma_Y]\,d\delta \tag{23}$$

where the exposed surface area $A = 2w_0 t_0$, and t_0 is the thickness in the out-of-plane direction. In order to use the separation curves obtained in the previous section, Γ_0 may be rewritten as

$$\frac{\Gamma_0}{2w_0\sigma_Y} = G\int_0^{\varepsilon^*} [\sigma_s(\varepsilon)/\sigma_Y]\exp(\varepsilon)\,d\varepsilon \tag{24}$$

where ε^* denotes the true strain at the boundary of the unit cell corresponding to the displacement U_1^*. The functional form of $\sigma_s(\varepsilon)/\sigma_Y$ can be obtained from the stress-separation curves in the previous section. Since the separation curves are nearly independent of the material constant σ_Y/E, the normalized initial fracture resistance should then be nearly independent of σ_Y/E.

Equation (24) shows that the value of the normalized work of separation per unit area is determined by the integral on the right-hand side only. The integration can be carried out for given values of N, V_f, and G. The dependence of the normalized work of separation on the geometrical parameter, G, is shown in Fig. 9(*a*). For strain-hardening exponent $N = 0.1$ and the void density $V_f = 0.1$, the normalized initial fracture resistance $\Gamma_0/2w_0\sigma_Y$ has a rather weak dependence on $G(=h_0/w_0)$. The value of the normalized Γ_0 changes only from 0.95 to 1.17 when G varies from 0.5 to 3.0. Furthermore, when the value of G is larger than about 2.0, the normalized initial fracture resistance reaches an asymptotic value independent of G.

The dependence of the normalized fracture resistance $\Gamma_0/2w_0\sigma_Y$ on the hardening exponent and on the initial void density is shown in Figs. 9(*b*)–(*c*). They demonstrate that Γ_0 is very sensitive to the strain-hardening exponent, N, as shown in Fig. 9(*c*), but moderately sensitive to the void density, V_f, as shown in Fig. 9(*b*). For given void density V_f, as N changes from the nonhardening case ($N=0$) to a strong hardening case ($N=0.3$), the value of Γ_0 increases nearly sixfold. However, for given N, Γ_0 decreases moderately as V_f increases.

From Fig. 9(*a*)–(*c*) we can see that the value of the normalized Γ_0 for the two-dimensional case is in the range of 0.5–1.75 for $N=0-0.2$. This is higher than the values (0.35–0.82) predicted by Tvergaard and Hutchinson [8]. Tvergaard and Hutchinson abruptly terminated their calculations when the void area fraction equals 0.2, and neglected the contribution of the stress-separation curve beyond that point. This undoubtedly results in a lower work of fracture than that predicted by our calculations.

5 Concluding Remarks

In the present paper, we have developed a new technique to solve solid mechanics problems involving large plastic deformation for which closed-form solutions are difficult or sometimes impossible to obtain. The technique involves using the fluid flow field from existing fluid mechanics solutions to approximate the plastic flow field, and using the principle of virtual work to satisfy the equilibrium condition. It should be pointed out that the major difference between flow of fluids and plastic flow of solids is the following: In the potential flow of fluids, there is no shear stress; whereas in plastic flow of solids, the shear stress (the Mises equivalent stress in the present model) is constant for nonhardening materials or nearly constant for weakly hardening materials. In general, the method we developed here may be applied to a variety of solid mechanics problems, as long as the fluid mechanics solution is readily available.

This technique is used here to address a particular mechanism of interfacial fracture—cavitation in a thin ductile layer in a sandwich structure. Although this problem may be solved using numerical methods, as has indeed been done by Tvergaard [14] using the finite element method, the method developed here has the clear advantage when carrying out a parametric study involving extremely large deformation. Solving the two-dimensional cavity growth problem using the present method is much less time-consuming compared to, e.g., that using the finite element method with remeshing.

Two configurations have been considered in this paper: a spherically symmetric case for which a closed-form solid mechanics solution exists; and a two-dimensional, plane strain problem for which no closed form solution exists. In both cases, the potential fluid flow field from a point source is used to approximate the plastic deformation field surrounding the growing cavity in a representative unit cell. The current method gives the exact solution for the spherically symmetric cavity growth problem. For the two-dimensional plane strain cavity growth problem, reliable stress-separation relations have been obtained. The results show that the initial size of the void and the strain-hardening exponent are two important parameters that strongly affect the stress-separation curves. The stress-separation relations are also dependent on such geometrical parameters as the layer thickness and the void spacing. The interfacial fracture resistance due to this mechanism, evaluated by integrating the stress-separation curves, is strongly dependent on the strain-hardening exponent, is moderately dependent on cavity density, and is nearly independent on the geometrical parameter G.

In the present paper, void coalescence due to the necking of ligaments is not addressed. The interaction between adjacent voids is modeled by simply specifying the Neumann boundary conditions for each representative unit cell, i.e., the normal displacement at the boundaries of the unit cell between adjacent voids vanishes. Therefore the model is not accurate when necking instability and coalescence take place between adjacent voids. Furthermore, necking would result in accelerated separation of the fracture planes, leading to a rapid load drop in the stress-separation curves. Hence the final stage of the stress-separation curves predicted by the present model may not be very accurate. Fortunately, this stage of the separation process has only insignificant contributions to the overall fracture resistance.

There are several implications of the results by the present model. As shown in Fig. 9, as the void density increases, i.e., as the spacing between voids decreases, the fracture resistance decreases monotonically. This result implies that, in an adhesively bonded structure, a finer interfacial microstructure with more densely distributed void nucleation sites will give rise to a lower interfacial fracture toughness. However, a finer microstructure also means a smaller initial cavity size; thus it requires a higher peak stress to reach instability during void growth. Therefore, there could exist a preferred combination of pore size and pore spacing of the interfacial microstructure that will result in an optimal performance of the sandwich structure. Identification of such optimal combination will provide guiding principles for preparing the surfaces of structural components (e.g., aluminum panels) for adhesive bonding.

Acknowledgment

The work has been supported by the U. S. Department of Energy through Grant No. DEFG02-96ER45607. The authors are grateful for helpful discussions with Prof. H. Aref.

References

[1] Bowlin, D. T., Scheeline, A., and Pearlstein, A. J., 1997, ''Current Oscillations in Potentiostatic Electro-Oxidation of Aluminum in Sulfuric and Phosphoric Acids,'' Electrochim. Acta, **43**, pp. 417–421.
[2] Gao, H., Scheeline, A., and Pearlstein, A. J., 1999, ''Demonstration of a Novel Rotating Cylindrical Electrode in Growth of Oxide Films With Spatially Controlled Microstructural Variation on Al 6061,'' *Proceedings of the International Symposium on New Directions in Electroanalytical Chemistry II*, 1999, J. Leddy, P. Vanysek, and M. D. Porter, eds., The Electrochemical Society, Pennington, NJ, pp. 109–115.
[3] Ashby, M. F., Blunt, F. J., and Bannister, M., 1989, ''Flow Characteristics of Highly Constrained Metal Wires,'' Acta Metall., **37**, pp. 1847–1857.
[4] Evans, A. G., and Dalgleish, B. J., 1992, ''The Fracture Resistance of Metal-Ceramic Interfaces,'' Acta Metall. Mater., **40**, Suppl., pp. S295–S306.
[5] Turner, M. R., Dalgleish, B. J., He, M. Y., and Evans, A. G., 1995, ''A Fracture Resistance Measurement Method for Bimaterial Interfaces Having Large Debond Energy,'' Acta Metall. Mater., **43**, pp. 3459–3465.
[6] Turner, M. R., and Evans, A. G., 1996, ''An Experimental Study of the Mechanisms of Crack Extension Along an Oxide/Metal Interface,'' Acta Mater., **44**, pp. 863–871.
[7] He, M. Y., Evans, A. G., and Hutchinson, J. W., 1996, ''Interface Cracking Phenomena in Constrained Metal Layers,'' Acta Mater., **44**, pp. 2963–2971.
[8] Tvergaard, V., and Hutchinson, J. W., 1992, ''The Relation Between Crack Growth Resistance and Fracture Process Parameters in Elastic-Plastic Solids,'' J. Mech. Phys. Solids, **40**, pp. 1377–1397.
[9] Tvergaard, V., and Hutchinson, J. W., 1996, ''Effect of Strain-Dependent Cohesive Zone Model on Predictions of Crack Growth Resistance,'' Int. J. Solids Struct., **33**, pp. 3297–3308.
[10] McClintock, F. A., 1968, ''A Criterion for Ductile Fracture by the Growth of Holes,'' ASME J. Appl. Mech., **35**, pp. 363–371.
[11] Rice, J. R., and Tracey, D. M., 1969, ''On the Ductile Enlargement of Voids in Triaxial Stress Fields,'' J. Mech. Phys. Solids, **17**, pp. 201–217.
[12] Needleman, A., 1972, ''Void Growth in an Elastic-Plastic Medium,'' ASME J. Appl. Mech., **39**, pp. 964–970.
[13] Andersson, H., 1977, ''Analysis of a Model for Void Growth and Coalescence Ahead of a Moving Crack Tip,'' J. Mech. Phys. Solids, **25**, pp. 217–233.
[14] Tvergaard, V., 1997, ''Studies of Void Growth in a Thin Ductile Layer Between Ceramics,'' Comput. Mech., **20**, pp. 186–191.
[15] Huang, Y., Hu, K. X., Yeh, C. P., Li, N.-Y., and Hwang, K. C., 1996, ''A Model Study of Thermal Stress-Induced Voiding in Electronic Packages,'' ASME J. Electron. Packag., **118**, pp. 229–234.
[16] Huang, Y., Hutchinson, J. W., and Tvergaard, V., 1991, ''Cavitation Instabilities in Elastic-Plastic Solids,'' J. Mech. Phys. Solids, **39**, pp. 223–241.
[17] Tvergaard, V., and Hutchinson, J. W., 1993, ''Effect of Initial Void Shape on the Occurrence of Cavitation Instabilities in Elastic-Plastic Solids,'' ASME J. Appl. Mech., **60**, pp. 807–812.
[18] Hou, H.-S., and Abeyaratne, R., 1992, ''Cavitation in Elastic and Elastic-Plastic Solids,'' J. Mech. Phys. Solids, **40**, pp. 571–592.

Theory of Boundary Eigensolutions in Engineering Mechanics

A. R. Hadjesfandiari

G. F. Dargush
Mem. ASME

Department of Civil Engineering,
University at Buffalo,
State University of New York,
Buffalo, NY 14260

A theory of boundary eigensolutions is presented for boundary value problems in engineering mechanics. While the theory is quite general, the presentation here is restricted to potential problems. Contrary to the traditional approach, the eigenproblem is formed by inserting the eigenparameter, along with a positive weight function, into the boundary condition. The resulting spectra are real and the eigenfunctions are mutually orthogonal on the boundary, thus providing a basis for solutions. The weight function permits effective treatment of nonsmooth problems associated with cracks, notches and mixed boundary conditions. Several ideas related to the convergence characteristics are also introduced. Furthermore, the connection is made to integral equation methods and variational methods. This paves the way toward the development of new computational formulations for finite element and boundary element methods. Two numerical examples are included to illustrate the applicability. [DOI: 10.1115/1.1331059]

1 Introduction

Orthogonal functions in the form of trigonometric series and their generalizations have been used to solve boundary value problems since at least the 19th century with a rigorous mathematical foundation provided by Fourier, Dirichlet, and others. In this classical approach the basis functions are orthogonal over the problem domain. Detailed accounts of related concepts can be found in the works of Carslaw [1], Courant and Hilbert [2], Morse and Feshbach [3], Tolstov [4], and Lanczos [5].

However, it is clear from integral equation representations that the bounding surface is actually paramount in the solution of linear boundary value problems. Consequently, it is perhaps more appropriate to employ basis functions that are orthogonal over the boundary. We will see that these functions can be generated by solving an eigenproblem in which the eigenparameter appears in the boundary condition. The new concepts that emerge from this approach seem to have significance for the general theory of boundary value problems as well as for computational mechanics.

The approach is particularly well suited for nonsmooth problems, providing a unified treatment of such problems. A boundary value problem is considered nonsmooth if the boundary of the domain is nonsmooth or mixed boundary conditions are specified. In these cases, the solution is nonanalytic at some points on the boundary. Thus, the characteristic feature of these nonsmooth boundary value problems is the presence of singularities in the flux or higher order derivatives on the boundary. Since most of the mathematical problems posed in engineering mechanics involve either nonsmooth geometries or mixed boundary conditions, we attempt to provide a unified treatment that encompasses these nonsmooth problems.

Although the theory is applicable to a wide range of problems, we explore this idea within the context of potential theory. Thus, we are interested in the solution of the Laplace equation $\nabla^2 u = 0$ in domain V, subject to boundary conditions on S. The domain V can be two or three-dimensional, simply or multiply connected. The boundary S is a contour or set of contours in two-dimensional problems and a closed surface or surfaces for three-dimensional domains. Let q represent the normal boundary flux, where $q = \partial u / \partial n$ with n as the outward unit normal to S. Then either Dirichlet ($u = \overline{u}$ on S), Neumann ($q = \overline{q}$ on S), or mixed ($u = \overline{u}$ on S_u and $q = \overline{q}$ on S_q with $S_u \cup S_q = S$ and $S_u \cap S_q = \varnothing$) boundary conditions may be specified. As we know from the theory of boundary value problems, u is analytic in domain V. Although in our work, u is assumed to be continuous, it need not be analytic on boundary S. Furthermore, at nonsmooth points, q is not defined. In general, q is a piecewise continuous function on the boundary.

With this background in mind, we begin by defining the boundary eigenproblem in the following section. Further details on certain aspects can be found in Hadjesfandiari [6], which also includes the extension to the theory of elasticity.

2 Boundary Eigenproblem for Potential Theory

Consider the *boundary eigenproblem* for potential theory defined as follows: Find the nonzero function u such that in the domain V

$$\nabla^2 u = 0 \tag{2.1a}$$

and on the boundary S

$$q = \lambda \phi u \tag{2.1b}$$

where the parameter λ is an eigenvalue. Additionally, the *weight function* ϕ is integrable on S, but does not change sign. For simplicity, we take ϕ to be always positive on S. Note that this permits ϕ to be discontinuous or even singular at some points (i.e., piecewise continuous). However, from the *fundamental boundary condition* (2.1b), we note that now q is always continuous on the boundary whenever ϕ is continuous, even if there are geometrically nonsmooth points (e.g., edges, corners).

With the classical approach, the eigenparameter is introduced into the governing differential equation and a specific set of homogeneous boundary conditions are prescribed. However, in the boundary eigenproblem (2.1), the differential operator remains intact, while the eigenvalue is inserted into the boundary condition. This subtle difference has significant consequences. For example, the eigenfunctions associated with (2.1) are indeed harmonic in V, unlike the classical eigenfunctions which are actually solutions to a corresponding Helmholtz problem. Furthermore, the infinite sequence of eigenfunctions for (2.1) can be used as a basis for all

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, July 19, 1999; final revision, July 7, 2000. Associate Editor: B. Moran. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

solutions to boundary value problems in domain V governed by the Laplace equation with arbitrary well-defined boundary conditions on S.

Hilbert [7] considered the eigenproblem (2.1) with $\phi=1$. Courant and Hilbert [2], in a supplement to their work on vibration and eigenvalue problems, asserted that the properties of the eigensolutions are similar to those of other eigenproblems such as Sturm-Liouville. It seems, however, that the significance of the boundary eigenproblem in engineering mechanics has not been recognized.

The theory of pseudodifferential operators (e.g., [8,9]) is broad enough to encompass (2.1) as a special case. This provides the mathematical foundation and permits the introduction of a theorem defining the behavior of solutions.

THEOREM. *The boundary eigenproblem defined by (2.1) has the following properties*:

(i) At least one eigensolution (λ,u) exists.
(ii) All of the eigenvalues λ are real.
(iii) The infinite sequence of eigenfunctions u_n for $n=1,2,\ldots$ are boundary orthonormal with respect to ϕ, that is

$$\int_S \phi(x)u_m(x)u_n(x)dS(x)=\delta_{mn} \tag{2.2}$$

where δ_{mn} is the Kronecker delta.
(iv) All nonzero eigenvalues are positive.
(v) The infinite collection of eigenvalues form an increasing sequence, $\lambda_1\leqslant\lambda_2\leqslant\ldots\leqslant\lambda_n\leqslant\ldots$ such that λ_n become infinite for $n\to\infty$.
(vi) The system of eigenfunctions is complete.

Proof. It is easy to see that the equipotential solution $u=$constant is an eigenfunction corresponding to $\lambda=0$. By integrating both sides of the boundary condition (2.1*b*) over the surface and using Green's first theorem, we have for each eigensolution

$$\int_S q dS=0 \tag{2.3a}$$

and for $\lambda\neq 0$

$$\int_S \phi u dS=0. \tag{2.3b}$$

If λ is a complex number, say $\lambda=\alpha+i\beta$, u can be complex, say $u=v+iw$. It is easily seen that the complex conjugate of the eigensolution, $(\bar{\lambda},\bar{u})$ is also an eigensolution, where $\bar{\lambda}=\alpha-i\beta$ and $\bar{u}=v-iw$. By using the reciprocal theorem (Green's second identity) for u and $\bar{u}$, we obtain

$$\int_S (u\bar{q}-\bar{u}q)dS=0.$$

Substituting the fundamental boundary condition (2.1*b*) along with its complex conjugate produces

$$(\bar{\lambda}-\lambda)\int_S \phi\bar{u}u dS=0.$$

However, ϕ is positive on S, u is not zero everywhere, and therefore the surface integral is a positive number. We conclude that $\bar{\lambda}-\lambda=0$ and therefore λ is real, as asserted in item (ii). If the eigenfunction u is complex, then its real and imaginary parts are both eigensolutions. Thus, we may choose to take only real eigenfunctions.

Let (λ_1,u_1) and (λ_2,u_2) be two different eigensolutions with $\lambda_1\neq\lambda_2$. By applying the reciprocal theorem for u_1 and u_2

$$\int_S (u_1q_2-u_2q_1)dS=0.$$

Using the fundamental boundary conditions

$$(\lambda_2-\lambda_1)\int_S^{\phi} u_1u_2dS=0.$$

However, since λ_1 and λ_2 are different eigenvalues, we have

$$\int_S^{\phi} u_1u_2dS=0$$

and item (iii) is proved for distinct eigenvalues. Gram-Schmidt orthogonalization can be used for eigenfunctions associated with nondistinct eigenvalues. Additionally, (2.3*b*) is now seen as a special case of orthogonality with respect to the constant eigenfunction corresponding to $\lambda=0$.

Next we find an expression for the eigenvalue λ. Multiplying both sides of (2.1*b*) by u and integrating over the boundary, we obtain

$$\lambda=\frac{\int_S qudS}{\int_S \phi u^2dS}. \tag{2.4}$$

If we substitute $q=u_{,i}n_i$, then the numerator can be written

$$\int_S qudS=\int_S u_{,i}un_idS.$$

Using the divergence theorem

$$\int_S qudS=\int_V (uu_{,i})_{,i}dV \tag{2.5}$$

or

$$\int_S qudS=\int_V (u_{,i}u_{,i}+uu_{,ii})dV.$$

However, since $u_{,ii}=\nabla^2u=0$, we have

$$\int_S qudS=\int_V u_{,i}u_{,i}dV$$

and for λ in terms of the eigenfunction u, we obtain

$$\lambda=\frac{\int_V u_{,i}u_{,i}dV}{\int_S \phi u^2dS}. \tag{2.6}$$

For any nonzero eigenvalue, the numerator and denominator of (2.6) are both positive. Therefore, the eigenvalue is positive as specified in item (iv). The properties listed as items (v) and (vi) can be inferred via analogy with related eigenproblems, such as the Sturm-Liouville problem. However, rigorous proof is still needed. Additional mathematical concepts, such as those provided by Hilbert-Schmidt theory, may be appropriate.

Based upon the characteristics of the boundary eigenproblem, we can express the potential u as an infinite series of boundary eigensolutions u_n

$$u=\sum_{n=1}^{\infty} A_nu_n \quad \text{in } V\cup S. \tag{2.7}$$

Multiplying both sides with ϕu_m where u_m is also an eigensolution and integrating on the boundary, we obtain

$$\int_S \phi uu_mdS=\sum_{n=1}^{\infty}\int_S A_n\phi u_nu_mdS.$$

Due to orthogonality all of the terms are zero, except the one corresponding to A_m. Then

$$\int_S \phi u u_m dS = A_m \int_S \phi u_m^2 dS$$

and the coefficient A_m can be written

$$A_m = \frac{\int_S \phi u u_m dS}{\int_S \phi u_m^2 dS}. \tag{2.8}$$

The A_m represent generalized Fourier coefficients and are here called the fundamental coefficients. If the eigenfunctions are orthonormal, then obviously

$$A_m = \int_S \phi u u_m dS. \tag{2.9}$$

As mentioned previously, the potential u is assumed continuous. Consequently, the fundamental expansion converges uniformly to u in the domain V and boundary S. In the potential problem, u is analytic in the domain V. Thus, the first (and higher order) derivatives of u converge uniformly in the domain V. On the other hand, u is not necessarily analytic on the boundary S.

However, the completeness of the set of boundary eigenfunctions enables the representation of any mean square integrable function w defined on S (i.e., $\int_S \phi w^2 dS < \infty$) as an infinite series $\Sigma c_n u_n$ where $c_n = \int_S \phi w u_n dS$. The series converges in the mean to w on S. The value of the series is w, wherever w is continuous. Otherwise, the series converges to the principal mean value ([6]).

Therefore, the infinite series in (2.7) converges in the mean to u on S. Now what can be said concerning the fundamental expansion of q? We write

$$q = \frac{\partial u}{\partial x_i} n_i = \sum_{n=1}^{\infty} A_n \frac{\partial u_n}{\partial x_i} n_i. \tag{2.10}$$

However,

$$q_n = \frac{\partial u_n}{\partial x_i} n_i = \lambda_n \phi u_n.$$

Then

$$q = \phi \sum_{n=1}^{\infty} A_n \lambda_n u_n \quad \text{on } S. \tag{2.11}$$

Next we can define the *weighted flux* q^ϕ, where

$$q = \phi q^\phi. \tag{2.12}$$

Therefore,

$$q^\phi = \sum_{n=1}^{\infty} A_n \lambda_n u_n \quad \text{on } S. \tag{2.13}$$

If the function q^ϕ is piecewise continuous on the boundary S, then the fundamental expansion (2.13) converges in the mean. Besides, (2.13) converges uniformly to q^ϕ in every closed set on S containing no discontinuity. Conversely, the Nth partial sum of the fundamental expansion of q^ϕ

$$q_N^\phi = \sum_{n=1}^{N} A_n \lambda_n u_n \quad \text{on } S \tag{2.14}$$

cannot approach the function $q^\phi(x)$ uniformly over any set containing a point or line of discontinuity of q^ϕ. If q^ϕ is piecewise regular, then this is a generalized form of Gibbs' phenomenon. Further ideas concerning convergence behavior are discussed in Hadjesfandiari [6].

In the theory above, we considered ϕ to be positive everywhere on S. We should emphasize this property is a necessary condition for having a complete set of eigenmodes for representing all given potential problems. If instead ϕ is only non-negative on some parts of the boundary, then the eigenmodes follow all of the previously mentioned properties, but are complete for representing only those problems for which $q = 0$ on segments corresponding to zero values of ϕ.

3 Integral Equation Method

As is well known, every boundary value problem can be transformed into an integral equation. For the direct integral equation method

$$c(\xi)u(\xi) + \int_S F(x,\xi)u(x)dS(x) = \int_S G(x,\xi)q(x)dS(x) \tag{3.1}$$

where kernel $G(x,\xi)$ is the potential at point ξ generated by a unit source at boundary point x. Thus,

$$G(x,\xi) = \begin{cases} \dfrac{1}{2\pi} \ln \dfrac{1}{r} & \text{in two dimensions} \\ \dfrac{1}{4\pi r} & \text{in three dimensions} \end{cases}$$

where r is the distance between points x and ξ. Meanwhile

$$F(x,\xi) = \frac{\partial G(x,\xi)}{\partial n(x)}.$$

For boundary point ξ, the integral on the left-hand side of (3.1) is considered as a Cauchy principal value. If ξ is on a smooth boundary, then $c(\xi) = 1/2$. In the more general case, including geometrically nonsmooth points

$$c(\xi) = -\int_S F(x,\xi)dS(x).$$

By substituting the fundamental boundary condition (2.1*b*) into (3.1), we obtain the boundary eigenproblem in integral form

$$c(\xi)u(\xi) + \int_S F(x,\xi)u(x)dS(x) = \lambda \int_S G(x,\xi)\phi(x)u(x)dS(x). \tag{3.2}$$

This is an integral representation of (2.1). The solution of (3.2) has all of the characteristics defined in Section 2. The eigensolutions of (3.2) are real, with non-negative eigenvalues and boundary orthogonal eigenfunctions. Consequently, the spectrum of the direct integral equation representation of the potential problem is real for every positive, integrable boundary weight function ϕ, and the eigenfunctions form an orthogonal set. It seems that a spectrum analysis of the direct integral equation has not appeared before in the literature.

We should not forget that the direct integral equation (3.1) is derived from the reciprocal theorem between the singular fundamental solution and the potential u. Recall that the reciprocal theorem also played a key role in the theory of boundary eigensolutions.

Furthermore, we can introduce the *weighted flux* q^ϕ, where $q(x) = \phi(x)q^\phi(x)$. Then, (3.1) can be rewritten

$$c(\xi)u(\xi) + \int_S F(x,\xi)u(x)dS(x) = \int_S G(x,\xi)\phi(x)q^\phi(x)dS(x). \tag{3.3}$$

In nonsmooth problems involving flux singularities, the weight function $\phi(x)$ can be chosen to capture the asymptotic behavior

of the flux near the singular point. The integral equation (3.3) then involves only bounded solution variables $u(x)$ and $q^{\phi}(x)$. In other words, q^{ϕ} is a piecewise regular function.

In a practical sense for engineering applications, we may wish to solve discretized versions of (3.2) and (3.3) by using the boundary element method (e.g., [10]). Numerical solution of (3.2) allows us to study the character of the discretized integral equation representation of the potential problem, while the computational algorithms associated with (3.3) permit the direct solution of nonsmooth boundary value problems. Numerical examples of this latter approach are included in Section 7. Further details concerning the boundary element formulations and implementations are provided in Hadjesfandiari and Dargush [11].

4 Variational Method

Consider the functional $\lambda[u]$ defined as follows:

$$\lambda[u]=\frac{\int_V u_{,i}u_{,i}dV}{\int_S \phi u^2 dS}. \tag{4.1}$$

This is the Rayleigh quotient associated with the eigenproblem (2.1). We can see from (2.6) that for any boundary eigensolution, say (λ_n, u_n), the functional $\lambda[u_n]=\lambda_n$. Furthermore, it is easy to show that the Rayleigh quotient is extremum for boundary eigenfunctions. Taking the first variation of $\lambda[u]$ from (4.1), we obtain

$$\delta\lambda=\frac{2\int_V \frac{\partial u}{\partial x_i}\delta\frac{\partial u}{\partial x_i}dV\int_S \phi u^2 dS-2\int_V \frac{\partial u}{\partial x_i}\frac{\partial u}{\partial x_i}dV\int_S \phi u\,\delta u dS}{\int_S \phi u^2 dS^2}.$$

Substituting (4.1) again produces

$$\delta\lambda=2\frac{\int_V \frac{\partial u}{\partial x_i}\delta\frac{\partial u}{\partial x_i}dV-\lambda\int_S \phi u\,\delta u dS}{\int_S \phi u^2 dS}$$

or

$$\delta\lambda=2\frac{\int_V \frac{\partial}{\partial x_i}\left(\frac{\partial u}{\partial x_i}\delta u\right)dV-\int_V \nabla^2 u\,\delta u dV-\lambda\int_S \phi u\,\delta u dS}{\int_S \phi u^2 dS}.$$

Then using the divergence theorem

$$\delta\lambda=2\frac{\int_S \frac{\partial u}{\partial n}\delta u dS-\int_V \nabla^2 u\,\delta u dV-\lambda\int_S \phi u\,\delta u dS}{\int_S \phi u^2 dS}$$

or

$$\delta\lambda=2\frac{\int_S (q-\lambda\phi u)\,\delta u dS-\int_V \nabla^2 u\,\delta u dV}{\int_S \phi u^2 dS}.$$

Now δu is an arbitrary variation in the domain and on the boundary. For an extremum of $\lambda[u]$

$$\delta\lambda=0, \tag{4.2}$$

and we must have

$$\nabla^2 u=0 \quad \text{in } V$$

and

$$q=\lambda\phi u \quad \text{on } S.$$

Therefore, every eigenfunction of the boundary eigenproblem (2.1) extremizes the Rayleigh quotient (4.1), and the value of this quotient equals the eigenvalue corresponding to the specified eigenfunction.

Variational methods can, of course, be used to formulate finite element methods (e.g., Bathe [12]). In Hadjesfandiari and Dargush [11] a discretized version of (4.1) is used to develop a finite element formulation for the boundary eigenproblem. Furthermore, the above variational framework leads to the development of a flux-oriented finite element method that has some distinct advantages over existing approaches for the solution of general smooth and nonsmooth boundary value problems. Details of this finite element formulation and the associated numerical implementation are also presented in Hadjesfandiari and Dargush [11].

5 Boundary Eigensolutions as Basis for Boundary Value Problems

As noted previously, the boundary eigensolutions can be used as a basis for solutions to general boundary value problems (BVPs) in potential theory. The three primary boundary value problem types (Dirichlet, Neumann, and mixed) are considered below. In all cases, the potential u and normal flux q are defined by the series (2.7) and (2.11), respectively.

Dirichlet Problem. Assume the value of u is prescribed everywhere on the boundary such that $u=f(x)$ on S, where $f(x)$ is an integrable continuous function. Using (2.9), we obtain the fundamental coefficients as

$$A_m=\int_S \phi f u_m dS \tag{5.1}$$

assuming orthonormalized eigensolutions.

Neumann Problem. Assume the value of q is prescribed everywhere on the boundary such that $q=g(x)$ on S, where $g(x)$ is a piecewise continuous function satisfying the Gauss condition

$$\int_S g(x)dS=0. \tag{5.2}$$

Multiplying both sides of (2.11) with u_m and integrating over the boundary, we obtain

$$\int_S g u_m dS=\sum_{n=1}^{\infty}\lambda_n\int_S A_n\phi u_n u_m dS.$$

All of the terms on the right-hand side except the one involving A_m are zero. Therefore

$$\int_S g u_m dS=\lambda_m A_m\int_S \phi u_m^2 dS$$

and the fundamental coefficients become

$$A_m=\frac{1}{\lambda_m}\int_S g u_m dS \qquad \text{for} \quad m\neq 1 \tag{5.3}$$

assuming orthonormalized eigensolutions. The coefficient A_1, corresponding to $\lambda_1=0$, is undetermined due to the character of the Neumann potential problem.

Mixed Problem. In this case, the value of q is specified on some portion of the boundary and the value of u is specified on the rest of the boundary. In Section 2, we classified this problem as nonsmooth, along with general problems involving corners, notches, and cracks. The common feature of all of these problems is singularity of the solution. Most practical engineering problems are of this type. We may still use the relationships inherent in

(5.1) and (5.3), but we cannot obtain an uncoupled closed-form solution for the fundamental coefficients. This case is related to methods such as dual series equations for bounded domains and dual integral equations for unbounded domains ([13]), the Hilbert problem ([14]), and the Wiener-Hopf technique ([15]). We can, however, obtain a numerical solution of mixed problems. This is addressed more fully in Hadjesfandiari and Dargush [11] where new boundary element and finite element formulations are introduced.

6 Some Boundary Eigensolutions in Closed Form

In order to obtain a better understanding of the nature of the boundary eigensolutions, some problems are now solved in closed form. These solutions will be needed subsequently for comparative purposes when studying the performance of boundary element and finite element methods. Solutions of the boundary eigenproblem with $\phi=1$ are developed for the circle in the following subsection. Afterward, boundary eigensolutions, involving singular weight functions, are presented for an infinite wedge and for a circle with a notch. Additional closed-form boundary eigensolutions for an annulus and a sphere are provided in Hadjesfandiari [6] along with an application to conformal mapping.

Circle. We first consider a circle with radius a. By using the separation of variables method, the potential may be written in polar coordinates as

$$u=A_0+\sum_{n=1}^{\infty}\,[A_n\cos(n\theta)+B_n\sin(n\theta)]r^n. \tag{6.1}$$

Every term satisfies our boundary eigenproblem, as demonstrated below. On the circle $\partial u/\partial n=\partial u/\partial r$. Therefore,

$$\frac{\partial u}{\partial r}=\sum_{n=1}^{\infty}\,n[A_n\cos(n\theta)+B_n\sin(n\theta)]r^{n-1}.$$

It is obvious that A_0 is the eigenfunction corresponding to $\lambda=0$. For each additional term, we check $q=\lambda u$ on the boundary of the circle $r=a$. For each value of n

$$n[A_n\cos(n\theta)+B_n\sin(n\theta)]a^{n-1}$$
$$=\lambda[A_n\cos(n\theta)+B_n\sin(n\theta)]a^n$$

and the boundary eigenvalues are easily established as

$$\lambda_n=\frac{n}{a}\quad\text{where}\quad n=1,2,\ldots\,. \tag{6.2}$$

We see that for every eigenvalue, λ_n, there are two different eigenfunctions

$$u_n^{(1)}=r^n\cos(n\theta), \tag{6.3a}$$

$$u_n^{(2)}=r^n\sin(n\theta). \tag{6.3b}$$

The circle is a special case in which all of the nonzero eigenvalues have multiplicity two. The orthogonality and completeness of this set of eigenfunctions is well established. We are most interested in the property of boundary orthogonality, which in this case is also satisfied.

All of the eigenfunctions, except u_0, oscillate along the circumference and decay to zero at the center. As n increases, the u_n oscillate and decay more rapidly. Thus, for large n, the response is essentially confined to the near surface region. Notice also that if the weight function is instead selected as $\phi=1/a$, then the boundary eigenvalues are simply the integers.

Infinite Wedge. In the first example, closed-form solutions were obtained for the boundary eigenproblem with constant ϕ. Here we examine the problem of an infinite wedge occupying the domain $0<\theta<\alpha$ and introduce a variable weight function $\phi>0$ to provide a means for obtaining closed-form expressions for the eigensolutions.

Consider the analytic function $f(z)$ given by the following:

$$f(z)=e^{ikz^{\pi/\alpha}}$$

where k is an arbitrary real number. Alternatively, this can be written in terms of polar coordinates as

$$f(z)=e^{-kr^{\pi/\alpha}\sin(\pi\theta/\alpha)}[\cos(kr^{\pi/\alpha}\cos(\pi\theta/\alpha))$$
$$+i\sin(kr^{\pi/\alpha}\cos(\pi\theta/\alpha))]. \tag{6.4}$$

Both the real and imaginary parts of $f(z)$ satisfy the boundary eigenproblem with weight function

$$\phi=\frac{\pi}{\alpha}r^{(\pi/\alpha)-1}. \tag{6.5}$$

We demonstrate this for the real part which is renamed u, where

$$u=e^{-kr^{\pi/\alpha}\sin(\pi\theta/\alpha)}\cos(kr^{\pi/\alpha}\cos(\pi\theta/\alpha))$$

with corresponding flux

$$q_\theta=\frac{1}{r}\frac{\partial u}{\partial\theta}=-\frac{\pi}{\alpha}r^{(\pi/\alpha)-1}e^{-kr^{\pi/\alpha}\sin(\pi\theta/\alpha)}\cos(kr^{\pi/\alpha}\cos(\pi\theta/\alpha)).$$

By noticing that

$$q=\begin{cases}-q_\theta & \text{at}\quad\theta=0\\ +q_\theta & \text{at}\quad\theta=\alpha\end{cases}$$

we see that the fundamental boundary condition $q=\lambda\phi u$ is satisfied on the boundary, if the weight function is given by (6.5) and the eigenvalues are $\lambda=k$. A similar proof can be obtained for the imaginary part.

For the infinite wedge, the spectrum is continuous and the potential can be represented in terms of boundary eigenfunctions as

$$u=\int_0^{\infty}u_k\,dk \tag{6.6}$$

with

$$u_k=e^{-kr^{\pi/\alpha}\sin(\pi\theta/\alpha)}[A(k)\cos(kr^{\pi/\alpha}\cos(\pi\theta/\alpha))$$
$$+B(k)\sin(kr^{\pi/\alpha}\cos(\pi\theta/\alpha))]. \tag{6.7}$$

Once again with increasing k, these eigenfunctions decay more rapidly toward the interior of the domain.

It is interesting to note that for $\alpha=\pi$, the weight function $\phi=1$ and the potential becomes analytic in $V\cup S$. The solution for u then reduces to that for a semi-infinite domain. On the other hand, for $\alpha>\pi$, the weight function ϕ and the boundary flux q are singular, in agreement with the singularity present at notches and cracks. However, the weighted boundary flux q^ϕ, where $q=\phi q^\phi$, remains bounded and continuous.

Circle With Notch (Finite Wedge). For the previous two examples, boundary eigensolutions were derived with a positive weight function everywhere on the boundary. Here we examine a finite wedge occupying the domain $0<\theta<\alpha$ and $r<a$. We introduce $\phi=1/a$ on the circular arc and $\phi=0$ on the two sides corresponding to $\theta=0$ and $\theta=\alpha$, as shown in Fig. 1. Thus, the eigenfunctions must be compatible with $q=0$ on the sides.

Clearly the constant eigenfunction $u_0=1$ with eigenvalue $\lambda_0=0$ satisfies the boundary eigenproblem. Now consider the harmonic functions

$$u_n=\mathrm{Re}\{z^{n\pi/\alpha}\}=r^{n\pi/\alpha}\cos\left(\frac{n\pi}{\alpha}\theta\right)\qquad\text{for}\quad n=1,2,\ldots \tag{6.8}$$

with gradient

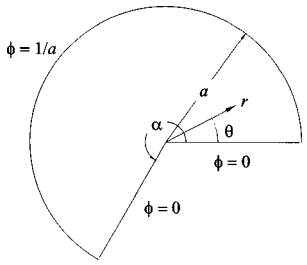


Fig. 1 Circle with notch—boundary eigenproblem definition

$$q_{\theta n}=\frac{1}{r}\frac{\partial u_n}{\partial \theta}=-\frac{n\pi}{\alpha}r^{(n\pi/\alpha-1)}\sin\left(\frac{n\pi}{\alpha}\theta\right)$$

$$q_{rn}=\frac{\partial u_n}{\partial r}=\frac{n\pi}{\alpha}r^{(n\pi/\alpha-1)}\cos\left(\frac{n\pi}{\alpha}\theta\right).$$

Notice that $q=-q_{\theta n}=0$ on $\theta=0$ and that $q=q_{\theta n}=0$ on $\theta=\alpha$. On the circular arc $r=a$, we have

$$q=q_{rn}=\frac{n\pi}{\alpha}a^{(n\pi/\alpha-1)}\cos\left(\frac{n\pi}{\alpha}\theta\right).$$

Consequently, the u_n defined in (6.8) are boundary eigenfunctions, with the corresponding eigenvalues given as

$$\lambda_n=\frac{n\pi}{\alpha}\quad \text{for}\quad n=1,2,\ldots. \tag{6.9}$$

We should emphasize that for $\alpha>\pi$, the gradients are singular at the tip of the notch for the first mode $n=1$. Furthermore, in this special example, the eigensolutions are exactly those that we can derive from a local analysis about the notch tip.

7 Application of Theory to Nonsmooth Boundary Value Problems

In this section, we consider the application of the theory of boundary eigensolutions to two nonsmooth potential problems. The first example utilizes the boundary eigensolutions just obtained to solve a boundary value problem for the circle with a notch. The solution is obtained as an infinite series of boundary eigensolutions. Results are also obtained numerically using a boundary element formulation based on the integral equation method of Section 3. A second example is then provided, involving potential flow in a square region with a diamond-shaped cutout for which an analytical solution is not possible.

Circle With Notch. Consider the Neumann problem illustrated in Fig. 2 with the following boundary conditions:

$$q=\begin{cases}+q_0 & \text{for}\quad \alpha_1<\theta<\alpha_2\\ -q_0 & \text{for}\quad \alpha-\alpha_2<\theta<\alpha-\alpha_1.\\ 0 & \text{elsewhere}\end{cases}$$

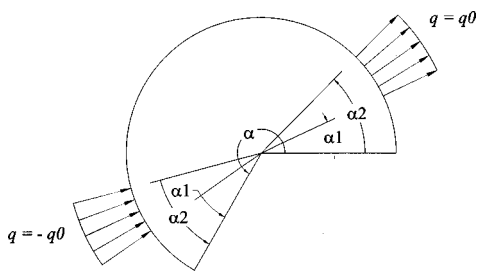


Fig. 2 Circle with notch—boundary value problem definition

Table 1 Circle with notch

Case	α (deg)	α_1 (deg)	α_2 (deg)	$q^{\phi}_{BE}(0)$	$K_{\mathrm{III}\mid BE}$	K_{III} Exact
Crack	360	67.5	112.5	0.351	0.880	0.881
Reentrant corner	270	22.5	67.5	0.571	1.43	1.43

Notice that the Gauss condition for equilibrium of flux (5.2) is satisfied.

Based on the theory of fundamental boundary eigenexpansion

$$u=\sum_{n=0}^{\infty}A_n u_n \tag{7.1}$$

where u_n are given by (6.8). From Section 5, the coefficients are established as

$$A_n=\frac{\int_S q u_n dS}{\lambda_n\int_S \phi u_n^2 dS}\quad \text{for}\quad n=1,2,\ldots. \tag{7.2}$$

Due to the character of the Neumann problem, the coefficient A_0 is undetermined. After carrying out the integrations in (7.2), we have

$$A_n=\frac{4q_0\alpha}{n^2\pi^2 a^{n\pi/\alpha-1}}\left[\sin\left(\frac{n\pi\alpha_2}{\alpha}\right)-\sin\left(\frac{n\pi\alpha_1}{\alpha}\right)\right]$$
$$\text{for}\quad n=1,3,5,\ldots \tag{7.3}$$

while $A_n=0$ for $n=2,4,\ldots$ due to the antisymmetric boundary conditions. Thus,

$$u=A_0+A_1 r^{\pi/\alpha}\cos\left(\frac{\pi\theta}{\alpha}\right)+\sum_{n=3,5}^{\infty}A_n r^{n\pi/\alpha}\cos\left(\frac{n\pi\theta}{\alpha}\right) \tag{7.4}$$

$$q_\theta=-\frac{\pi}{\alpha}A_1 r^{\pi/\alpha-1}\sin\left(\frac{\pi\theta}{\alpha}\right)-\sum_{n=3,5}^{\infty}\frac{n\pi}{\alpha}A_n r^{n\pi/\alpha-1}\sin\left(\frac{n\pi\theta}{\alpha}\right). \tag{7.5}$$

We are most interested in A_1 corresponding to $\lambda_1=\pi/\alpha$. In light of the tearing stress intensity factor (Mode III) in fracture mechanics, we can define a generalized flux intensity factor via the following:

$$K_{\mathrm{III}}=\lim_{r\to 0}\sqrt{2\pi}r^{1-\pi/\alpha}q_\theta(r,\theta=\alpha/2).$$

Then, for the present problem

$$K_{\mathrm{III}}=\sqrt{2\pi}\frac{\pi}{\alpha}A_1=4\sqrt{\frac{2}{\pi}}q_0 a^{1-\pi/\alpha}\left[\sin\left(\frac{\pi\alpha_2}{\alpha}\right)-\sin\left(\frac{\pi\alpha_1}{\alpha}\right)\right].$$

Now we use a boundary element method, based upon (3.3), to determine the generalized flux intensity factor at the notch tip numerically. Additional details can be found in Hadjesfandiari [6] and Hadjesfandiari and Dargush [11]. We model only half of the body and impose $u=0$ on the symmetry cut. Furthermore, from the closed-form eigenfunction, we also choose $\phi=r^{\pi/\alpha-1}$ along that symmetry boundary. We consider two specific cases involving a crack and a reentrant corner. Geometric and loading details are provided in Table 1. For each case, the model employs 30 quadratic boundary elements on the surface with a very fine mesh near the singular point.

Results obtained for the weighted flux q^{ϕ} along the symmetry cut are shown in Fig. 3 for $\alpha=2\pi$ (i.e., a crack). The oscillations in q^{ϕ} near the tip are the result of nonuniform convergence, and thus represent a generalized form of Gibbs' phenomenon. This is

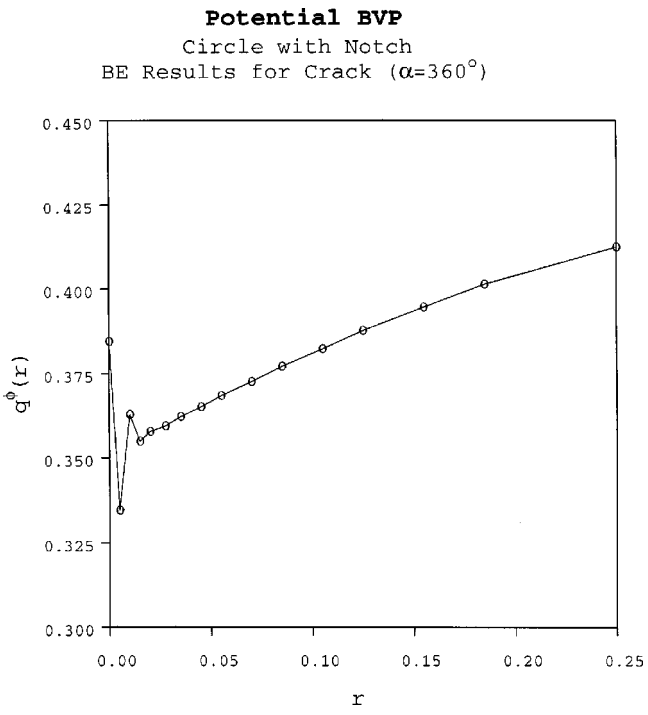


Fig. 3 Circle with notch—weighted flux versus distance from crack tip

discussed further in the next example. Here we use an extrapolation procedure to estimate the value of the generalized flux intensity factors, where

$$K_{\mathrm{III}|BE} = \lim_{r \to 0} \sqrt{2\pi} q^{\phi}(r).$$

A comparison of the boundary element results with the analytical solution is presented in Table 1 for both the crack and reentrant corner. The errors in the numerical solutions are approximately 0.1 percent.

It may seem that in order to solve any nonsmooth problem, we must have the boundary eigensolutions. Fortunately, we only need the asymptotic behavior near the singular point, which can be found from a local analysis. The next example shows this clearly.

Square With Diamond-Shaped Cutout. We examine the problem of potential flow in a square region with a diamond-shaped cutout illustrated in Fig. 4. Let $h=10$, $a=1$ and $b=2$. All

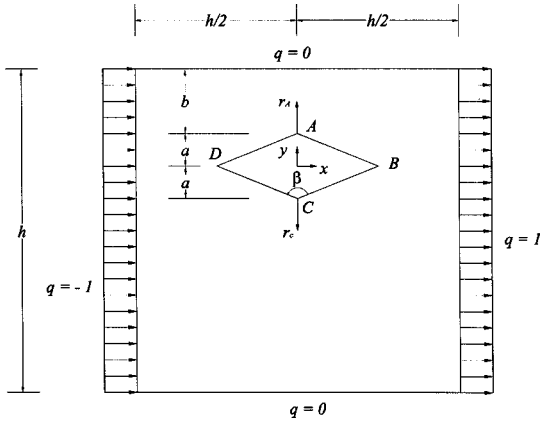
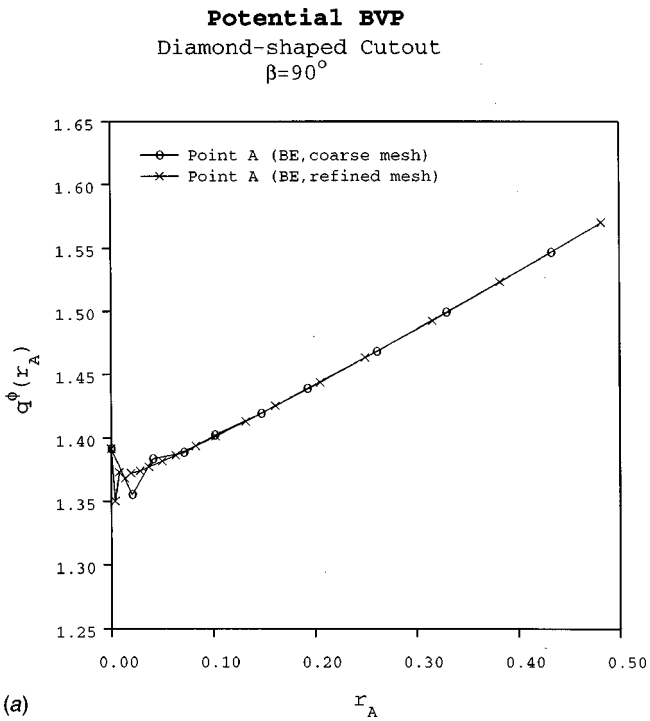


Fig. 4 Square with a diamond-shaped cutout—problem definition



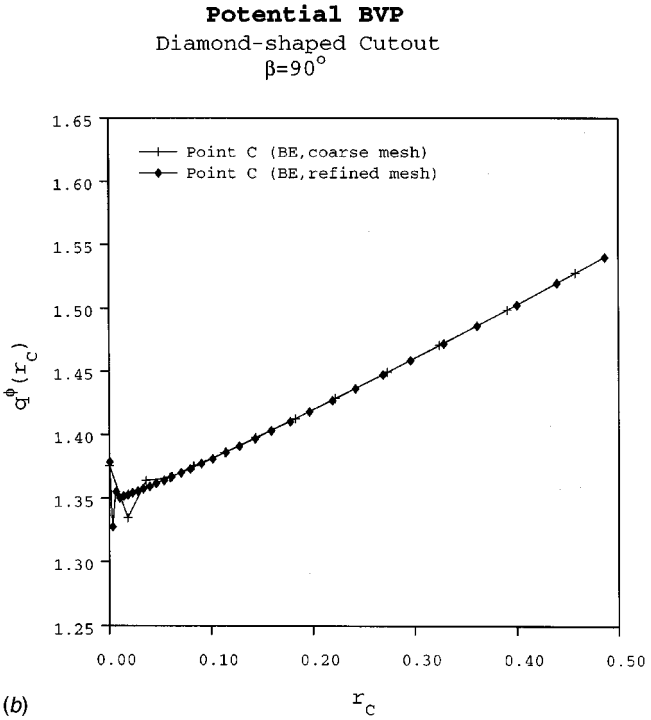


Fig. 5 Square with a diamond-shaped cutout—convergence of weighted flux versus distance from singular point

of the boundary conditions are specified in the diagram. Due to the nature of the cutout, singularities in flux occur at points $A-D$. Here we focus on the singularities at points A and C. We invoke symmetry conditions about $x=0$ and introduce the following positive weight function on that boundary:

$$\phi = \begin{cases} \phi_A = r_A^{\gamma-1} & \text{for } y > a \\ \phi_C = r_C^{\gamma-1} & \text{for } y < -a \end{cases}$$

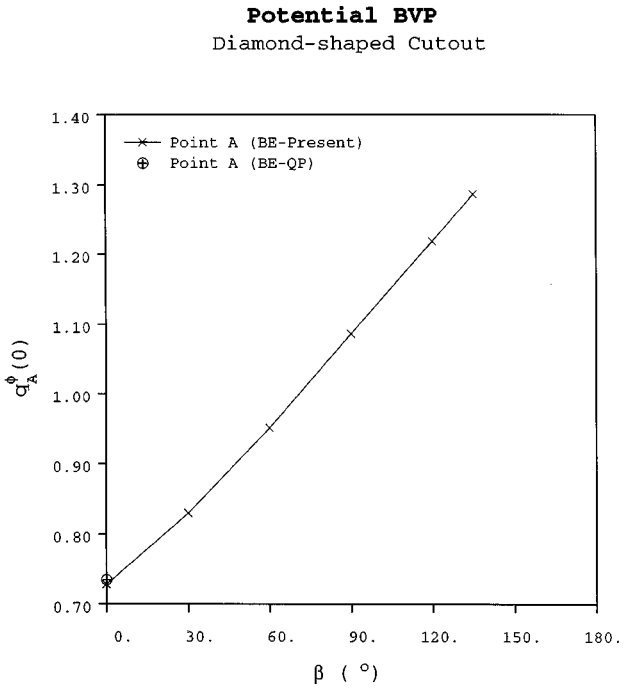


Fig. 6 Square with a diamond-shaped cutout—generalized flux intensity factor versus internal cutout angle

with $r_A = y - a$ and $r_C = -y - a$ as shown in the figure. From an analysis of local asymptotic behavior based upon the results of Section 6, $\gamma = \pi/(2\pi - \beta)$, where β is the internal cutout angle. Then, as in (2.12), let $q = \phi q^\phi$. Although q is singular at A and C, q^ϕ remains bounded, and the eigenproblem defined by (3.2) has a real spectrum and a set of boundary orthogonal eigenfunctions. These eigensolutions cannot be found in closed form; however, we only need their asymptotic behavior near the notch tips in order to solve the boundary value problem numerically.

Here the problem is again solved by employing a boundary element method based upon the integral equation (3.3). Due to symmetry, only the left half of the problem domain is modeled and $u = 0$ is enforced on the symmetry boundary $x = 0$. Two levels of mesh refinement are considered by using 51 and 81 quadratic boundary elements along the surface. The mesh is graded to include a finer discretization near the singular points.

Results obtained for q^ϕ with $\beta = \pi/2$ are displayed in Figs. 5(*a*,*b*). The solutions away from the singular points have converged even with a relatively coarse mesh. Oscillations in q^ϕ appear near the points A and C. Once again these oscillations are a generalized form of Gibbs' phenomenon due to nonuniform convergence. Notice that the amplitude of the oscillation is insensitive to the level of mesh refinement. This is the analog of the classical Gibbs' phenomenon in which the oscillation amplitude is independent of the number of terms included in a series representation of a function with a jump discontinuity. Here it is the infinite series of boundary eigenfunctions that has been truncated.

In practice, filtering techniques or extrapolation algorithms can be used to extract the generalized flux intensity factors. The latter approach is adopted to provide the results presented in Fig. 6 for the coarse boundary element mesh. The quantities $q_A^\phi(0)$ and $q_C^\phi(0)$ are related to generalized flux intensity factors that quantify the singularity at points A and C, respectively. Figure 6 illustrates the variation of $q_A^\phi(0)$ with internal cutout angle. In all cases, the flux q is infinite. However, by introducing the singular function ϕ, smooth variations of $q_A^\phi(0)$ and $q_C^\phi(0)$ with β are obtained. The limiting case, $\beta = 0$ represents a crack. Boundary element results obtained for this limiting case using quarter-point elements (e.g., [10]) are also shown in Fig. 6, and indicate very good correlation with the present approach.

8 Concluding Remarks

In this paper, we have explored the concept of boundary eigensolutions to boundary value problems. The resulting theory furnishes new insight into the solutions of BVPs. In addition, we find that there is a connection among the theory of boundary eigensolutions, integral equation methods and variational methods. In the domain of computational mechanics, this provides a relationship between boundary element methods and finite element methods.

Hilbert [7] has mentioned the boundary eigensolutions with $\phi = 1$ long ago, and has even given their relation with the calculus of variations. He did not notice the relation with the direct integral equation. This theory has been further developed here by introducing a general positive weight function ϕ, which then provides a unified treatment for nonsmooth problems in engineering mechanics and allows for meaningful solutions to be obtained. The simple numerical examples considered in Section 7, based on a boundary element formulation, illustrate the attractiveness of this new methodology.

References

[1] Carslaw, H. S., 1950, *An Introduction to the Theory of Fourier's Series and Integrals*, Dover, New York.
[2] Courant, R., and Hilbert, D., 1953, *Methods of Mathematical Physics*, John Wiley and Sons, New York.
[3] Morse, P. M., and Feshbach, H., 1953, *Methods of Theoretical Physics*, McGraw-Hill, New York.
[4] Tolstov, G. P., 1962, *Fourier Series*, Dover, New York.
[5] Lanczos, C., 1966, *Discourse on Fourier Series*, Oliver & Boyd, Edinburgh.
[6] Hadjesfandiari, A. R., 1998, ''Theoretical and Computational Concepts in Engineering Mechanics,'' Ph.D. dissertation, University at Buffalo, State University of New York, Buffalo, NY.
[7] Hilbert, D., 1912, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, B. G. Teubner, Leipzig.
[8] Shubin, M. A., 1987, *Pseudodifferential Operators and Spectral Theory*, Springer-Verlag, Berlin.
[9] Grubb, G., 1996, *Functional Calculus of Pseudodifferential Boundary Problems*, Birkhauser, Boston.
[10] Banerjee, P. K., 1994, *The Boundary Element Methods in Engineering*, McGraw-Hill, London.
[11] Hadjesfandiari, A. R., and Dargush, G. F., 2000, ''Computational Mechanics Based on the Theory of Boundary Eigensolutions,'' Int. J. Numer. Methods Eng., in press.
[12] Bathe, K. J., 1996, *Finite Element Procedures*, Prentice-Hall, Englewood Cliffs, NJ.
[13] Sneddon, I. N., 1966, *Mixed Boundary Value Problems in Potential Theory*, North Holland, Amsterdam.
[14] Muskhelishvili, N. I., 1953, *Some Basic Problems of the Mathematical Theory of Elasticity*, P. Noordhoff, Groningen, Holland.
[15] Noble, B., 1958, *Methods Based on the Wiener-Hopf Technique for the Solution of Partial Differential Equations*, Pergamon, New York.

A State-Space-Based Stress Analysis of a Multilayered Spherical Shell With Spherical Isotropy

W. Q. Chen[1]
e-mail: caijb@ccea.zju.edu.cn

H. J. Ding

Department of Civil Engineering,
Zhejiang University,
Hangzhou 310027, P. R. China

This paper presents an exact static stress analysis of a multilayered elastic spherical shell (hollow sphere) completely based on three-dimensional elasticity for spherical isotropy. Two independent state equations are derived after introducing three displacement functions and two stress functions. In particular, a variable substitution technique is used to derive the state equations with constant coefficients. Matrix theory is then employed to obtain the relationships between the state variables at the upper and lower surfaces of each lamina. By virtue of the continuity conditions between two adjacent layers, a second-order linear algebraic equation and a fourth-order one about the boundary variables at the inner and outer surfaces of a multilayered spherical shell are obtained. Numerical examples are presented to show the effectiveness of the present method.
[DOI: 10.1115/1.1343913]

1 Introduction

Spherical isotropy is a special kind of transverse isotropy that was introduced in 1865 by Saint-Venant, who gave an exact solution of a spherically isotropic spherical shell subjected to both internal and external uniform pressures ([1,2]). Hu [3] first initiated to use a separation method and presented a general theory of elasticity for a spherically isotropic medium. Many subsequently important analyses were inspired by and based on Hu's elegant method. For instance, Chen [4] utilized Hu's method ([3]) to investigate some static problems such as a concentrated force in an infinite medium, stress concentration due to a spherical cavity, and a steadily rotating shell. Puro [5] generalized the separation method to the inhomogeneous case. Shul'ga et al. [6] considered the free-vibration problem of a nonhomogeneous spherically isotropic spherical shell. Chau [7] recently extended Hu's formula to consider the toroidal vibration of a spherically isotropic solid sphere. In the monograph of Ding et al. [8], there is a detailed description on the coupled vibrations of spherically isotropic hollow spheres.

The interest of the study of spherically isotropic materials comes not only from the academic tradition, but also from the fact that they have been widely applied in aerospace and many other industries ([9,10]). More importantly, the latest investigation of geophysics showed that an appropriate model of the Earth should be a multilayered spherical shell including layers with spherical isotropy ([11]). Using such a model, Ding et al. [12] studied the effect of anisotropy on the tidal stress of the Earth.

The state-space-based method (also known as the method of initial function) is a powerful tool for solving problems of laminated structures ([13,14]). It can effectively reduce the order of the final solving matrix and greatly improve the computational precision. It is mentioned here that Shul'ga et al. [6] firstly presented two separated state equations with varying coefficients in spherical coordinates for spherically isotropic elasticity. However, they completely have overlooked the superiority of the state-space-based method in the analysis of laminated structures and they treated the two state equations just as the intermediate equations in simpler forms that were solved by a numerical method.

This paper presents two separated state equations by employing the separation formulae for displacements and shear stresses. A variable substitution method is then employed to transfer the resulting equations to the ones with constant coefficients. By employing the matrix theory and utilizing the continuity conditions at each interface, two relationships are obtained between the boundary variables at the inner and outer surfaces of a laminated spherical shell. The numbers of the final solving equations corresponding to the two separated state equations are only one and two, respectively, for a specified boundary value problem. Numerical example is given for a three-layered spherical shell subjected to an external distributed pressure.

2 Basic Equations

The basic equations of a spherically isotropic elastic body are well described in the monograph of Lekhnitskii [2] or more recently in the book of Ding et al. [8]. For the sake of the followed analysis, we give these equations in this section in a slightly different way. Assuming the center of the spherical isotropy coincident with the origin of spherical coordinates (r,θ,ϕ), the linear constitutive relations can be rewritten as follows:

$$\begin{cases} \Sigma_{\theta\theta}=r\sigma_{\theta\theta}=c_{11}S_{\theta\theta}+c_{12}S_{\phi\phi}+c_{13}S_{rr}, \\ \Sigma_{\phi\phi}=r\sigma_{\phi\phi}=c_{12}S_{\theta\theta}+c_{11}S_{\phi\phi}+c_{13}S_{rr}, \\ \Sigma_{rr}=r\sigma_{rr}=c_{13}S_{\theta\theta}+c_{13}S_{\phi\phi}+c_{33}S_{rr}, \\ \Sigma_{r\theta}=r\sigma_{r\theta}=2c_{44}S_{r\theta}, \\ \Sigma_{r\phi}=r\sigma_{r\phi}=2c_{44}S_{r\phi}, \\ \Sigma_{\theta\phi}=r\sigma_{\theta\phi}=2c_{66}S_{\theta\phi}, \end{cases} \tag{1}$$

where σ_{ij} is the stress tensor, c_{ij} are elastic constants, and the relation $c_{11}=c_{12}+2c_{66}$ holds for spherical isotropy. S_{ij} in Eq. (1) is the ''generalized strain tensor'' determined by

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, August 26, 1999; final revision, June 7, 2000. Associate Technical Editor: R. C. Benson. Discussion on the paper should be addressed to the Technical Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

$$\begin{cases} S_{rr}=rs_{rr}=\nabla_2 u_r, S_{\theta\theta}=rs_{\theta\theta}=\dfrac{\partial u_\theta}{\partial\theta}+u_r, \\ S_{\phi\phi}=rs_{\phi\phi}=\dfrac{1}{\sin\theta}\dfrac{\partial u_\phi}{\partial\phi}+u_r+u_\theta\cot\theta, \\ 2S_{r\theta}=2rs_{r\theta}=\dfrac{\partial u_r}{\partial\theta}+\nabla_2 u_\theta-u_\theta, \\ 2S_{r\phi}=2rs_{r\phi}=\dfrac{1}{\sin\theta}\dfrac{\partial u_r}{\partial\phi}+\nabla_2 u_\phi-u_\phi, \\ 2S_{\theta\phi}=2rs_{\theta\phi}=\dfrac{1}{\sin\theta}\dfrac{\partial u_\theta}{\partial\phi}+\dfrac{\partial u_\phi}{\partial\theta}-u_\phi\cot\theta, \end{cases} \tag{2}$$

where $\nabla_2=r\partial/\partial r$, s_{ij} is the strain tensor, $u_i\,(i=r,\theta,\phi)$ are three displacement components. The equations of equilibrium in terms of stresses can easily be transformed into the following forms:

$$\begin{cases} \nabla_2\Sigma_{r\theta}+\csc\theta\dfrac{\partial\Sigma_{\theta\phi}}{\partial\phi}+\dfrac{\partial\Sigma_{\theta\theta}}{\partial\theta}+2\Sigma_{r\theta} \\ \quad+(\Sigma_{\theta\theta}-\Sigma_{\phi\phi})\cot\theta=0, \\ \nabla_2\Sigma_{r\phi}+\csc\theta\dfrac{\partial\Sigma_{\phi\phi}}{\partial\phi}+\dfrac{\partial\Sigma_{\theta\phi}}{\partial\theta}+2\Sigma_{r\phi} \\ \quad+2\Sigma_{\theta\phi}\cot\theta=0, \\ \nabla_2\Sigma_{rr}+\csc\theta\dfrac{\partial\Sigma_{r\phi}}{\partial\phi}+\dfrac{\partial\Sigma_{r\theta}}{\partial\theta}+\Sigma_{rr}-\Sigma_{\theta\theta} \\ \quad-\Sigma_{\phi\phi}+\Sigma_{r\theta}\cot\theta=0. \end{cases} \tag{3}$$

3 The State-Space-Based Formulations

It is not difficult to establish the corresponding state equation by directly choosing $(u_r,u_\theta,u_\phi,\Sigma_{rr},\Sigma_{r\theta},\Sigma_{r\phi})$ as the state variables. It has, however, been shown that, by employing certain separation formulae ([3,6,15]), not only can the basic equations be decoupled with order reduced, the subsequent solving procedure also becomes simpler. It is thus assumed that

$$u_\theta=-\frac{1}{\sin\theta}\frac{\partial\psi}{\partial\phi}-\frac{\partial G}{\partial\theta},\quad u_\phi=\frac{\partial\psi}{\partial\theta}-\frac{1}{\sin\theta}\frac{\partial G}{\partial\phi},\quad u_r=w, \tag{4}$$

and

$$\Sigma_{r\theta}=-\frac{1}{\sin\theta}\frac{\partial\Sigma_1}{\partial\phi}-\frac{\partial\Sigma_2}{\partial\theta},\quad \Sigma_{r\phi}=\frac{\partial\Sigma_1}{\partial\theta}-\frac{1}{\sin\theta}\frac{\partial\Sigma_2}{\partial\phi}, \tag{5}$$

where w, G and ψ are three displacement functions while Σ_1 and Σ_2 are two stress functions.

Utilizing Eqs. (4) and (5), through some lengthy manipulations, one can obtain the following equations from Eqs. (1)–(3):

$$\nabla_2\begin{Bmatrix}\Sigma_1\\ \psi\end{Bmatrix}=\begin{bmatrix}-2 & -c_{66}(\nabla_1^2+2)\\ \dfrac{1}{c_{44}} & 1\end{bmatrix}\begin{Bmatrix}\Sigma_1\\ \psi\end{Bmatrix}, \tag{6}$$

$$\nabla_2\begin{Bmatrix}\Sigma_{rr}\\ \Sigma_2\\ G\\ w\end{Bmatrix}=\begin{bmatrix}2\beta-1 & \nabla_1^2 & k_1\nabla_1^2 & -2k_1\\ \beta & -2 & k_2\nabla_1^2-2c_{66} & -k_1\\ 0 & \dfrac{1}{c_{44}} & 1 & 1\\ \dfrac{1}{c_{33}} & 0 & \beta\nabla_1^2 & -2\beta\end{bmatrix}\begin{Bmatrix}\Sigma_{rr}\\ \Sigma_2\\ G\\ w\end{Bmatrix}, \tag{7}$$

where $\nabla_1^2=\partial^2/\partial\theta^2+\cot\theta(\partial/\partial\theta)+\csc^2\theta(\partial^2/\partial\phi^2)$ is the two-dimensional Laplacian on a spherical surface and

$$\beta=c_{13}/c_{33},\quad k_1=2c_{13}\beta-(c_{11}+c_{12}),\quad k_2=k_1/2-c_{66}.$$

To eliminate the partial operator ∇_1^2 contained in Eqs. (6) and (7), it can be assumed for a closed spherical shell

$$\begin{cases} \Sigma_1=\displaystyle\sum_{m=0}^{n}\sum_{n=0}^{\infty}\Sigma_{1n}(r)S_n^m(\theta,\phi), \\ \psi=\displaystyle\sum_{m=0}^{n}\sum_{n=0}^{\infty}\psi_n(r)S_n^m(\theta,\phi), \\ \Sigma_{rr}=\displaystyle\sum_{m=0}^{n}\sum_{n=0}^{\infty}\Sigma_{rn}(r)S_n^m(\theta,\phi), \\ \Sigma_2=\displaystyle\sum_{m=0}^{n}\sum_{n=0}^{\infty}\Sigma_{2n}(r)S_n^m(\theta,\phi), \\ G=\displaystyle\sum_{m=0}^{n}\sum_{n=0}^{\infty}G_n(r)S_n^m(\theta,\phi), \\ w=\displaystyle\sum_{m=0}^{n}\sum_{n=0}^{\infty}w_n(r)S_n^m(\theta,\phi), \end{cases} \tag{8}$$

where $S_n^m(\theta,\phi)=P_n^m(\cos\theta)e^{im\phi}$ are spherical harmonics and $P_n^m(x)$ are the associated Legendre polynomials, and n and m are integers. From the derivations in the following, it will be shown that the integer m will not appear in the resulting ordinary differential equations about the unknown functions $\Sigma_{1n}(r)$ and $\psi_n(r)$, etc., so that we needn't indicate it in the subscript of these functions in Eq. (8) and hereafter. From Eqs. (4) and (5), it is clear that Σ_{10}, ψ_0, Σ_{20}, and G_0 all vanish in the final expressions of displacements and stresses, so that they can be assumed zero. From Eqs. (6) and (7), it is obtained that

$$r\frac{d}{dr}\begin{Bmatrix}\Sigma_{1n}\\ \psi_n\end{Bmatrix}=\begin{bmatrix}-2 & c_{66}(l-2)\\ \dfrac{1}{c_{44}} & 1\end{bmatrix}\begin{Bmatrix}\Sigma_{1n}\\ \psi_n\end{Bmatrix}, \tag{9}$$

$$r\frac{d}{dr}\begin{Bmatrix}\Sigma_{rn}\\ \Sigma_{2n}\\ G_n\\ w_n\end{Bmatrix}=\begin{bmatrix}2\beta-1 & -l & -k_1l & -2k_1\\ \beta & -2 & -k_2l-2c_{66} & -k_1\\ 0 & \dfrac{1}{c_{44}} & 1 & 1\\ \dfrac{1}{c_{33}} & 0 & -\beta l & -2\beta\end{bmatrix}\times\begin{Bmatrix}\Sigma_{rn}\\ \Sigma_{2n}\\ G_n\\ w_n\end{Bmatrix}, \tag{10}$$

where $l=n(n+1)$. It can be seen that Eqs. (9) and (10) are two separated state equations with varying coefficients.

Considering a p-ply spherical shell, Fig. 1, for the ith layer, the following variable substitution is taken:

$$r=a_ie^{\xi},\quad (i=1,2,\cdots,p;0\leqslant\xi\leqslant\xi_i), \tag{11}$$

where a_i and b_i are the inner and outer radii of the ith layer, respectively, and $\xi_i=\ln(b_i/a_i)$. Substituting Eq. (11) into Eqs. (9) and (10) gives

$$\frac{d}{d\xi}\mathbf{T}_{1ni}=\mathbf{M}_{1ni}\mathbf{T}_{1ni},\quad (n=1,2,3,\cdots), \tag{12}$$

$$\frac{d}{d\xi}\mathbf{T}_{2ni}=\mathbf{M}_{2ni}\mathbf{T}_{2ni},\quad (n=0,1,2,\cdots), \tag{13}$$

where

$$\mathbf{T}_{1ni}=[t_{1n1i},t_{1n2i}]^T=[t_{1n1},t_{1n2}]_i^T,$$

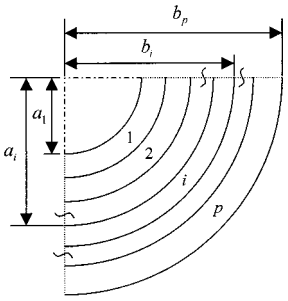


Fig. 1 The geometry of a *p*-ply spherical shell

$$\mathbf{T}_{2ni}=[t_{2n1i},t_{2n2i},t_{2n3i},t_{2n4i}]^T=[t_{2n1},t_{2n2},t_{2n3},t_{2n4}]_i^T,$$

and

$$t_{1n1}=\Sigma_{1n}/(a_1c_{44}^{(1)}),\quad t_{1n2}=\psi_n/a_1,$$

$$t_{2n1}=\Sigma_{rn}/(a_1c_{44}^{(1)}),\quad t_{2n2}=\Sigma_{2n}/(a_1c_{44}^{(1)}),$$

$$t_{2n3}=G_n/a_1,\quad t_{2n4}=w_n/a_1, \tag{14}$$

$$\mathbf{M}_{1ni}=\begin{bmatrix} -2 & \dfrac{(l-2)c_{66}}{c_{44}^{(1)}} \\ \dfrac{c_{44}^{(1)}}{c_{44}} & 1 \end{bmatrix},$$

$$\mathbf{M}_{2ni}=\begin{bmatrix} 2\beta-1 & -l & -\dfrac{k_1 l}{c_{44}^{(1)}} & -\dfrac{2k_1}{c_{44}^{(1)}} \\ \beta & -2 & \dfrac{-k_2 l-2c_{66}}{c_{44}^{(1)}} & -\dfrac{k_1}{c_{44}^{(1)}} \\ 0 & \dfrac{c_{44}^{(1)}}{c_{44}} & 1 & 1 \\ \dfrac{c_{44}^{(1)}}{c_{33}} & 0 & -\beta l & -2\beta \end{bmatrix}, \tag{15}$$

where $c_{44}^{(1)}$ represents the elastic constant in the first layer. It is noted here that a nondimensionalization procedure has been used to derive the state Eqs. (12) and (13). Thus in each lamina, we have established two separated state equations with constant coefficients in a dimensionless form. The solutions to these two equations will be presented in the next section.

4 The Solution Method

Utilizing the matrix theory, solutions to Eqs. (12) and (13) are

$$\mathbf{T}_{1ni}(\xi)=\exp(\mathbf{M}_{1ni}\xi)\mathbf{T}_{1ni}(0),\quad (n=1,2,3,\cdots;0\leqslant\xi\leqslant\xi_i), \tag{16}$$

$$\mathbf{T}_{2ni}(\xi)=\exp(\mathbf{M}_{2ni}\xi)\mathbf{T}_{2ni}(0),\quad (n=0,1,2,\cdots;0\leqslant\xi\leqslant\xi_i), \tag{17}$$

where the exponential matrices $\exp(\mathbf{M}_{1ni}\xi)$ and $\exp(\mathbf{M}_{2ni}\xi)$ are known as the transfer matrices, which can be expressed in terms of polynomials about the matrices $\mathbf{M}_{1ni}$ and $\mathbf{M}_{2ni}$, respectively, through the use of Cayley-Hamilton theorem ([16]).

Setting $\xi=\xi_i$ in Eqs. (16) and (17) gives

$$\mathbf{T}_{1ni}(\xi_i)=\exp(\mathbf{M}_{1ni}\xi_i)\mathbf{T}_{1ni}(0),\quad (n=1,2,3,\cdots;i=1,2,\cdots,p), \tag{18}$$

$$\mathbf{T}_{2ni}(\xi_i)=\exp(\mathbf{M}_{2ni}\xi_i)\mathbf{T}_{2ni}(0),\quad (n=0,1,2,\cdots;i=1,2,\cdots,p). \tag{19}$$

Thus we have established relations between the state variables of the inner and outer surfaces of the ith layer. Further allowing for the continuity conditions at each interface, we can finally get

$$\mathbf{T}_{1np}(\xi_p)=\mathbf{S}_{1n}\mathbf{T}_{1n1}(0),\quad (n=1,2,3,\cdots), \tag{20}$$

$$\mathbf{T}_{2np}(\xi_p)=\mathbf{S}_{2n}\mathbf{T}_{2n1}(0),\quad (n=0,1,2,\cdots), \tag{21}$$

where $\mathbf{S}_{1n}=\Pi_{i=p}^{1}\exp(\mathbf{M}_{1ni}\,\xi_i)$ and $\mathbf{S}_{2n}=\Pi_{i=p}^{1}\exp(\mathbf{M}_{2ni}\xi_i)$ are the second-order and fourth-order square matrices, respectively. Through these two matrices, the boundary variables at the inner and outer surfaces of a multilayered spherical shell are connected directly. For a specified boundary value problem, one does not need to solve a second-order and/or a fourth-order algebraic equation as shown by Eq. (20) and/or Eq. (21). For example, when the stresses are specified, i.e., $t_{1n11}(0)$, $t_{1n1p}(\xi_p)$ and $t_{2nj1}(0)$, $t_{2njp}(\xi_p)$ $(j=1,2)$ are known, one can get from Eqs. (20) and (21)

$$S_{1n12}t_{1n21}(0)=t_{1n1p}(\xi_p)-S_{1n11}t_{1n11}(0),\quad (n=1,2,3,\cdots), \tag{22}$$

$$\begin{bmatrix} S_{2n13} & S_{2n14} \\ S_{2n23} & S_{2n24} \end{bmatrix}\begin{Bmatrix} t_{2n31}(0) \\ t_{2n41}(0) \end{Bmatrix}=\begin{Bmatrix} t_{2n1p}(\xi_p) \\ t_{2n2p}(\xi_p) \end{Bmatrix}-\begin{bmatrix} S_{2n11} & S_{2n12} \\ S_{2n21} & S_{2n22} \end{bmatrix}\times\begin{Bmatrix} t_{2n11}(0) \\ t_{2n21}(0) \end{Bmatrix},\quad (n=0,1,2,\cdots), \tag{23}$$

where S_{1nij} and S_{2nij} are the elements on the ith row and jth column of the matrices $\mathbf{S}_{1n}$ and $\mathbf{S}_{2n}$, respectively. After the state variables of the inner surface are solved, the state variables at any interior point can be obtained by using the following formulae:

$$\mathbf{T}_{1nj}(\xi)=\exp(\mathbf{M}_{1nj}\xi)\prod_{i=j-1}^{1}\exp(\mathbf{M}_{1ni}\xi_i)\mathbf{T}_{1n1}(0),\quad (n=1,2,3,\cdots;0\leqslant\xi\leqslant\xi_j), \tag{24}$$

$$\mathbf{T}_{2nj}(\xi)=\exp(\mathbf{M}_{2nj}\xi)\prod_{i=j-1}^{1}\exp(\mathbf{M}_{2ni}\xi_i)\mathbf{T}_{2n1}(0),\quad (n=0,1,2,\cdots;0\leqslant\xi\leqslant\xi_j). \tag{25}$$

The induced variables $\Sigma_{\theta\theta}$, $\Sigma_{\phi\phi}$, and $\Sigma_{\theta\phi}$ are determined by

$$\begin{cases} \Sigma_{\theta\theta}-\Sigma_{\phi\phi}=2c_{66}\left(\nabla_1^2G-2\dfrac{\partial^2G}{\partial\theta^2}+2\cot\theta\csc\theta\dfrac{\partial\psi}{\partial\phi}-2\csc\theta\dfrac{\partial^2\psi}{\partial\theta\partial\phi}\right), \\ \Sigma_{\theta\theta}+\Sigma_{\phi\phi}=2\beta\Sigma_{rr}+k_1\nabla_1^2G-2k_1w, \\ \Sigma_{\theta\phi}=-c_{66}\left(\nabla_1^2\psi-2\dfrac{\partial^2\psi}{\partial\theta^2}-2\cot\theta\csc\theta\dfrac{\partial G}{\partial\phi}+2\csc\theta\dfrac{\partial^2G}{\partial\theta\partial\phi}\right) \end{cases}. \tag{26}$$

5 Numerical Examples

Consider a three-layered spherical shell subjected to distributed uniform pressure q over the ranges $0 \leqslant \theta \leqslant \theta_0$ and $\pi - \theta_0 \leqslant \theta \leqslant \pi$ at the outer surface $r=b$ (see Fig. 2). From Fig. 2, one has $a=a_1$, $b=b_3$, and $h=b/k=(1-\cos\theta_0)b$. Obviously when $k=1$, the whole outer surface will undergo a uniform pressure, while for $k\to\infty$, the spherical shell will be subjected to a couple of concentrated forces applied at the two poles. Because the problem considered is axisymmetric for which one has $m=0$ in Eq. (8), the distributed pressure can be expanded in the form of $\Sigma_{n=0}^{\infty}\alpha_n P_n(\cos\theta)$, where the coefficients α_n are given by

$$\alpha_n=\begin{cases}\dfrac{q}{m}, & n=0,\\[2ex] [1-(-1)^{n+1}]\left[P_{n-1}\left(\dfrac{m-1}{m}\right)-P_{n+1}\left(\dfrac{m-1}{m}\right)\right]\dfrac{q}{2}, & n>0.\end{cases} \tag{27}$$

Figures 3 and 4 display the distributions of the nondimensional stress σ_{rr}/q and the nondimensional radial displacement $\overline{u_r}=c_{44}^{(1)}w/(bq)$ for a spherical shell subjected to external uniform pressure ($k=1$). The following three cases are considered: (1) the three layers are of the same isotropic material; (2) the three layers are of the same anisotropic material; and (3) the inner and the outer layers are anisotropic and the intermediate one is isotropic. The elastic constants of the two materials are listed in Table 1, where for the isotropic material, the elastic constants, c_{ij}, are determined by

$$c_{11}=c_{33}=\frac{E(1-\nu)}{(1+\nu)(1-2\nu)},\quad c_{12}=c_{13}=\frac{E\nu}{(1+\nu)(1-2\nu)},$$

$$c_{44}=c_{66}=\frac{E}{2(1+\nu)}, \tag{28}$$

where E and ν are the Young's modulus and the Poisson ratio, respectively.

Throughout the calculation, we shall take

$$a_1=a=0.5b,\quad a_2=b_1=0.7b,\quad a_3=b_2=0.8b,\quad b_3=b.$$

The problem of a homogeneous spherical shell subjected to uniform internal and external pressures is spherically symmetric to which the solution has been given by Saint-Venant ([1,2]). Our results of cases (1) and (2) are found identical to Saint-Venant's solution.

It can be seen from Fig. 3 that, though the difference between the normal stresses of cases (1) and (2) is very small (for other materials, the difference may become obvious), in the case of a laminate, i.e., for case (3), the distribution of the normal stress changes greatly. Not only the stress gradient has a sudden change at the interface, the stress level is also raised. Such a fact is not a good thing to the engineering design. However, Fig. 4 shows that

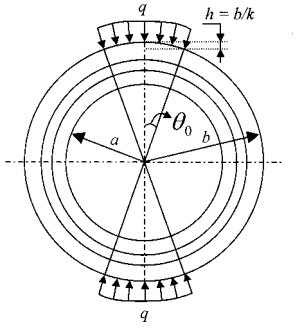


Fig. 2 A three-layered spherical shell under distributed pressures

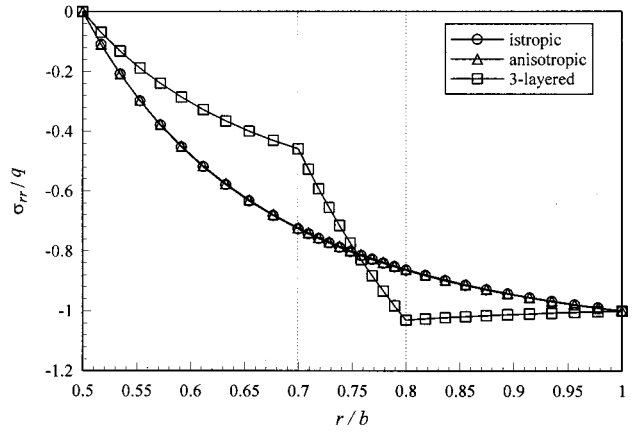


Fig. 3 Distribution of the nondimensional radial stress σ_{rr}/q in the radial direction ($k=1$)

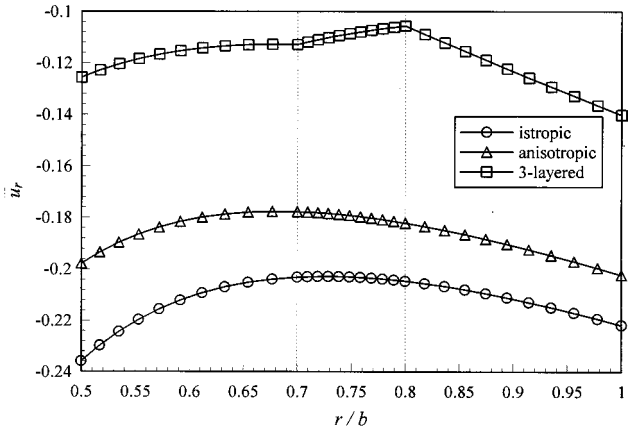


Fig. 4 Distribution of the nondimensional radial displacement $\overline{u_r}=c_{44}^{(1)}w/(bq)$ in the radial direction ($k=1$)

Table 1 Elastic constants

isotropic material	$E=20.7\times10^{10}$ Pa, $\nu=0.29$
spherically isotropic material	$c_{11}=5.97\times10^{10}$ Pa, $c_{12}=2.62\times10^{10}$ Pa, $c_{13}=2.17\times10^{10}$ Pa, $c_{33}=6.17\times10^{10}$ Pa, $c_{44}=1.64\times10^{10}$ Pa

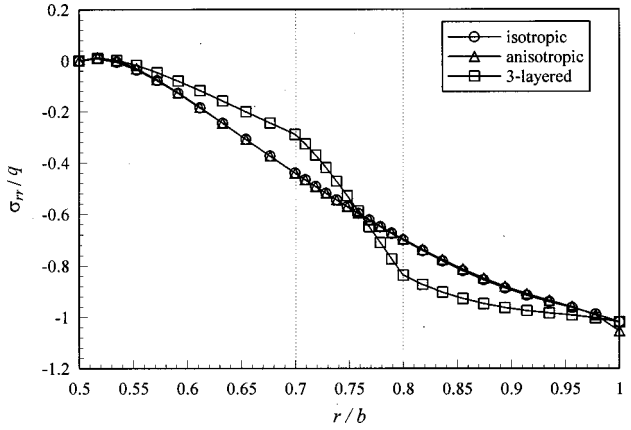


Fig. 5 Distribution of σ_{rr}/q in the radial direction ($k=4$, $\theta=\pi/6$)

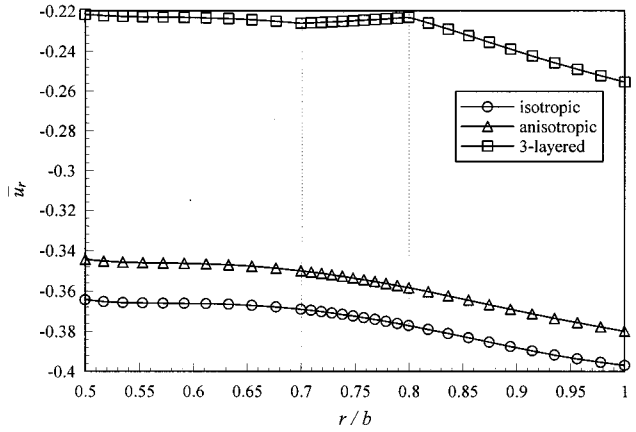


Fig. 6 Distribution of $\bar{u}_r$ in the radial direction ($k=4$, $\theta=\pi/6$)

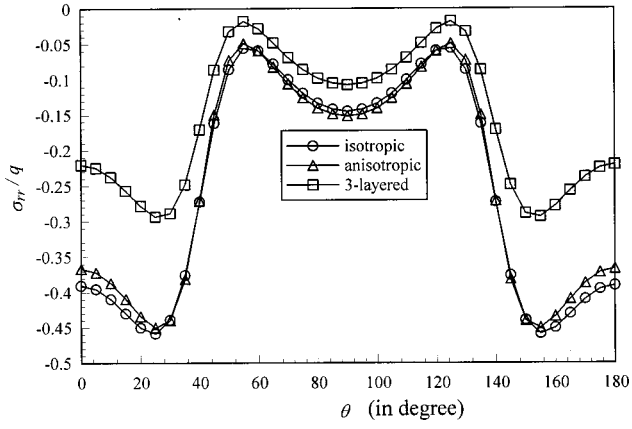


Fig. 7 Distribution of σ_{rr}/q in the circumferential direction ($k=4$, $r=0.7b$)

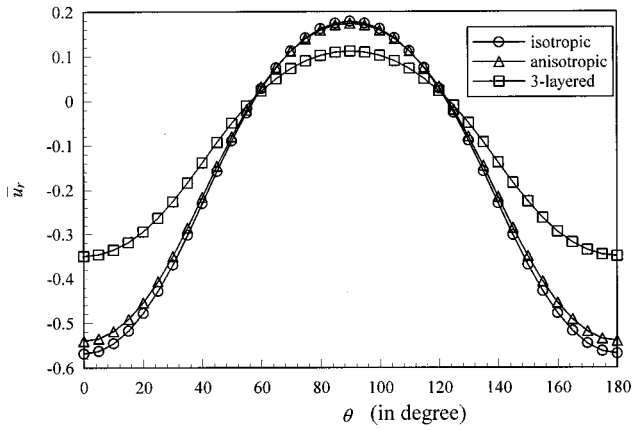


Fig. 8 Distribution of $\bar{u}_r$ in the circumferential direction ($k=4$, $r=0.7b$)

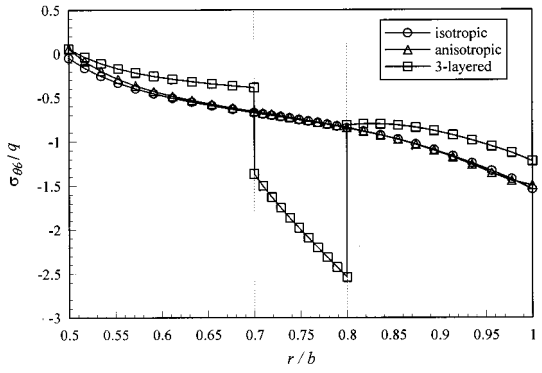


Fig. 9 Distribution of $\sigma_{\theta\theta}/q$ in the radial direction ($k=4$, $\theta=\pi/6$)

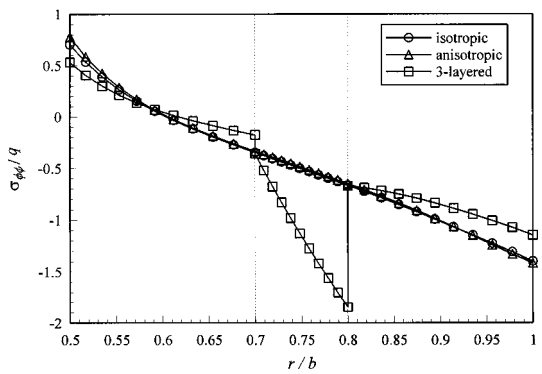


Fig. 10 Distribution of $\sigma_{\phi\phi}/q$ in the radial direction ($k=4$, $\theta=\pi/6$)

utilizing the laminate structures, one can improve the anti-deformation ability of the spherical shell effectively. Thus in practice, a proper design should be made based on a thorough evaluation of relative factors as mentioned above.

In the case of nonuniform pressure, i.e., $k\neq 1$, we take $k=4$ in the numerical calculation. Figures 5 and 6 give the radial distributions of σ_{rr}/q and $\bar{u}_r$ when $\theta=\pi/6$. Figures 7 and 8 give the circumferential distributions of σ_{rr}/q and $\bar{u}_r$ at the interface $r=a_2=0.7b$. Figures 9 and 10 display the radial distributions of $\sigma_{\theta\theta}/q$ and $\sigma_{\phi\phi}/q$ when $\theta=\pi/6$. It can be seen that for a multi-layered spherical shell, both $\sigma_{\theta\theta}$ and $\sigma_{\phi\phi}$ have a sudden jump across the material interface.

As mentioned earlier, when $k\rightarrow\infty$, one will obtain the solution to the problem of a spherical shell subjected to a couple of balanced concentrated forces applied at two poles. Table 2 shows such a procedure, where $p=P/(2\pi b^2)=2\pi bhq/(2\pi b^2)=qh/b=q/k$, and P is the radial resultant of the unilateral distributed pressure. It can be seen that the solution for a distributed pressure over a very small spherical surface ($k=1024$) agrees well with

Table 2 The variations of σ_{rr}/p and $c_{44}^{(1)}w/(bp)$ with the parameter k for $\theta=\pi/2$

radial coordinate		0.5 b	0.7 b		0.8 b	
items		$c_{44}^{(1)}w/(bp)$	σ_{rr}/p	$c_{44}^{(1)}w/(bp)$	σ_{rr}/p	$c_{44}^{(1)}w/(bp)$
k	1	-0.125645	-0.459286	-0.112921	-1.030560	-0.105698
	2	0.240259	-0.259969	0.264295	-0.301933	0.267229
	4	0.427916	-0.429636	0.443751	-0.338056	0.441715
	8	0.516154	-0.488604	0.527793	-0.339883	0.523666
	16	0.558718	-0.511249	0.568294	-0.337479	0.563223
	32	0.579603	-0.520971	0.588158	-0.335500	0.582640
	64	0.589946	-0.525446	0.597992	-0.334317	0.592258
	128	0.595092	-0.527590	0.602885	-0.333677	0.597043
	256	0.597659	-0.528638	0.605325	-0.333356	0.599431
	512	0.598941	-0.529156	0.606544	-0.333205	0.600623
	1024	0.599582	-0.529414	0.607153	-0.333082	0.601219
concentrated force		0.600222	-0.529671	0.607761	-0.332837	0.601814

the one for a concentrated force that is obtained by directly expanding the δ-function in terms of the Legendre polynomials ([17]).

6 Discussion and Conclusion

As mentioned earlier, Σ_{10}, ψ_0, Σ_{20}, and G_0 all vanish in the final expressions of displacements and stresses. In fact, when $n=0$, Eqs. (9) and (10) degenerate to the following equation:

$$r\frac{d}{dr}\begin{Bmatrix}\Sigma_{r0}\\ w_0\end{Bmatrix}=\begin{bmatrix}2\beta-1 & -2k_1\\ \dfrac{1}{c_{33}} & -2\beta\end{bmatrix}\begin{Bmatrix}\Sigma_{r0}\\ w_0\end{Bmatrix}. \tag{29}$$

It can be shown that the Saint-Venant's solution can be exactly derived from Eq. (29). Following the procedure described in Section 3 and Section 4, one can also get a series of state-space-based formulae for $n=0$. Noticing that in this case, the transfer matrix will be of second order. However, our calculation for the Saint-Venant's problem based on the fourth-order one, which does not distinguish whether $n=0$ or not, is shown valid. Thus one can uniformly adopt the formulations presented in Section 3 and Section 4 for all n.

The present state-space-based method is superior to other conventional solution methods. Numerical examples show that, especially for a spherical shell with many layers, the method can greatly reduce the computing time. Since it is completely based on three-dimensional elasticity for spherical isotropy, it can be a benchmark to check the validity of any two-dimensional approximate shell theories or numerical methods.

Acknowledgment

The work was supported by the National Natural Science Foundation of China (No. 10002016).

References

[1] Love, A. E. H., 1927, *A Treatise on the Mathematical Theory of Elasticity*, Cambridge University Press, Cambridge.
[2] Lekhnitskii, S. G., 1981, *Theory of Elasticity of an Anisotropic Body*, Mir Publishers, Moscow.
[3] Hu, H. C., 1954, ''On the General Theory of Elasticity for a Spherically Isotropic Medium,'' Acta Sci. Sin., **3**, pp. 247–260.
[4] Chen, W. T., 1966, ''On Some Problems in Spherically Isotropic Elastic Materials,'' ASME J. Appl. Mech., **33**, pp. 539–546.
[5] Puro, A. E., 1980, ''Variable Separation in Elasticity—Theory Equations for Spherically Transversely Isotropic Inhomogeneous Bodies,'' Soviet Applied Mechanics, **16**, pp. 117–120.
[6] Shul'ga, N. A., Grigorenko, A. Y., and Efimova, T. L., 1988, ''Free Nonaxisymmetric Oscillations of a Thick-Walled, Nonhomogeneous, Transversely Isotropic, Hollow Sphere,'' Soviet Applied Mechanics, **24**, pp. 439–444.
[7] Chau, K. T., 1998, ''Toroidal Vibrations of Anisotropic Spheres with Spherical Isotropy,'' ASME J. Appl. Mech., **65**, pp. 59–65.
[8] Ding, H. J., Liang, J., Zou, D. Q., and Chen, W. Q., 1997, Transversely Isotropic Elasticity, Zhejiang University Press, Hangzhou (in Chinese).
[9] Maiti, M., 1975, ''Stresses in Anisotropic Nonhomogeneous Sphere,'' ASME J. Eng. Mech., **101**, pp. 101–108.
[10] Raju, P. P., 1975, ''On Shallow Shells of Transversely Isotropic Materials,'' ASME J. Pressure Vessel Technol., **97**, pp. 185–191.
[11] Montagner, J. P., and Anderson, D. L., 1989, ''Constrained Reference Mantle Model,'' Phys. Earth Planet. Inter., **54**, pp. 205–227.
[12] Ding, Z. Y., Zou, D. Q., and Ding, H. J., 1996, ''A Study of Effects of the Earth's Radial Anisotropy on the Tidal Stress Field,'' Tectonophysics, **258**, pp. 103–114.
[13] Sosa, H. A., and Castro, M. A., 1993, ''Electroelastic Analysis of Piezoelectric Laminated Structures,'' ASME Appl. Mech. Rev., **46**, pp. 21–28.
[14] Ding, H. J., Xu, R. Q., Chen, W. Q., and Chi, Y. W., 1998, ''Free Axisymmetric Vibration of Transversely Isotropic Laminated Circular Plates,'' Acta Mechanica Solida Sinica, **11**, pp. 209–215.
[15] Ding, H. J., Ren, Y. J., Zou, D. Q., and Chen, W. Q., 1994, ''Displacement Method of Elasticity Problems in Spherically Isotropic Media,'' Acta Mech. Sin., **26**, pp. 186–197 (in Chinese).
[16] Bellman, R., 1970, *Introduction to Matrix Analysis*, McGraw-Hill, New York.
[17] Hayek, S., 1966, ''Vibration of a Spherical Shell in an Acoustic Medium,'' J. Acoust. Soc. Am., **40**, pp. 342–348.

C. O. Horgan
Applied Mechanics Program,
Department of Civil Engineering,
University of Virginia,
Charlottesville, VA 22904
e-mail: coh8p@virginia.edu
Fellow ASME

G. Saccomandi
Dipartimento di Ingegneria dell'Innovazione,
Università degli Studi di Lecce,
73100 Lecce, Italy
e-mail: giuseppe.saccomandi@unile.it

Large Deformations of a Rotating Solid Cylinder for Non-Gaussian Isotropic, Incompressible Hyperelastic Materials

The purpose of this research is to investigate the steady rotation of a solid cylinder for a class of strain-energy densities that are able to describe hardening phenomena in rubber. It is well known that use of the classic neo-Hookean strain energy gives rise to physically unrealistic *response in this problem. In particular, solutions exist only for a sufficiently small angular velocity. As the velocity approaches this limiting value, the analysis predicts that the rotating cylinder collapses to a disk. It is shown here that this nonphysical behavior does not occur when generalized neo-Hookean models, which exhibit hardening at large deformations, are used.* [DOI: 10.1115/1.1349418]

1 Introduction

Usually elastomeric materials are conveniently represented in terms of a strain-energy density function W. Thus, given an undeformed reference state, the state of strain is characterized by the principal stretches $\lambda_1,\lambda_2,\lambda_3$ of the deformation or equivalently by introducing a strain measure such as the left Cauchy-Green tensor $\mathbf{B}=\mathbf{F}\mathbf{F}^T$, where $\mathbf{F}$ is the gradient of the deformation. For an isotropic material, W is a function of the strain invariants

$$I_1=\mathrm{tr}\mathbf{B},\quad I_2=\mathrm{tr}\mathbf{B},\quad I_3=\det\mathbf{B}. \tag{1}$$

Rubber can be considered to behave in an incompressible manner as long as the hydrostatic stress does not become too large. Thus, it is common to adopt the assumption of incompressibility so that the admissible deformations must be isochoric, i.e., $\det\mathbf{F}=1$ so that $I_3=1$.

The basic strain-energy densities for rubber elasticity are the neo-Hookean strain-energy

$$W=\frac{\mu}{2}(I_1-3), \tag{2}$$

where μ is the constant shear modulus for infinitesimal deformations and the Mooney-Rivlin strain-energy

$$W=\mu_1(I_1-3)+\mu_2(I_2-3), \tag{3}$$

where μ_1 and μ_2 are constant parameters. The theoretical predictions based on these strain-energy density functions do not adequately describe experimental data especially at high values of strain. For example, the strain energies in (2) and (3) are not able to describe the characteristic *S*-shaped load versus stretch curve exhibited in simple tension experiments.

To model the typical hardening at large deformations, a number of alternative models have been proposed. In the molecular theory of elasticity (see, e.g., [1]) these models are usually called non-Gaussian, because they introduce a distribution function for the end-to-end distance of the polymeric chain which is not Gaussian. From the phenomenological point of view the non-Gaussian models can be divided into two classes: models with limiting chain extensibility and strain-hardening models.

The simplest example of the first class is the model due to Gent [2] who proposed the strain-energy density

$$W_I=-\frac{\mu}{2}J_m\ln\left(1-\frac{I_1-3}{J_m}\right), \tag{4}$$

where μ is the shear modulus and J_m is the constant limiting value for I_1-3, taking into account limiting polymeric chain extensibility. The response of this material in simple extension is described in [2]. This strain-energy density gives theoretical predictions similar to the more complicated Arruda and Boyce eight chain model ([3]). Note that from the strain-energy (4) we recover the neo-Hookean model on taking the limit as $J_m\to\infty$. For further discussion of (4) and related constitutive models, see [4–6] where solutions to the torsion, axial shear and circular shear problems have been obtained.

An example of the second class is the power-law material first proposed by Knowles in [7] in the context of anti-plane shear

$$W=\frac{\mu}{2b}\left[\left(1+\frac{b}{n}(I_1-3)\right)^n-1\right], \tag{5}$$

where μ is the shear modulus, and b and n are positive material constants. When $n=1$ in (5) we recover the neo-Hookean model. It was shown by Knowles [7] that the material modeled by (5) is *hardening* in simple shear if $n>1$. A similar model to (5) has been derived by Erman and Mark [8] in the framework of the molecular theory of elasticity using a generalized Fixman-Alben distribution for the end-to-end length of the molecular chains. The Fixman-Alben distribution function also models hardening at high strains without considering limiting chain extensibility.

The aim of this paper is to consider the deformation of a steady rotating solid cylinder of radius A for the materials (4) and (5). The rotating cylinder problem is investigated in an interesting paper by Chadwick et al. [9] (see also the book [10]). As shown in [9,10] and discussed briefly below in Section 2, the neo-Hookean model (2) gives rise to *physically unrealistic* response in this problem. In particular, solutions exist only for a sufficiently small angular velocity. As the velocity approaches this limiting value, the analysis predicts that the rotating cylinder collapses to a disk. Our purpose here is to show that this nonphysical behavior does *not* occur when the non-Gaussian models (4) and (5) are used. It should be noted that the nonphysical predictions of the neo-Hookean model may be also avoided on using the Mooney-Rivlin

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received and accepted by the ASME Applied Mechanics Division, June 8, 2000. Associate Editor: L. T. Wheeler. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

strain-energy (3). However, our goal here is to demonstrate the advantages of using generalized neo-Hookean models with $W=W(I_1)$ which exhibit material hardening at large deformations.

2 Steady Rotation of a Solid Circular Cylinder

We consider the steady rotation (at constant angular velocity ω about its central axis), of a solid circular cylinder, of radius A in the reference configuration, composed of a homogeneous incompressible isotropic hyperelatic material. The lateral surface of the cylinder is traction-free and the cylinder is assumed sufficiently long so that end effects are ignored. Thus the traction-free boundary conditions at the ends are to be satisfied globally rather than pointwise. This problem has been formulated and extensively investigated by Chadwick et al. ([9]) where references to earlier work may be found.

The kinematics of the deformation is described by

$$r=\frac{R}{\lambda^{1/2}}, \quad \theta=\Theta+\omega t, \quad z=\lambda Z, \tag{6}$$

where the material and spatial cylindrical polar coordinates are denoted by (R,Θ,Z) and (r,θ,z). Here λ and ω are positive constants and t denotes the time. In this case the principal stretches are given by

$$\lambda_1=\lambda^{-1/2}, \quad \lambda_2=\lambda^{-1/2}, \quad \lambda_3=\lambda \tag{7}$$

and, as shown in [9], the balance equations reduce to the scalar equation

$$M(\lambda)\equiv\lambda^2\hat{W}'(\lambda)=-\frac{\rho}{4}\omega^2A^2, \tag{8}$$

where $\hat{W}'(\lambda)\equiv d\hat{W}/d\lambda$ and $\hat{W}(\lambda)=W(\lambda^{-1/2},\lambda^{-1/2},\lambda)$.

For the *static* problem of a circular shaft under axial loading, as is discussed by Ogden ([10], p. 305), Haughton and Ogden ([11] p. 253), one would expect *extension* of the shaft to occur when the loading is *tensile* while *contraction* should occur when the loading is *compressive*. Thus the strain-energy density $\hat{W}(\lambda)$ is assumed to satisfy the constitutive assumptions

$$\hat{W}'(\lambda)\begin{cases}>0 & \text{if} \quad \lambda>1,\\ =0 & \text{if} \quad \lambda=1,\\ <0 & \text{if} \quad \lambda<1.\end{cases} \tag{9}$$

By virtue of (8), we see that the latter inequality holds for the rotating shaft when $\omega\neq 0$. Thus, as pointed out in ([10,11])) rotation is accompanied by *shortening* of the cylinder.

For example, for the neo-Hookean material (2) we have

$$\hat{W}(\lambda)=\frac{\mu}{2}(2\lambda^{-1}+\lambda^2-3) \tag{10}$$

so that

$$M(\lambda)=\mu(\lambda^3-1). \tag{11}$$

Thus from (8) one obtains λ in terms of ω as

$$\lambda=\left(1-\frac{\omega^2}{\omega_0^2}\right)^{1/3}, \tag{12}$$

where $\omega_0^2=4\mu/\rho A^2$. As was pointed out in [9], since $\lambda>0$, the solution (12) is meaningful only for $\omega<\omega_0$. The features of this simple solution are physically unrealistic because it implies that the cylinder collapses to a disk (i.e., $\lambda\to 0$) for a finite value of the angular velocity. Moreover it is not clear why solutions should not exist for all possible values of the velocity. Thus, the predictions of the neo-Hookean model are not physically realistic for this problem.

For the Gent material (4), one finds that

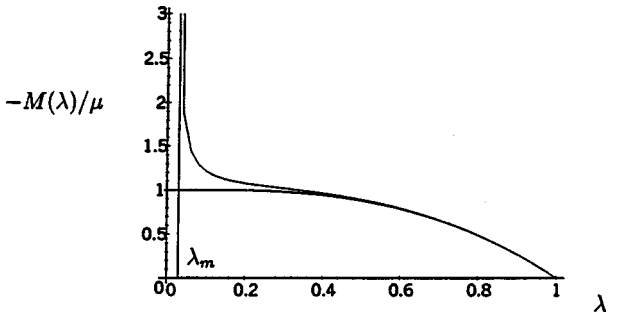


Fig. 1 Plot of $-M(\lambda)/\mu$ versus λ for the neo-Hookean and Gent material. The curve for the Gent material has the vertical line as an asymptote as $\lambda\to\lambda_m$.

$$\hat{W}(\lambda)=-\frac{\mu}{2}J_m\ln\left(1-\frac{2\lambda^{-1}+\lambda^2-3}{J_m}\right). \tag{13}$$

The second and third of the inequalities in (9) now show that λ is restricted to the range

$$\lambda_m<\lambda\leqslant 1, \tag{14}$$

where λ_m is the smallest positive root of the cubic

$$\lambda^3-(3+J_m)\lambda+2=0. \tag{15}$$

This root is given by

$$\lambda_m=2\sqrt{\frac{J_m+3}{3}}\sin\left[\frac{1}{3}\arcsin\left(3\sqrt{\frac{3}{(J_m+3)^3}}\right)\right]. \tag{16}$$

As $J_m\to 0$ we have $\lambda_m\to 1$, whereas for $J_m>0$ we have $\lambda_m<1$.

Equation (8) now reads

$$\frac{M(\lambda)}{\mu}\equiv\frac{\lambda^3-1}{1-\dfrac{2\lambda^{-1}+\lambda^2-3}{J_m}}=-\frac{\rho}{4\mu}\omega^2A^2. \tag{17}$$

In Fig. 1, we have plotted $-M(\lambda)/\mu$ versus λ for the neo-Hookean material and for the Gent material. The value $J_m=97.2$ is chosen since this is the value obtained by Gent from fitting with uniaxial data ([2]). Thus $\lambda_m=0.01996$.

The contrasting behavior predicted by the two material models is evident from Fig. 1. For the Gent material, a unique solution exists for all ω^2. Furthermore, since $\lambda>\lambda_m$, the nonphysical prediction of collapse to a disk is now eliminated.

For the power-law material (5) we have

$$\hat{W}(\lambda)=\frac{\mu}{2b}\left[\left(1+\frac{b}{n}(2\lambda^{-1}+\lambda^2-3)\right)^n-1\right]. \tag{18}$$

In contrast with the situation for the Gent material, the constitutive inequalities (9) do not impose any minimum allowable value of λ in this case.

Equation (8) reads

$$\frac{M(\lambda)}{\mu}\equiv\left(1+\frac{b}{n}\left(\frac{2}{\lambda}+\lambda^2-3\right)\right)^{n-1}(\lambda^3-1)=-\frac{\rho}{4\mu}\omega^2A^2. \tag{19}$$

There are now several possibilities. Consider, for example, the cases $n=3/2$ and $n=1/4$, respectively, and for simplicity we take $b=1$. The results are plotted in Fig. 2. For $n=3/2$ (solid curve), the behavior is similar to that predicted by the Gent material in Fig. 1. For $n=1/4$ (dashed curve), the behavior is quite different.

The response depicted in Fig. 2 is typical of that exhibited for values of $n>1$ and $n<1$, respectively. We see from (19) that

$$-\frac{M(1)}{\mu}=0 \tag{20}$$

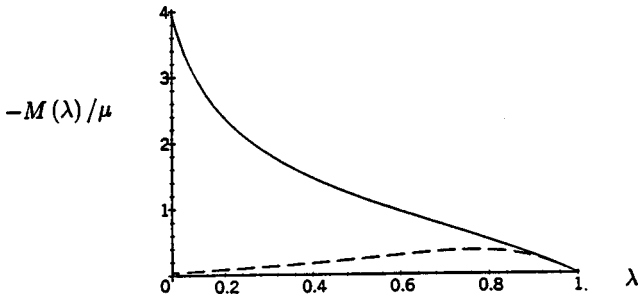


Fig. 2 Plot of $-M(\lambda)/\mu$ versus λ for the power-law material $n=3/2$ (solid curve) and $n=1/4$ (dashed curve). The solid curve has the vertical axis as an asymptote as $\lambda \to 0^+$.

and

$$\lim_{\lambda \to 0^+} \left(-\frac{M(\lambda)}{\mu} \right) = \begin{cases} +\infty & \text{if} \quad n>1, \\ 1 & \text{if} \quad n=1, \\ 0 & \text{if} \quad n<1. \end{cases} \tag{21}$$

Consider first the case of a *hardening* material for which $n>1$. It can be shown from (19) that $-M(\lambda)/\mu$ is monotone increasing in λ for $0<\lambda<1$, and so, corresponding to a given ω, there exists a *unique* λ. As $\omega \to \infty$ in (19), we see from (21) that $\lambda \to 0^+$ so that infinite velocity is required to shorten the cylinder to a disk. The results for $n>1$ are similar to those obtained in [9] for the Mooney-Rivlin material (3). For a *softening* material ($n<1$) the $M(\lambda)/\mu$ versus λ curve is no longer monotone and existence is ensured only for values of the velocity less than or equal to a $\omega_{\max}$. In fact given a $\omega<\omega_{\max}$, there are two values for the corresponding stretch λ. Furthermore, as $\omega \to 0$, it follows from (19)–(21) that $\lambda \to 1^-$ or $\lambda \to 0^+$, the latter case corresponding to collapse to a disk. The neo-Hookean material (10), corresponding to $n=1$ in (18), is now seen to be an intermediate special case.

3 The Stress Response

It is shown in [9] that, for an arbitrary incompressible hyperelastic material, the stresses are given by

$$T_{rr}=T_{\theta\theta}=\frac{\rho\omega^2}{2\lambda}(A^2-R^2),$$

$$T_{zz}=T_{rr}+\lambda \hat{W}'(\lambda) \tag{22}$$

where, for a prescribed ω, the stretch λ is obtained from (8). On using $(22)_1$ and (8) in $(22)_2$, we find that

$$T_{zz}=\frac{\rho\omega^2}{2\lambda}\left(\frac{A^2}{2}-R^2\right). \tag{23}$$

Observe that the axial stress is tensile on the inner core where $0 \leq R<A/\sqrt{2}$ and compressive on the outer annulus where $A/\sqrt{2}<R\leq A$, with zero *resultant* axial force. The radial dependence of the stress response is explicitly given in (22), (23). The parameter λ in the multiplicative factor $(\rho\omega^2)/2\lambda$ in (22), (23) can be eliminated on using the ω/λ relationship (8) discussed in Section 2. Observe from (22), (23) that the universal relation

$$\frac{T_{rr}}{T_{zz}}=2\left(\frac{A^2-R^2}{A^2-2R^2}\right) \tag{24}$$

holds. This result is valid for all incompressible isotropic hyperelastic materials.

The stresses induced in rotating cylinders within the theory of linearized elasticity are well known (see, e.g., [12], pp. 384–386). For an *incompressible* linearly elastic solid, these stresses can be obtained on setting Poisson's ratio $\nu=1/2$. Alternatively, they can be found from (22), (23) on setting $\lambda=1$.

4 Concluding Remarks

The preceding results have been obtained under the assumptions (6) that the radial deformation depends only on the radial coordinate. The possibility of bifurcation from such a configuration has been investigated in [11] and illustrated for the Ogden strain-energy density. We shall not pursue such considerations here. One of the advantages of the rotating cylinder problem is that it may provide a means of obtaining experimental data in compression while avoiding common instabilities such as buckling. Such data is usually difficult to obtain (see, e.g., [13] for recent results). Thus, as we have suggested in the context of torsion ([4]), axial shear ([5]), and azimuthal shear ([6]), it is hoped that the present results may help to provide guidelines for future experimental work on large deformations of rubber-like materials.

Acknowledgments

The research of C.O.H. was supported by the U.S. Air Force Office of Scientific Research under Grant AFOSR-F49620-98-1-0443. The work of G. S. was partially supported by GNFM of Italian INDAM, Progetto Giovani Ricercatori MURST of 1999 through Università degli Studi di Perugia and by ex-60 percent research grant of the University of Lecce.

References

[1] Mark, J. E., and Erman, B., 1988, *Rubberlike Elasticity: A Molecular Primer*, John Wiley and Sons, New York.
[2] Gent, A. N., 1996, ''A New Constitutive Relation for Rubber,'' Rubber Chem. Technol., **69**, pp. 59–61.
[3] Boyce, M. C., 1996, ''Direct Comparison of the Gent and the Arruda-Boyce Constitutive Models of Rubber Elasticity,'' Rubber Chem. Technol., **69**, pp. 781–785.
[4] Horgan, C. O., and Saccomandi, G., 1999, ''Simple Torsion of Isotropic, Hyperelastic, Incompressible Materials With Limiting Chain Extensibility,'' J. Elast., **56**, pp. 159–170.
[5] Horgan, C. O., and Saccomandi, G., 1999, ''Pure Axial Shear of Isotropic Incompressible Nonlinearly Elastic Materials With Limiting Chain Extensibility,'' J. Elast., **57**, pp. 307–319.
[6] Horgan, C. O., and Saccomandi, G., 2001, ''Pure Azimuthal Shear of Isotropic Incompressible Hyperelastic Materials With Limiting Chain Extensibility,'' Int. J. Non-Linear Mech., **36**, pp. 465–475.
[7] Knowles, J. K., 1977, ''The Finite Anti-Plane Shear Field Near the Tip of a Crack for a Class of Incompressible Elastic Solids,'' Int. J. Fract., **13**, pp. 611–639.
[8] Erman, B., and Mark, J. E., 1988, ''Use of Fixman-Alben Distribution Function in the Analysis of Non-Gaussian Rubber-Like Elasticity,'' J. Chem. Phys., **89**, pp. 3314–3316.
[9] Chadwick, P., Creasy, C. F. M., and Hart, V. G., 1977, ''The Deformation of Rubber Cylinders and Tubes by Rotation,'' J. Aust. Math. Soc. B, Appl. Math., **20**, Series 13, pp. 62–96.
[10] Ogden, R. W., 1984, *Non-linear Elastic Deformations*, Ellis Horwood, Chichester, UK, reprinted by Dover, New York, 1997.
[11] Haughton, D. M., and Ogden, R. W., 1980, ''Bifurcation of Finitely Deformed Rotating Cylinders,'' Q. J. Mech. Appl. Math., **33**, pp. 251–265.
[12] Hunter, S. C., 1976, *Mechanics of Continuous Media*, Ellis Horwood, Chichester, UK.
[13] Mott, P. H., and Roland, C. M., 1996, ''Elasticity of Natural Rubber Networks,'' Macromolecules, **29**, pp. 6941–6945.

P. Song
Graduate Student,
e-mail: pengs@grip.cis.upenn.edu

P. Kraus
Graduate Student,
e-mail: pkraus@grip.cis.upenn.edu

V. Kumar
e-mail: kumar@grip.cis.upenn.edu
Mem. ASME

GRASP Laboratory,
University of Pennsylvania,
3401 Walnut Street,
Philadelphia, PA 19104

P. Dupont
Department of Aerospace and
Mechanical Engineering,
Boston University,
Boston, MA 02215
e-mail: pierre@bu.edu
Mem. ASME

Analysis of Rigid-Body Dynamic Models for Simulation of Systems With Frictional Contacts

The use of Coulomb's friction law with the principles of classical rigid-body dynamics introduces mathematical inconsistencies. Specifically, the forward dynamics problem can have no solutions or multiple solutions. In these situations, compliant contact models, while increasing the dimensionality of the state vector, can resolve these problems. The simplicity and efficiency of rigid-body models, however, provide strong motivation for their use during those portions of a simulation when the rigid-body solution is unique and stable. In this paper, we use singular perturbation analysis in conjunction with linear complementarity theory to establish conditions under which the solution predicted by the rigid-body dynamic model is stable. We employ a general model of contact compliance to derive stability criteria for planar mechanical systems. In particular, we show that for cases with one sliding contact, there is always at most one stable solution. Our approach is not directly applicable to transitions between rolling and sliding where the Coulomb friction law is discontinuous. To overcome this difficulty, we introduce a smooth nonlinear friction law, which approximates Coulomb friction. Such a friction model can also increase the efficiency of both rigid-body and compliant contact simulation. Numerical simulations for the different models and comparison with experimental results are also presented. [DOI: 10.1115/1.1331060]

1 Introduction

There are many applications in an industrial setting where it is beneficial to understand the dynamics of systems with frictional contacts. Examples include part-feeding systems ([1]) and automatic assembly of mechanical components ([2]). Examples of mechanical systems with frictional contacts include multifingered grippers ([3]), multiarm manipulation systems ([4]), legged locomotion systems, and wheeled robots on uneven terrain ([5]). In order to successfully design and optimize such mechanical systems or manufacturing processes, a method for modeling and simulating mechanical systems with frictional contacts is necessary ([6]).

In a forward dynamics problem, it is well known that in the frictionless case there is always a unique solution for the accelerations. When the constraints are not all independent, the system is statically indeterminate and the constraint forces cannot be uniquely determined. In the frictional case, if all contacts are known to be rolling (sticking), the existence of a solution can be shown if the constraints are independent ([7]). In all other cases, the initial value problem can be shown to have no solution or multiple solutions for special choices of initial conditions ([8,9]). The major difficulty of proving existence and uniqueness arises when rigid-body models are combined with friction laws coupling normal and tangential contact forces. In these situations, it is attractive to pursue models in which the contact forces are explicit functions of the state variables. For example, a continuum model for modeling the deformations at each contact is described in ([10]). Each contact is modeled as frictional elastic or viscoelastic, and the contact force distribution across the contact patch is calculated using a finite element mesh. This general approach is further refined by [11]. Existence and uniqueness is shown for the special case in which the maximum tangential force at each point is a priori known.

The empirical nature of friction models can cause additional difficulties with dynamic simulation. The most widely employed model, for example, is Coulomb friction. When used in combination with a rigid-body contact model, the tangential force is a discontinuous function of the sliding velocity and independent of tangential displacement. Furthermore, this model does not predict such phenomena as microslip, hysteresis, and local adhesion ([12]). Both these difficulties can be overcome by combining the Coulomb friction model with a simple lumped model of compliance (e.g., the Kelvin-Voigt model [13]). At very small displacements, the tangential force opposes the tangential displacement, simulating an approximately linear spring. For small oscillatory displacements, hysteric behavior is exhibited as in [14]. With a suitable modification to the Coulomb friction model, the steady-state friction force can be made to decrease with increasing velocity thus simulating the development of a lubricant film ([12]). However, while the difficulty with discontinuities is eliminated, such Coulomb-like friction laws are generally not smooth. The laws are described by separate equations for rolling and sliding contact and are not differentiable at transitions between rolling and sliding. We will overcome this difficulty by introducing a friction model that depends on normal force, but which is continuously differentiable.

In this paper, we derive a simple *compliant contact model* that (a) provides a framework for analyzing frictional forces for constraint dynamic systems; and (b) establishes a unique solution for initial value problems in dynamic simulation. We use methods from singular perturbation analysis to establish conditions under which the solution predicted by the rigid-body model is *stable*. We argue that rigid-body dynamic simulation is meaningful only when the solution of the compliant contact model converges to the solution of the rigid-body model. Experimental results and numerical simulations are illustrated to verify the stability analysis. We also describe stability results using a *smooth nonlinear friction law* which is an alternative to the Coulomb's friction model.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, June 23, 1999; final revision, June 16, 2000. Associate Editor: A. A. Ferri. Discussion on the paper should be addressed to the Editor, Professor Lewis T. Wheeler, Department of Mechanical Engineering, University of Houston, Houston, TX 77204-4792, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Rigid-Body Models

The dynamic equations of motion for a mechanical system comprised of rigid bodies subject to Coulomb friction can be written in the form

$$M(q)\ddot{q}+h(q,\dot{q})=u+\Phi_q^T\lambda \tag{1}$$

where $q\in\Re^n$ is the vector of generalized coordinates, $M(q)$ is an $n\times n$ positive-definite symmetric inertia matrix, $h(q,\dot{q})$ is a $n\times 1$ vector of nonlinear inertial forces, u is the vector of applied (external) forces and torques, and λ is the vector of constraint forces. The system is subject to k unilateral constraints:

$$\Phi(q)=(\phi_1(q),\cdots,\phi_k(q))^T\geqslant 0 \tag{2}$$

and Φ_q in Eq. (1) is the $k\times n$ Jacobian matrix, $\partial\Phi/\partial q$. We will assume, without loss of generality, that this does not include bilateral, holonomic constraints. Further, for the sake of simplicity, we will assume that nonholonomic constraints are not present.

Suppose there are c contacts, consisting of r rolling contacts and s sliding contacts. Let the subscripts N and T denote quantities in the normal and tangential contact directions and the subscripts S and R denote sliding and rolling contacts, respectively. The Jacobian matrix and constraint forces in Eq. (1) are given by

$$\Phi_q^T=(\hat{\Phi}_{S_q}^T\Phi_{NR_q}^T\Phi_{TR_q}^T),\quad \hat{\Phi}_{S_q}^T=(\Phi_{NS_q}^T+\Phi_{TS_q}^T\mu_s), \tag{3}$$

$$\lambda=(\lambda_{NS}^T\lambda_{NR}^T\lambda_{TR}^T)^T, \tag{4}$$

where $\mu_s=-\mathrm{diag}(\mu\,\mathrm{sign}(\hat{\Phi}_{TS}))$, μ is a $s\times s$ diagonal matrix that contains all the coefficients of friction at the sliding contacts, $\hat{\Phi}_{S_q}$ is a $s\times n$ matrix, Φ_{NR_q} and Φ_{TR_q} are both $r\times n$ matrices, and the total number of constraints $k=2r+s$. λ_{NS} is the s-dimensional vector of normal forces at sliding contacts, while λ_{NR} and λ_{TR} are the $r\times 1$ vectors of normal and tangential forces at rolling contacts, respectively.

Contacts between rigid bodies generate complementary constraints on the position (or velocity or acceleration) variables and the corresponding force variables. In the normal direction, if no new contact becomes active over a finite time interval, then in that interval, there is a complementary equation satisfied by the relative normal acceleration, $\ddot{\phi}_{N,i}$, and the normal force, $\lambda_{N,i}$ ([8])

$$\ddot{\phi}_{N,i}\geqslant 0,\quad \lambda_{N,i}\geqslant 0,\quad \ddot{\phi}_{N,i}\lambda_{N,i}=0,\quad i=1,\ldots,c. \tag{5}$$

This complementary constraint is valid for all sliding contacts (indexed by the subscript $i=1,\ldots,s$) and rolling contacts (indexed by $i=s+1,\ldots,c$). Subscripts R and S are omitted for convenience. This condition allows active contacts to become inactive. The case of inactive contacts becoming active is modeled by rigid-body impacts and is treated elsewhere ([13]). Similar complementarity constraints can be found in the tangential direction by assuming Coulomb's friction. We refer the reader to ([15]) for details.

The problem of determining contact forces can be reduced to a linear complementarity problem (LCP) that has the form ([7])

$$x\geqslant 0,\quad y=Ax+B>0,\quad y^Tx=0. \tag{6}$$

The LCP has a unique solution for all vectors B if an only if the matrix A is a P matrix ([16]). However, even if A is not a P matrix, the LCP may have unique solution for special choices of B. For other choices of B, Eq. (6) may have no solution or multiple solutions. To overcome these inconsistencies, we consider more sophisticated models of contact interactions in the next section.

3 Compliant Contact Models

Our contact model of compliance assumes that the principles of rigid-body dynamics are valid and the gross motion of the dynamic system is described by the state variables $(q,\dot{q})$. However, in addition to the gross motion, there are small (local) deformations at each contact. Thus a rigid body can be modeled as a rigid core surrounded by a very thin deformable layer the inertia of which is considered to be negligible, as shown in the schematic in Fig. 1. The gross rigid-body motion determines the relative displacement at the contact point (ϕ_T,ϕ_N). The actual relative displacement of the contact point is given by $(\phi_T+\delta_T,\phi_N+\delta_N)$. The contact forces are related to the normal and tangential deformations (δ_N,δ_T) of the deformable layer and their derivatives $(\dot{\delta}_N,\dot{\delta}_T)$ through the material properties of the deformable layer.

A general viscoelastic model for contact compliance is shown in Fig. 1. At contact i, the normal and tangential contact forces ($\lambda_{N,i}$ and $\lambda_{T,i}$) between the two contacting bodies may be modeled as

$$\lambda_{N,i}=f_{N,i}(\delta_{N,i})+g_{N,i}(\delta_{N,i},\dot{\delta}_{N,i}),\quad i=1,\ldots,c, \tag{7}$$

$$\lambda_{T,i}=f_{T,i}(\delta_{T,i})+g_{T,i}(\delta_{T,i},\dot{\delta}_{T,i}),\quad i=1,\ldots,c, \tag{8}$$

where the functions $f_{N,i}$ and $f_{T,i}$ are the elastic stiffness terms and $g_{N,i}$ and $g_{T,i}$ are the damping terms in the normal and tangential directions, respectively. These functions depend on the geometry and material properties of the two bodies in contact and may be nonlinear. We have decoupled the modeling of the contact forces (i.e., the force at a contact is only dependent on the deformation at that contact). We will consider the case where the tangential force obeys Coulomb's frictional law:

$$|\lambda_{T,i}|\leqslant\mu_i\lambda_{N,i}. \tag{9}$$

An alternative frictional model is discussed in Section 7.

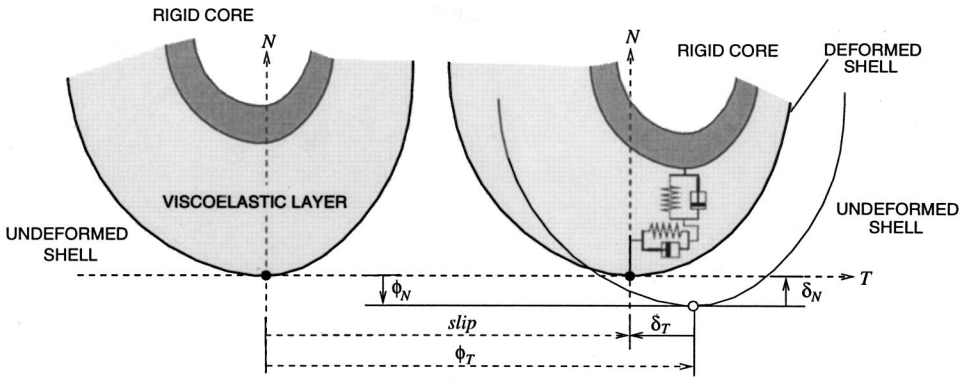


Fig. 1 A simple model of contact compliance

The simplest viscoelastic model is the Kelvin-Voigt model given by

$$f_i = K_i \delta_i, \quad g_i = C_i \dot{\delta}_i, \quad i = 1, \ldots, c, \tag{10}$$

where K_i and C_i are stiffness and damping coefficients (in the normal or tangential directions) respectively. The coefficients can be estimated using linear elastic and viscoelastic theory for half-spaces ([17]). A more sophisticated model due to Hunt and Crossley ([18]) incorporates nonlinear elastic and dissipation terms:

$$f_i = K_i \delta_i^{\beta}, \quad g_i = \frac{3}{2} \alpha K_i \delta_i^{\beta} \dot{\delta}_i, \quad i = 1, \ldots, c, \tag{11}$$

where α and β are functions of the material properties and the local geometry.

In any of the above models, the normal deformations are directly related to the constraints in the normal direction. The normal deformations and constraint forces are given by

$$\delta_{N,i} = \max\{0, -\phi_{N,i}(q)\}, \tag{12}$$

if $\delta_{N,i} > 0$,

$$\begin{cases} \dot{\delta}_{N,i} = -\dot{\phi}_{N,i}(q), \quad i = 1, \ldots, c \\ \lambda_{N,i} = \max\{0, f_{N,i}(\delta_{N,i}) + g_{N,i}(\delta_{N,i} + \dot{\delta}_{N,i})\}. \end{cases} \tag{13}$$

In the tangential direction we define a new variable, σ_i, to denote the relative sliding velocity between the (deformed) contact points at contact *i*. This quantity is the *slip rate*, the sum of the tangential rigid-body velocity at the contact and the rate of tangential deformation:

$$\sigma_i = \dot{\phi}_{T,i} + \dot{\delta}_{T,i}.$$

For rolling contacts, we have

$$\lambda_{T,i} = f_{T,i}(\delta_{T,i}) + g_{T,i}(\delta_{T,i}, \dot{\delta}_{T,i}) \tag{14}$$

$$\dot{\delta}_{T,i} = -\dot{\phi}_{T,i}(q), \quad i = s+1, \ldots, c, \tag{15}$$

in conjunction with the frictional inequality of Eq. (9). For sliding contacts,

$$\dot{\delta}_{T,i} = h_{T,i}(\lambda_{T,i} - f_{T,i}(\delta_{T,i})), \tag{16}$$

$$\lambda_{T,i} = -\mu \lambda_{N,i} \operatorname{sign}(\sigma_i), \quad i = 1, \ldots, s, \tag{17}$$

where $h_{T,i}(.)$ is the inverse of the function $g_{T,i}$ in Eq. (8) for a given $\delta_{T,i}$. For both sliding and rolling contacts, we track the tangential deformations by integrating the expression for its derivative:

$$\delta_{T,i} = \int_{t_0}^{t} \dot{\delta}_{T,i} dt + \delta_{T,i}(t_0). \tag{18}$$

In order to determine which set of equations apply, we start with the assumption that any contact is rolling. If the tangential force from Eq. (15) violates the frictional constraint in (9), the contact is sliding and Eqs. (16)–(17) yield the correct force with $\operatorname{sign}(\sigma_i)$ taken to be the opposite of the sign of the tangential force in Eq. (15). It is clear that Eqs. (12)–(18) always provide a unique answer for the normal and tangential contact forces and the positive-definiteness of M in Eq. (1) yields a unique solution for $\ddot{q}$.

There are two disadvantages of the compliant contact model. First it is clear that we now need to model the contacts and this increases the possibility of modeling errors. Second, and more importantly from an computational standpoint, there is a need to extend the dimension of the state space from $2n-2(c+r)$ to $2n+c$ in order to track the tangential deformation, $\delta_{T,i}$, at each contact. The three main advantages, which outweigh the disadvantages, are: (a) The normal and tangential forces are now uniquely determined and there is no question of static indeterminacy; (b) The difficulties with uniqueness and existence no longer arise; and (c) A model with tangential contact compliance is more realistic and can better explain physical observations ([13]).

We do not wish to promote unnecessary model complexity, however, and in those situations when a compliant contact model is not needed, it would be desirable to retain the simpler rigid-body model. The popularity of rigid-body models can be attributed not only to their simplicity, but also to the fact that they produce adequate results in a broad range of applications. Clearly, rigid-body models can only be used when a unique solution can be determined without any additional ad hoc assumptions. But even when this is the case, it is meaningful to use the reduced-order rigid-body model only when the solution from the more accurate compliant contact model converges to the solution obtained from the rigid-body model. In the next section, we will use singular perturbation theory to investigate the *stability* of the solutions obtained from the rigid-body model.

4 Singular Perturbation Analysis

The rigid-body model leads to a set of differential-algebraic equations as shown in Section 2. In the compliant contact model, the deformations at the contact points are at least an order of magnitude smaller than the gross motions of the mechanical system. By setting these small deformations to zero (or by allowing the corresponding stiffnesses to be infinitely large), we recover the equations of the rigid-body model. This suggests that we can use singular perturbation theory to decompose the system model into reduced (slow time scale) and boundary layer (fast time scale) models ([19]). In mechanical systems described by Eq. (1), the slow time scale corresponds to the reduced-order rigid-body model dynamics and the fast time scale is the time scale that characterizes the contact dynamics ([20,21]). The response of the system then consists of a slow response and a fast transient. If the boundary layer model is exponentially stable, the fast transient will exponentially converge to zero and it is reasonable to neglect the high-frequency contact dynamics. In such a situation, the reduced-order model obtained by neglecting the compliance is robust to the unmodeled dynamics. If the boundary layer model is not stable, we cannot neglect these terms and it is necessary to use the complete dynamic model given by Eqs. (12)–(18).

We first partition the generalized coordinates q into the fast variables q_1, related to the contact deformations, and the remaining slow variables, q_2. We accordingly define a new set of variables:

$$p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} \Phi_N(q_1, q_2) \\ \Phi_{TR}(q_1, q_2) \\ q_2 \end{pmatrix} \in \Re^n,$$

where $p_1, q_1 \in \Re^k$ and $p_2, q_2 \in \Re^{n-k}$. Recall that k is the total number of constraints. In order to make p a valid choice of coordinates, the implicit function theorem requires that the Jacobian matrix

$$\Gamma = \begin{pmatrix} \Phi_{Nq(c\times n)} \\ \Phi_{TRq(r\times n)} \\ 0_{(n-k)\times k} \quad I_{(n-k)\times(n-k)} \end{pmatrix} \in \Re^{n\times n}$$

be nonsingular, that is, the contact normals and the rolling tangents have to be linearly independent. If these conditions are satisfied, we may write

$$\begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \end{pmatrix} = J(p_1, p_2) \begin{pmatrix} \dot{p}_1 \\ \dot{p}_2 \end{pmatrix}$$

where $J = \Gamma^{-1}$. Note that the choice of the p-coordinates is arbitrary as long as Γ^{-1} exists. The time variable and the p-coordinates can be nondimensionalized by letting

$$\overline{t} = \frac{t}{T}, \quad \overline{p_1} = D_1^{-1} p_1, \quad \overline{p_2} = D_2^{-1} p_2, \tag{19}$$

where T is the characteristic time scale and $\bar{t}$ is dimensionless. $\bar{p}_1$ and $\bar{p}_2$ are the nondimensionalized fast and slow variables, respectively. D_1 is a diagonal matrix whose components are the deformation length scales while D_2 is a diagonal matrix of the characteristic scales of the slow variables. For the sake of simplicity, all contacts are assumed to have similar physical properties, and the diagonal matrix of the deformation scales can be defined as $D_1 = d_1 \cdot I_{k\times k}$. We also define a parameter ϵ as the dimensionless ratio, d_1/L, where L is the length scale for gross rigid-body motions. As d_1 tends to zero, ϵ goes to zero, and the compliant contact model degenerates into the rigid-body model.

We use ϵ and the dimensionless variables in (19) to perform a second change of coordinates. Let

$$\mathbf{x} = (x_1, x_2)^T = (\bar{p}_2, \bar{p}_2')^T, \quad \mathbf{y} = (y_1, y_2)^T = (\bar{p}_1, \sqrt{\epsilon}\bar{p}_1')^T \tag{20}$$

be the new state variables and rewrite the dynamic Eqs. (1) in state space notation:

$$\begin{pmatrix} \sqrt{\epsilon}y_1' \\ x_1' \\ \sqrt{\epsilon}y_2' \\ x_2' \end{pmatrix} = \begin{pmatrix} y_2 \\ x_2 \\ (T^2 D^{-1} J^{-1} M^{-1} \Phi_q^T)\lambda + \left(T^2 D^{-1} J^{-1} M^{-1}(u-h) - TJ^{-1}\dot{J}\begin{pmatrix} \sqrt{\epsilon}y_2 \\ x_2 \end{pmatrix} \right) \end{pmatrix}, \tag{21}$$

where $'$ denotes the differentiation with respect to $\bar{t}$, $D = \begin{pmatrix} L\cdot I_{k\times k} & 0 \\ 0 & D_2 \end{pmatrix}$, and $\lambda_i = f_i(-\epsilon L y_{1,i}) + g_i(-\epsilon L y_{1,i}, -\sqrt{\epsilon}y_{2,i}L/T)$. Here we use the notation $y_{i,j}$ to refer the jth component of the vector y_i. $\mathbf{x}(\bar{t})$ and $\mathbf{y}(\bar{t})$ represent the dimensionless slow and fast trajectories, respectively.

The differential equations for the fast variables are given by

$$\sqrt{\epsilon}\mathbf{y}' = \begin{pmatrix} y_2 \\ A(\mathbf{x},\mathbf{y},\epsilon)\bar{\lambda}(\mathbf{y},\epsilon) + B(\mathbf{x},\mathbf{y},\epsilon) \end{pmatrix} \tag{22}$$

with

$$\bar{\lambda}(\mathbf{y},\epsilon) = \begin{pmatrix} \bar{\lambda}_N(\mathbf{y},\epsilon) \\ \bar{\lambda}_{TR}(\mathbf{y},\epsilon) \end{pmatrix} = \frac{T^2}{D_M L}\begin{pmatrix} \lambda_N \\ \lambda_{TR} \end{pmatrix}, \quad \bar{\lambda}_N(\mathbf{y},\epsilon) \geqslant 0,$$

$$A(\mathbf{x},\mathbf{y},\epsilon) = D_M L (D^{-1}J^{-1}M^{-1}\Phi_q^T)_{kk},$$

$$B(\mathbf{x},\mathbf{y},\epsilon) = (T^2 D^{-1}J^{-1}M^{-1})_{kn}(u-h) - (TJ^{-1}\dot{J})_{kn}\begin{pmatrix} \sqrt{\epsilon}y_2 \\ x_2 \end{pmatrix},$$

where D_M is the characteristic mass. Here, $(\cdot)_{\alpha\beta}$ refers to the submatrix in (.) consisting of the first α rows and the first β columns. As ϵ goes to zero, Eq. (22) degenerates into the following algebraic equations:

$$\begin{pmatrix} y_2 \\ A(\mathbf{x},\mathbf{y},0)\bar{\lambda}(\mathbf{y},0) + B(\mathbf{x},\mathbf{y},0) \end{pmatrix} = 0_{2k\times 1}. \tag{23}$$

We say that the singular perturbation model (21) is in standard form if and only if the above algebraic equations have at least one isolated real root for $\mathbf{y}$ in terms of $\mathbf{x}$. We will proceed with the stability analysis with the assumption that Eq. (23) has at least one feasible solution $\mathbf{y}_0(\mathbf{x})$. We now look at the solution to Eq. (21) with $\mathbf{y} = \mathbf{y}_0(\mathbf{x})$ and $\epsilon = 0$. This solution, denoted by $\mathbf{x}_0(\bar{t})$, is the solution of the reduced rigid-body system.

Assume that $\mathbf{x}_0(\bar{t})$ is defined for $\bar{t} \in [0, \bar{t}_1]$. At an arbitrary time instance $\bar{t}_0 \in [0, \bar{t}_1]$, the boundary layer system of (22) can be introduced through a ''stretch'' of the time scale, $\tau = (\bar{t} - \bar{t}_0)/\sqrt{\epsilon}$. In the stretched time scale τ, the variables $\bar{t}$ and $\mathbf{x}(\bar{t},\epsilon)$ are slowly varying. Since $\bar{t}_0$ is allowed to take any value in $[0, \bar{t}_1]$, the boundary layer system of (22) can be written with τ as the independent variable:

$$\mathbf{y}' = \begin{pmatrix} y_2 \\ A(\mathbf{x}_0,\mathbf{y},0)\bar{\lambda}(\mathbf{y},0) + B(\mathbf{x}_0,\mathbf{y},0) \end{pmatrix} \tag{24}$$

where $'$ now denotes differentiation with respect to τ. Let

$$\begin{pmatrix} z \\ z' \end{pmatrix} = \mathbf{y} - \mathbf{y}_0(\mathbf{x}_0).$$

Perform a linearization of the boundary layer model (24) around the equilibrium solution $\mathbf{y}_0(\mathbf{x}_0)$. We obtain the homogeneous boundary layer dynamics of the form

$$z'' + Pz' + Qz = 0_{k\times 1}, \tag{25}$$

with

$$P = -\left.\frac{\partial(A(\mathbf{x}_0,\mathbf{y},0)\bar{\lambda}(\mathbf{y},0) + B(\mathbf{x}_0,\mathbf{y},0))}{\partial y_2}\right|_{\mathbf{y}=\mathbf{y}_0(\mathbf{x}_0)},$$

$$Q = -\left.\frac{\partial(A(\mathbf{x}_0,\mathbf{y},0)\bar{\lambda}(\mathbf{y},0) + B(\mathbf{x}_0,\mathbf{y},0))}{\partial y_1}\right|_{\mathbf{y}=\mathbf{y}_0(\mathbf{x}_0)}.$$

The response of the above system equation, $z(\tau)$, is the transient that describes the dynamics associated with the compliance at the contact points. The stability of the system implies the convergence of the compliant contact model solution $\mathbf{x}(\bar{t},\epsilon)$ to the rigid-body model solution $\mathbf{x}_0(\bar{t})$. We can directly apply Tikhonov's theorem [19] to get the following result:

THEOREM 4.1 *Consider the system described by (21) and let* $\mathbf{y}_0(\mathbf{x})$ *be an isolated solution of (23). If the following three conditions are satisfied by* $(\mathbf{x}(\bar{t}),\mathbf{y}(\bar{t}),\epsilon)$ *for all* $\bar{t} \in [0,\bar{t}_1]$ *and* $\epsilon \in [0,\epsilon_0]$: *(a) the terms on the right-hand side of (21) and their first partial derivatives with respect to* $(\mathbf{x},\mathbf{y},\epsilon)$ *are bounded and continuous; (b) the origin of the boundary layer system (25) is exponentially stable; and (c)* $\mathbf{y}_0(\mathbf{x})$ *has continuous first partial derivatives with respect to its arguments, then the following are true:*

- *The reduced rigid-body model, obtained from* (21) *by substitution of* $\mathbf{y} = \mathbf{y}_0(\mathbf{x})$ *and* $\epsilon = 0$, *has a unique bounded solution,* $\mathbf{x}_0(\bar{t})$, *for all* $\bar{t} \in [\bar{t}_0, \bar{t}_1]$, *where* $\bar{t}_0 \in [0, \bar{t}_1]$.
- *There exist positive constants* δ_0 *and* ϵ_0 *such that for the initial conditions* $\mathbf{x}(\bar{t}_0,\epsilon)$ *and* $\mathbf{y}(\bar{t}_0,\epsilon)$ *satisfying* $\|\mathbf{y}(\bar{t}_0,\epsilon) - \mathbf{y}_0(\mathbf{x}(\bar{t}_0,0))\| < \delta_0$ *and* $0 < \epsilon < \epsilon_0$, *the singular perturbation problem has a unique solution* $\mathbf{x}(\bar{t},\epsilon)$ *and* $\mathbf{y}(\bar{t},\epsilon)$ *on the interval* $[\bar{t}_0, \bar{t}_1]$ *and*

$$\mathbf{x}(\bar{t},\epsilon) - \mathbf{x}_0(\bar{t}) = O(\sqrt{\epsilon}), \quad \mathbf{y}(\bar{t},\epsilon) - \mathbf{y}_0(\bar{t}) = O(\sqrt{\epsilon}).$$

Proof: The proof of this theorem follows directly from Tikhonov's theorem, and is a direct application of Theorem 9.1 in [19].

Remark 4.1

- The stability of the boundary layer system is determined by the matrices P and Q, or specifically, the eigenvalues of $\begin{pmatrix} 0_{k\times k} & I_{k\times k} \\ -Q & -P \end{pmatrix}$.

- Other than the general constitutive model described by Eqs. (7) and (8), no specific compliant contact models are introduced in the discussion, therefore the stability results will not change when different compliant models are employed.

It is worth noting that the requirements on the continuity of the first partial derivatives in Theorem 4.1 are not satisfied whenever there are transitions from rolling to sliding or sliding to rolling because of the nonsmooth nature of Coulomb's law. In the next section, we will apply Theorem 4.1 to planar mechanical systems with one contact and discuss the cases of sliding and rolling separately.

5 Planar Mechanical Systems With One Contact

Consider the planar rigid body depicted in Fig. 2 in contact with a horizontal surface, where L is the distance from the contact point to the center of mass (CM). The rigid body has mass m and centroidal moment of inertia I. $q=(y\ x\ \theta)^T$ represent the generalized coordinates for the rigid body which are the position of the CM and the angular orientation. (F_x,F_y) are the external forces acting on the body and F_θ is the external moment about the CM. μ is the coefficient of friction between the rigid body and surface. The equations of motion for the system with one contact are given by Eq. (1) with

$$q=\begin{pmatrix} y \\ x \\ \theta \end{pmatrix},\quad M=\begin{pmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & I \end{pmatrix},\quad u=\begin{pmatrix} F_y \\ F_x \\ F_\theta \end{pmatrix},\quad \text{and}\quad h=0_{3\times 1}. \tag{26}$$

For sliding contact:

$$\Phi_q^T=(1\ \ \mu_s\ \ \mu_s L\sin\theta - L\cos\theta)^T,\ \ \lambda=\lambda_N=\lambda_{NS}, \tag{27}$$

and for rolling contact:

$$\Phi_q^T=\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -L\cos\theta & L\sin\theta \end{pmatrix},\quad \lambda=\begin{pmatrix} \lambda_{NR} \\ \lambda_{TR} \end{pmatrix}, \tag{28}$$

where $\mu_s=-\mu\ \text{sign}(\dot{\Phi}_T)$.

5.1 Sliding Contact. For the sliding case, the rigid-body dynamics can be modeled as a LCP of the form

$$\ddot{\Phi}_N=A\lambda_N+B \tag{29}$$

where

$$\ddot{\Phi}_N\geqslant 0,\ \ \lambda_N\geqslant 0,\quad \text{and}\quad \ddot{\Phi}_N\lambda_N=0,$$

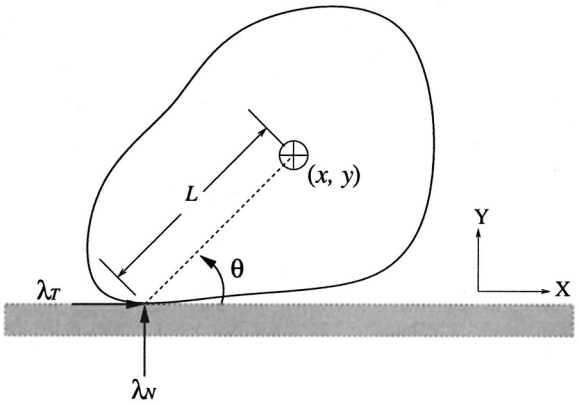


Fig. 2 Planar rigid body in contact with a rough surface

$$A=\frac{1}{m}+\frac{L^2\cos\theta}{I}(\cos\theta-\mu_s\sin\theta),$$

$$B=L\dot{\theta}^2\sin\theta+\frac{F_y}{m}-\frac{L\cos\theta F_\theta}{I}.$$

Note with $\mu_s=0$, $A>0$ and therefore A is a 1×1 P matrix, which guarantees a unique solution. If $B\geqslant 0$ we are guaranteed of the existence of a solution regardless of whether or not A is a P matrix.

We now proceed with the singular perturbation approach to the problem. A transformation to a system of fast and slow variables can be accomplished by making the change of variables as

$$q_1=y,\ \ q_2=\begin{pmatrix} x \\ \theta \end{pmatrix},\ \ p_1=\Phi_N,\quad \text{and}\quad p_2=q_2.$$

Use (19) and (20) to nondimensionalize the state variables with

$$D_1=d_1,\ \ D_2=\begin{pmatrix} L & 0 \\ 0 & 1 \end{pmatrix},\quad \text{and}\quad \epsilon=\frac{d_1}{L}.$$

Let the characteristic mass $D_M=m$, the mass of the rigid body. The standard singular perturbation form of the planar rigid-body system with one sliding contact is given by

$$\begin{cases} \sqrt{\epsilon}\mathbf{y}'=\begin{pmatrix} y_2 \\ A(\mathbf{x})\bar{\lambda}_N(\mathbf{y},\epsilon)+B(\mathbf{x}) \end{pmatrix} \\ \mathbf{x}'=\begin{pmatrix} x_2 \\ \mu_s\bar{\lambda}_N(\mathbf{y},\epsilon)+\bar{F}_x \\ \dfrac{mL^2}{I}(\mu_s\sin x_{1,2}-\cos x_{1,2})\bar{\lambda}_N(\mathbf{y},\epsilon)+\bar{F}_\theta \end{pmatrix} \end{cases} \tag{30}$$

where

$$\bar{\lambda}_N(\mathbf{y},\epsilon)=\frac{f_N(-\epsilon L y_1)+g_N\left(-\epsilon L y_1,-\dfrac{\epsilon L}{T}\mathbf{y}_2\right)}{\dfrac{mL}{T^2}}\geqslant 0,$$

$$\begin{gathered} A(\mathbf{x})=1+\frac{mL^2}{I}\cos x_{1,2}(\cos x_{1,2}-\mu_s\sin x_{1,2}), \\ B(\mathbf{x})=x_{2,2}^2\sin x_{1,2}-\bar{F}_\theta\cos x_{1,2}+\bar{F}_y, \\ \bar{F}_x=\frac{F_{\mathbf{x}}}{\dfrac{mL}{T^2}},\ \ \bar{F}_y=\frac{F_{\mathbf{y}}}{\dfrac{mL}{T^2}},\quad \text{and}\quad \bar{F}_\theta=\frac{F_\theta}{\dfrac{I}{T^2}}. \end{gathered} \tag{31}$$

Note that the symbol $x_{i,j}$ refers to the jth component of the vector x_i. By following through the same derivations given by Eqs. (21)–(25) in Section 4, the linearized boundary layer model of (30) can be obtained as

$$z''-A(\mathbf{x}_0)\left(\frac{\partial\bar{\lambda}_N(\mathbf{y},0)}{\partial y_2}\right)\Bigg|_{\mathbf{y}=\mathbf{y}_0(\mathbf{x}_0)} z'-A(\mathbf{x}_0) \times\left(\frac{\partial\bar{\lambda}_N(\mathbf{y},0)}{\partial y_1}\right)\Bigg|_{\mathbf{y}=\mathbf{y}_0(\mathbf{x}_0)} z=0, \tag{32}$$

where $\mathbf{y}_0(\mathbf{x})$ is a solution of the algebraic equations obtained by setting $\epsilon=0$ in (30), and $\mathbf{x}_0(t)$ is the solution of the reduced system of (30) corresponding to $\mathbf{y}_0$.

In a general viscoelastic model, it is reasonable to assume that $\bar{\lambda}_N$ is a monotonically decreasing function with respect to y_1 and y_2. With this assumption, if $A(\mathbf{x}_0,0)$ is positive, the boundary layer system (25) is stable, and the stability of the solution for the

Table 1 LCP and stability results for one sliding contact (*C* = contact, *NC* = no contact, *NS* = no solution, *RB* = rigid-body model, *CC* = compliant contact model; * indicates that the stability result comes from the fact that the noncontact solutions (free falling) are always stable)

Conditions		Solutions	Stability	Preferred Model
$A>0$	$B\geq 0$	*NC*	stable*	*RB*
$A>0$	$B<0$	*C*	stable	*RB*
$A=0$	$B>0$	*NC*	stable*	*RB*
$A=0$	$B=0$	∞ solns.	-	*CC*
$A=0$	$B<0$	*NS*	N/A	*CC*
$A<0$	$B\geq 0$	*C*	unstable	discard
		NC	stable*	*RB*
$A<0$	$B<0$	*NS*	N/A	*CC*

singular perturbation problem (21) is guaranteed if all other conditions in Theorem 4.1 are satisfied. In a single point sliding contact problem, the rigid-body LCP formulation (29) has a unique solution if and only if $A(\mathbf{x}_0,0)$ is positive. The above conclusion can be summarized into the following theorem.

THEOREM 5.1 *For a planar rigid body with a single sliding contact described by (1) and (27), the solution obtained from the compliant contact model converges to that obtained from the rigid-body model if and only if there exists a unique solution for the rigid body LCP formulation (29). This result is independent of the compliant contact models as long as the monotonicity condition,* $-\partial\overline{\lambda}_N(\mathbf{y}(t),0)/\partial y_{1,2}>0$, *is satisfied.*

As examples, we show that the stability results are the same for both the Kelvin-Voigt and the Hunt-Crossley models. From Eq. (10) and the expression in (31), the dimensionless normal contact force for Kelvin-Voigt model can be written as

$$\overline{\lambda}_N(\mathbf{y},\boldsymbol{\epsilon})=\overline{\lambda}_N(\mathbf{y})=-\overline{K}y_1-\overline{C}y_2. \tag{33}$$

where the nondimensional stiffness and the damping are defined as

$$\overline{K}=\frac{K}{\frac{1}{\epsilon L}\frac{mL}{T^2}}>0 \quad \text{and} \quad \overline{C}=\frac{C}{\frac{1}{\sqrt{\epsilon}L/T}\frac{mL}{T^2}}>0. \tag{34}$$

The boundary layer system is obtained as

$$z''+A(\mathbf{x}_0)\overline{C}z'+A(\mathbf{x}_0)\overline{K}z=0. \tag{35}$$

For Hunt-Crossley model, the normal contact force can be expressed as

$$\overline{\lambda}_N(\mathbf{y},\boldsymbol{\epsilon})=\overline{\lambda}_N(\mathbf{y})=\overline{K}(-y_1)^{\beta}-\overline{\alpha}\overline{K}(-y_1)^{\beta}y_2, \tag{36}$$

where

$$\overline{K}=\frac{K}{\frac{1}{(\epsilon L)^{\beta}}\frac{mL}{T^2}}>0 \quad \text{and} \quad \overline{\alpha}=\frac{3}{2}\frac{\alpha}{\frac{1}{\sqrt{\epsilon}L/T}}>0. \tag{37}$$

The linearized boundary layer system for Hunt-Crossley model is given by

$$z''+A(\mathbf{x}_0)\overline{\alpha}\overline{K}(-y_1)^{\beta}|_{\mathbf{y}=\mathbf{y}_0(\mathbf{x}_0)}z'+A(\mathbf{x}_0)\beta\overline{K}(-y_1)^{\beta-1}|_{\mathbf{y}=\mathbf{y}_0(\mathbf{x}_0)}z=0. \tag{38}$$

Since $y_1\leq 0$ for any active constraint, it is clear that the stability of the boundary layer dynamics, described by either (35) or (38), depends entirely on the value of $A(\mathbf{x}_0)$. Thus, independent of the choice of contact model, $A(\mathbf{x}_0)$ may used to test contact force stability in those situations where the LCP tells us that the contact is maintained. A summary of the results is given in Table 1. For the contact maintaining solutions, the result of the singular perturbation analysis states that stability only occurs where the quantity A in the LCP formulation is positive (P matrix). If the LCP reports an unique solution, we use the rigid-body model to simulate the dynamic motion. For the case when the LCP has two solutions ($A<0, B\geq 0$), we can still use the rigid-body model since the stability analysis shows a unique stable solution.

5.2 Rolling Contact. The rigid-body dynamics can once again be formulated as an LCP with the help of surplus and slack variables ([7]). The singular perturbation analysis proceeds in exactly the same way as in the previous section. The following is a partition of the generalized coordinates for the rolling case:

$$q_1=\begin{pmatrix} y \\ x \end{pmatrix}, \quad q_2=\theta, \quad p_1=\begin{pmatrix} \Phi_N \\ \Phi_T \end{pmatrix}, \quad \text{and} \quad p_2=q_2.$$

The linearized boundary layer model for this case is given by

$$z''-A(\mathbf{x}_0)\left(\frac{\partial\overline{\lambda}(\mathbf{y},0)}{\partial y_2}\right)\Bigg|_{\mathbf{y}=\mathbf{y}_0(\mathbf{x}_0)}z'-A(\mathbf{x}_0)\left(\frac{\partial\overline{\lambda}(\mathbf{y},0)}{\partial y_1}\right)\Bigg|_{\mathbf{y}=\mathbf{y}_0(\mathbf{x}_0)}z=0_{2\times 1}, \tag{39}$$

where

$$\overline{\lambda}=(\overline{\lambda}_{NR}^T \ \ \overline{\lambda}_{TR}^T)^T,$$

$$A(\mathbf{x})=\begin{pmatrix} 1+\frac{mL^2}{I}\cos^2 x_1 & -\frac{mL^2}{I}\sin x_1\cos x_1 \\ -\frac{mL^2}{I}\sin x_1\cos x_1 & 1+\frac{mL^2}{I}\sin^2 x_1 \end{pmatrix},$$

$$B(\mathbf{x})=\begin{pmatrix} \overline{F}_y-\overline{F}_\theta\cos x_1+x_2^2\sin x_1 \\ \overline{F}_x+\overline{F}_\theta\sin x_1+x_2^2\cos x_1 \end{pmatrix}. \tag{40}$$

In the above system, $A(\mathbf{x}_0)$ is symmetric, and its eigenvalues are given by $a_1=1$, $a_2=1+mL^2/I$ which are positive real numbers. Also if $\overline{\lambda}$ is a monotonically decreasing function with respect to $\mathbf{y}$, both $-\partial\overline{\lambda}(\mathbf{y}_0,0)/\partial y_1$ and $-\partial\overline{\lambda}(\mathbf{y}_0,0)/\partial y_2$ are diagonal matrices with positive entries. In this situation, the stability of the boundary layer system (39) follows from the Routh-Hurwitz criterion. The reason is that for rolling constraints, the contact model corresponds instantaneously to a frictionless (no dissipation) complaint pin joint. Viewed in this context, the contact forces correspond to the joint constraint forces. It is not surprising that, in the rigid-body limit, these forces are always stable. In contrast, the singular perturbation analysis of sliding included the dependence of tangential friction force on normal force. This dependence produced the potential for instability during sliding. Since the LCP and singular perturbation analyses for sliding both included this dependence, it was possible in Theorem 5.1 to relate the LCP existence and uniqueness results to the singular perturbation stability result.

There are three possible solutions for the LCP formulation of a system with a rolling contact: (a) breaking contact, (b) continued rolling, and (c) transition to sliding. The conditions of Theorem 4.1 for use of the rigid-body model include continuity and differentiability of the tangential contact forces. These conditions are not met during (a) or (c) because the contact forces need only be C^0 continuous at a transition. Therefore, we cannot derive a result similar to Theorem 5.1 for rolling contacts. It is possible, however, to state the more conservative result:

THEOREM 5.2 *For a planar rigid-body with a single rolling contact described by (1) and (28), the solution obtained from the compliant contact model converges to that obtained from the rigid-body model whenever the LCP formulation yields a unique*

solution corresponding to continued rolling. This result is independent of the compliant contact models as long as the monotonicity condition of λ *is satisfied.*

5.3 Extensions. In the treatment thus far, we considered the dynamics of a single rigid body in which the unilateral constraints were due to one contact (sliding or rolling) with a second fixed rigid body. When we consider multiple planar rigid bodies with bilateral constraints, but only one contact, a similar result can be derived. In such a case, the dynamics formulations in the Cartesian space and the constraints can still be described by (1) and (27), (28), if the operational space inertia matrix, M, exists. The only differences are that the inertia matrix, if it exists, is no longer diagonal but symmetric and still positive definite, and $h(q,\dot{q})$ is no longer zero. But these differences will not affect the properties of the A matrix in the boundary layer systems (32) and (39). Consequently, the basic ideas developed in this section are still valid, and the main results are applicable to any mechanical system in which the unilateral constraints are due to a single contact.

6 Results From Experiments and Simulations

In this section, we compare the results of numerical simulations with experimental observations. In the experiments, an aluminum rod with spherical ends is released from rest, while contacting a flat, rough, fixed surface, with different initial positions. We used the OPTOTRAK-3020 (Northern Digital, Inc.), a noncontact three-dimensional motion measurement system with an accuracy better than 0.1 mm in each coordinate direction and a tracking rate can be as high as 1000Hz. The experimental setup is shown in Fig. 3.

The numerical simulation is based on the dynamics given by Eqs. (1) and (27). The length and diameter of the tested rod are 0.468 m and 0.00948 m, respectively. The mass is 0.088 kg. The compliant model used in the simulations is the Hunt-Crossley model expressed by Eqs. (36)–(37). The only unknown parameter in the equations is the coefficient of friction. The coefficient of friction for the simulation is chosen to be the value that best approximates experimentally observed trajectories in a least-squares sense.

6.1 Case 1: The LCP has a Unique Solution. We first consider an experiment in which our rigid-body LCP predicts a unique solution throughout the duration of the experiment. The initial conditions of the rod are $\theta = 42.3$ deg, $\dot{\theta} = 0$, and $\dot{x} = \dot{y} = 0$. The external force are $F_x = 0$, $F_y = -\text{mg}$, and $M_z = 0$. The related parameters used in Hunt-Crossley model are $\overline{K} = 1$, $\overline{\alpha} = 1$, and $\beta = 2$. The trajectory corresponds to a condition of sliding where the contact point slides to the left. The sliding velocity decreases, and at approximately 0.205 sec, the sliding velocity changes direction so that the contact point slides to the right. In Fig. 4, we show the experimentally observed trajectory and the simulation results for (a) the trajectory of contact point, (b) the normal contact force, and (c) the tangential contact force. The simulation results are provided for the rigid-body LCP solution and for the compliant contact model for a range of ϵ values. The coefficient of friction used in the simulation is $\mu = 0.27$.

The first thing to note is that there is a close agreement between the experimental trajectory and the rigid-body LCP solution with the same initial condition as expected. The second issue to focus on is the set of results from the simulation of the compliant contact model. Even though the initial condition for the compliant contact model solution is different from the equilibrium solution, it quickly converges to the equilibrium solution. The convergence in an absolute time scale is faster as ϵ becomes smaller. This is also evident at the transition from reverse to forward sliding, which includes a very brief period of rolling. The discontinuity of the rigid-body dynamic model with Coulomb friction is seen in the contact force variation in Figs. 4(b) and 4(c). However, the compliant contact model yields a continuous solution that can be made to approach the solution of the LCP model arbitrarily closely by letting ϵ assume very small values.

6.2 Case 2: The LCP has Two Solutions. In Section 5 we showed that, in cases when the LCP formulation for sliding contact has two solutions, the model of the boundary layer system (32) predicts that the contact maintaining solution is unstable. In such cases, at any instant, the LCP predicts two possible outcomes. While it is possible to simulate either outcome using the compliant contact model, a simulation based on the rigid-body model involves making a choice at each such point.

An ideal uniform rod with $L = 1$ m and $m = 1$ kg is used in the simulation. The initial condition of the rod are $\theta = 70$ deg, $\dot{\theta} = 0$, and $\dot{x} = \dot{y} = 0$. The external forces are $F_x = 0$, $F_y = -\text{mg}$, and $M_z = -1$ Nm. The coefficient of friction is $\mu = 1$. The compliant contact model used in the simulation is the Hunt-Crossley model with the same parameters as in Case 1. Because this case can only be achieved at carefully chosen values of external forces or initial velocities, we were unable to reproduce this case experimentally.

In Fig. 5, we show the results of the rigid-body solution assuming that (a) the contact breaks at $t = 0$—the first solution; and (b) the contact is maintained at $t = 0$—the second solution, and at future time instants as well. The main point to be observed in Fig. 5 is the performance of the compliant contact model. As shown in Fig. 5(b), even when started from the condition of maintaining contact, the solution for the compliant contact model exponentially converges to the stable solution of no contact. The rate of

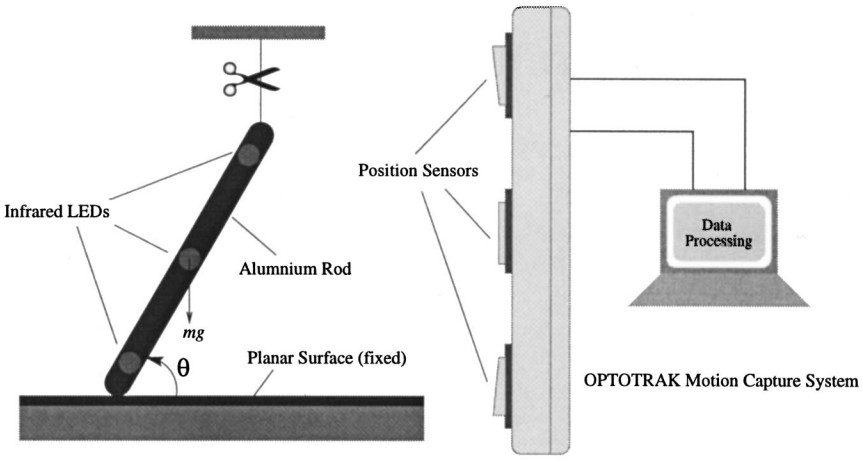
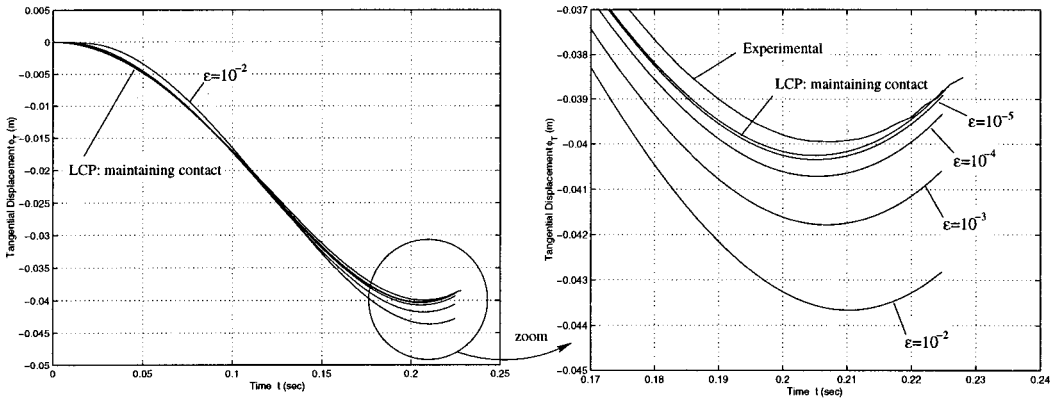
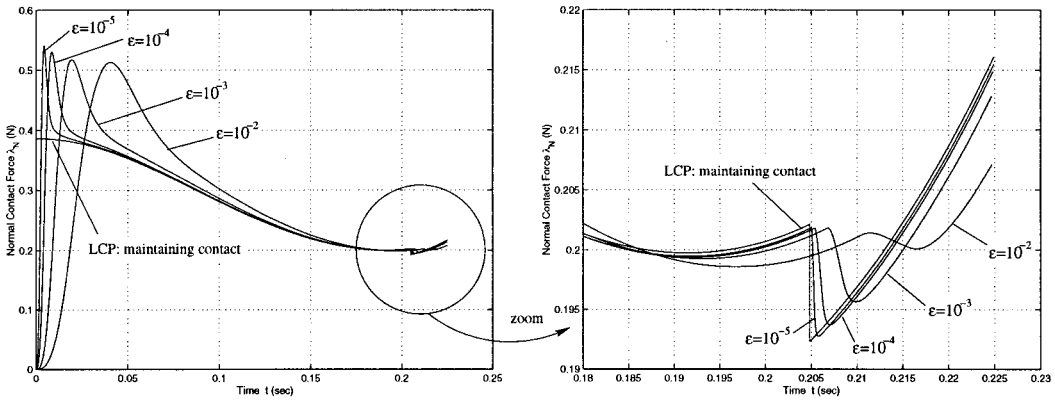


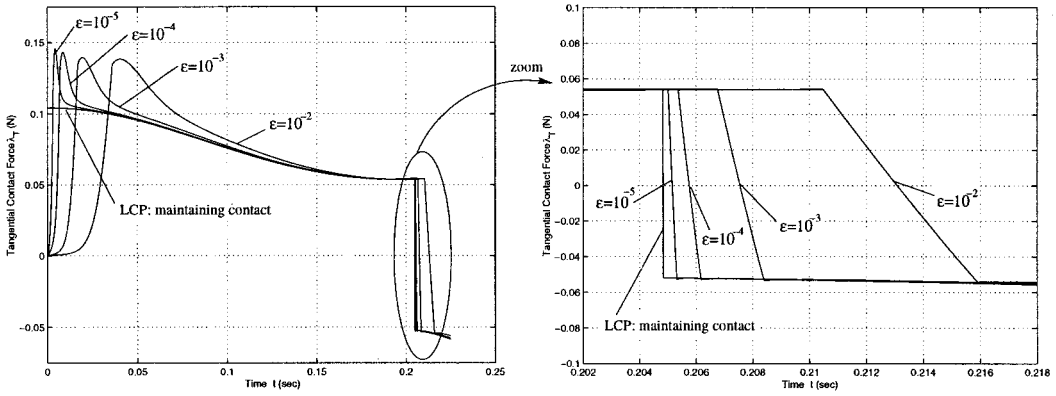
Fig. 3 The experimental setup



(a) Tangential displacement of the rod at the contact point



(b) History of the normal contact force



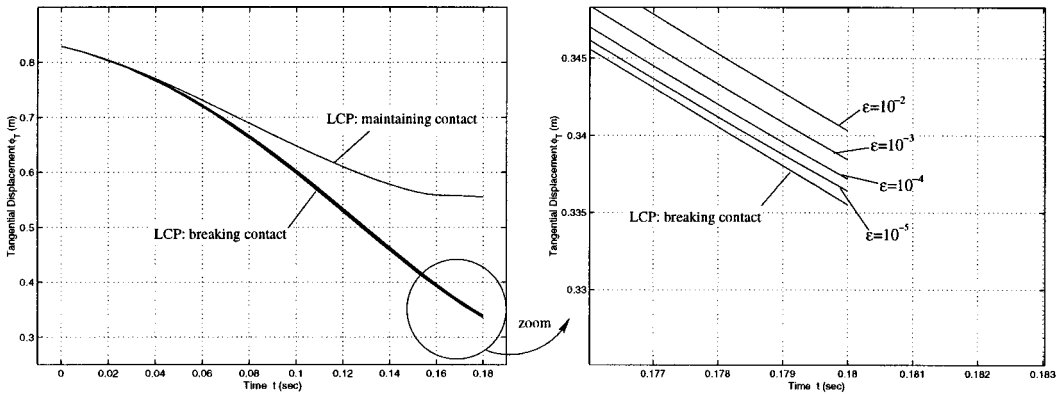
(c) History of the tangential contact force

Fig. 4 Case 1: The LCP has a unique solution and the compliant contact model solution converges to the rigid-body model solution as the perturbation parameter ϵ goes to 0

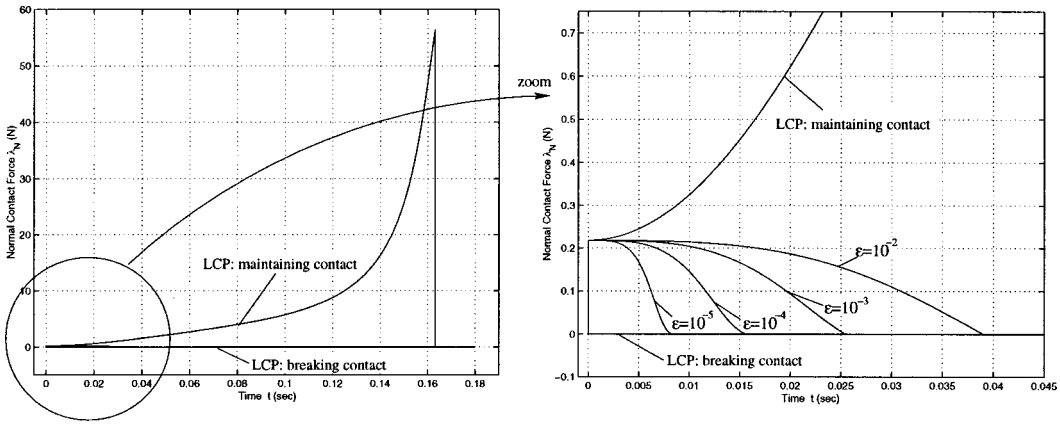
convergence increases with decreasing ϵ. In contrast, at $t = 0.163$ sec, the rigid-body solution corresponding to maintaining contact reaches a state where the LCP has a unique solution corresponding to contact separation. This can be seen in Fig. 5(*b*) as the discontinuous drop in normal contact force. The fact that the compliant model solution converges to the stable rigid-body model solution indicates that in cases when LCP has two solutions, one stable and one unstable, we can always choose the stable solution and use the rigid body model to continue the simulation.

7 Friction Models

There are many types of friction phenomena and equations to model them. Coulomb friction is one of the simplest and in many



(a) Tangential displacement of the rod at the contact point



(b) History of the normal contact force

Fig. 5 Case 2: The LCP has two solutions, maintaining contact (unstable) and separation (stable). If the compliant model solution is started with the unstable maintaining contact solution, it quickly converges to the separation solution (stable).

situations can adequately predict the system's behavior. Nevertheless, its mathematical properties complicate dynamic simulation for both rigid-body and compliant contact models. The difficulties caused by Coulomb's friction model in rigid-body dynamic simulation are due to the following issues: (1) the friction force is not smooth during rolling-sliding transitions; and (2) during rolling, the friction force cannot be directly determined from the state variables. When solving the forward dynamic problem, these two issues can either increase the complexity of the system or cause analytical difficulties. Specifically the rolling and sliding constraints need to be handled differently in the rigid-body formulations ([7]). This is also the main reason that our stability results in Section 4 are not applicable to transitions from rolling to sliding. Furthermore, cases arise in which a unique solution to the forward dynamics problem does not exist.

Since these difficulties are due to the Coulomb model, it is possible to overcome them by substituting a model with the requisite mathematical properties. In fact, nonclassical friction laws which are nonlinear and nonlocal have been found to be superior to pointwise Coulomb models from both a phenomenological and a computational viewpoint ([22]). A few of these models were developed specifically for rigid-body dynamics. For example, a discontinuous model that extends the Coulomb's stiction zone from zero velocity to a small neighborhood of zero velocity is suggested in ([23]). Many others ([12]), including Dahl's model and the bristle model, can be considered to be extensions of the compliant contact model.

Our interest is in the simplest friction law that approximates Coulomb friction and is a continuously differentiable function of the system states. Such a model would allow us to formulate the dynamics using either rigid-body models or compliant contact

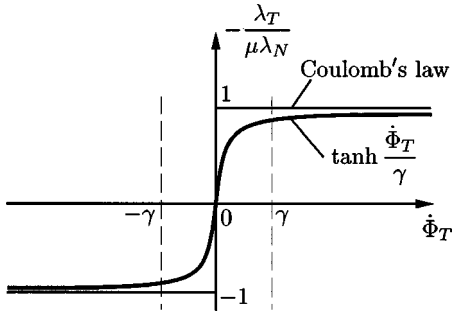
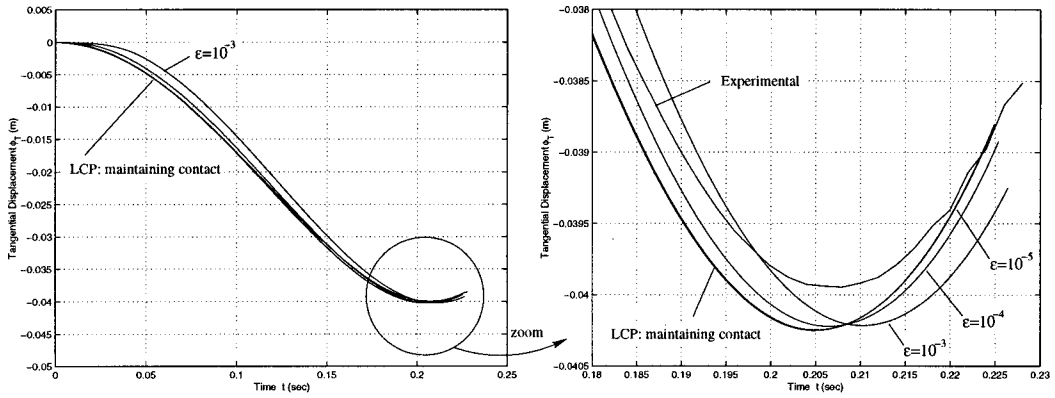
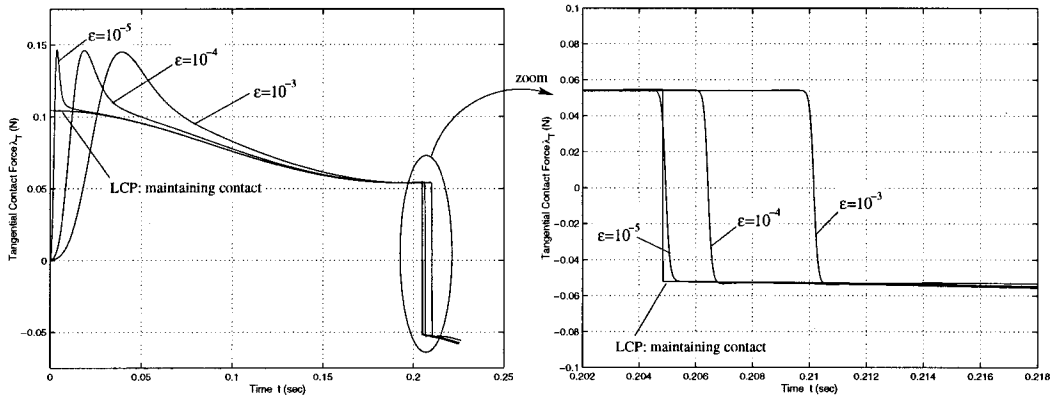


Fig. 6 A smooth, nonlinear friction law with two parameters γ, a characteristic speed, and μ, the coefficient of friction



(a) Tangential displacement of the rod at the contact point



(b) History of the tangential constraint force

Fig. 7 Results with the smooth nonlinear friction law ($\gamma = 10^{-3}$). The transition from reverse sliding to rolling to forward sliding at $t = 0.205$ sec is characterized by a smooth variation of contact forces.

models while improving the performance of both. A friction law with these properties is shown in Fig. 6. It has a one-to-one correspondence between the friction force λ_T and the relative tangential velocity $\dot{\Phi}_T$. The small parameter γ defines the extent of the ''rolling'' regime: $\dot{\Phi}_T \in [-\gamma, \gamma]$. By letting γ tend to zero, we can obtain an arbitrarily close approximation to Coulomb's law, albeit with some sliding in the ''rolling'' regime. Using this friction law with the rigid-body dynamic model expressed in Eqs. (1)–(2), we no longer need to differentiate between rolling and sliding contacts. Instead, the Jacobian matrix (Φ_q) in Eq. (3) always takes the form

$$\Phi_q^T = \left[\Phi_{N_q}^T - \Phi_{T_q}^T \operatorname{diag}\left(\mu \tanh\left(\frac{\dot{\Phi}_T}{\gamma} \right) \right) \right]. \tag{41}$$

This gives rise to a unified LCP formulation that works for both rolling and sliding constraints. The smooth friction law can also improve numerical performance of the compliant contact model, since we no longer need the deformation state vector, δ_T, to compute the tangential force which is uniquely defined by the normal contact force and the relative velocity.

$$\lambda_{T,i} = -\mu \tanh\left(\frac{\dot{\Phi}_{T,i}}{\gamma} \right) \lambda_{N,i} \tag{42}$$

As is the case with rigid-body dynamics and Coulomb's law, the LCP formulation with the smooth friction law will have situations with no solution or multiple solutions. And, as before, the compliant contact model given by Eqs. (12) and (42) resolves the difficulties with uniqueness and existence. Since the stability analysis in Section 4 can be easily applied to this new frictional model without worrying about transitions between rolling and sliding contacts, Theorem 5.1 can be directly extended to any planar mechanical system with a single rolling or sliding contact. Because the rolling constraint is now replaced by ''microsliding'' with the tangential contact velocity smaller than γ, the proof follows exactly the same lines of the proof in Section 5.1.

To illustrate the effect of the smooth friction law, we consider the same situation shown in Fig. 4 (Case 1 in Section 6). Recall the transition from reverse sliding to rolling to forward sliding in Fig. 4 at $t = 0.205$ sec. Figure 7 shows the results of the simulation with the smooth friction law with the same initial conditions. The rigid-body model predicts discontinuities in the contact forces. However, the compliant contact predicts a smooth transition from sliding to sticking (relative velocity less than the threshold γ) to sliding in the opposite direction.

The main disadvantage of the nonlinear friction law is due to the fact a static friction force can only be maintained through ''creep'' in the tangential direction. While the ''creep'' rate is less than γ, and γ can be set to a very small value, it is not a very attractive solution because it has the adverse effect of making the system of ODEs stiff. There is a natural tradeoff that must be considered in selecting the parameter γ.

8 Concluding Remarks

When rigid-body models are used in conjunction with Coulomb friction for dynamic simulation of systems with frictional contacts, there may be situations in which there are no solutions or multiple solutions for the contact forces and the accelerations. In this paper, we describe a contact model that models the small compliance in the normal and tangential directions. We show that this compliant contact model, when used with the rigid-body dynamic equations of motion, always yields a unique solution for the accelerations and the forces. While this model is superior to the traditional rigid-body model in terms of accuracy and robustness, it is also more complex and requires a larger number of parameters. Therefore, it is appealing to use rigid-body models, whenever concerns of uniqueness and existence do not arise.

The main contribution of this paper is the use of singular perturbation theory to establish conditions under which solutions from the rigid-body model are stable, or in other words, conditions in which the compliant contact model solution converges exponentially to the rigid-body model solution. In situations when rigid-body LCP analysis reveals multiple solutions, stability analysis can resolve the ambiguity. We can simply discard the unstable solutions and retain the stable one. The stability analysis shows when it is essential to pursue the more sophisticated compliant contact model, and when it is satisfactory to neglect the fast dynamics. The basic issues are illustrated with the help of a simple example with one contact that may be rolling, sliding, or separating. The case of rolling contacts poses an additional difficulty because of the fact the tangential forces obtained by Coulomb-like frictional laws, even when used with compliant contact models, are not smooth functions of the state. The second main contribution of the paper is the result that a smooth nonlinear friction law, inspired by Oden and Pires' nonlinear friction law ([22]), overcomes this difficulty. We show that in the case of planar mechanical systems with one contact, there are at most two solutions, and there is only one stable solution.

The basic ideas of this paper are applicable to any situation with frictional contacts. However, in order for the rigid-body model, and therefore the perturbation analysis to be applicable, we are limited to planar problems with three or less independent constraints and spatial problems with six or less independent constraints. Note the compliant contact model can always be applied without such limitations. Since not all of the constraints of the physical system are embodied in the rigid-body mathematical model, a study of the stability of these solutions based solely on the structure of the LCP itself is not justified. Existence and uniqueness problems suggest the inapplicability of the rigid-body model altogether and not simply uncertainty in or sensitivity to model parameter values.

Our future work addresses incorporating stability analysis as a diagnostic tool in real-time simulation where it is prudent to check for stability and warn the user in unstable regimes.

Acknowledgment

The support of NSF grants GRT-9355018, MIP-9617997 and CISE/CDS-9703220, and ARO grant MURI/DAAH04-9610007 are gratefully acknowledged.

References

[1] Mirtich, B., Zhuang, Y., Goldberg, K., Craig, J., Zanutta, R., Carlisle, B., and Canny, J., 1996, ''Estimating Pose Statistics for Robotic Part Feeders,'' *Proc. of the 1996 IEEE Int'l Conf. on Robotics and Automation*, pp. 1140–1146.

[2] Donald, B. R., and Pai, D. K., 1990, ''On the Motion of Compliantly Connected Rigid Bodies in Contact: A System for Analyzing Designs for Assembly,'' *Proc. of the 1990 IEEE Int'l Conf. on Robotics and Automation*, pp. 1756–1762.

[3] Howe, R. D., and Cutkosky, M. R., 1996, ''Practical Force-Motion Models for Sliding Manipulation,'' Int. J. Robot. Res., **15**, pp. 557–572.

[4] Song, P., Yashima, M., and Kumar, V., 2000, ''Dynamic Simulation for Grasping and Whole Arm Manipulation,'' *Proc. of the 2000 IEEE Int'l Conf. on Robotics and Automation*, Vol. 2, pp. 1082–1087.

[5] Kumar, V., and Waldron, K. J., 1989, ''Actively Coordinated Vehicle Systems,'' ASME J. Mech. Des., **111**, pp. 223–231.

[6] Sacks, E., and Joskowicz, L., 1995, ''Computational Kinematic Analysis of Higher Pairs With Multiple Contacts,'' ASME J. Mech. Des., **117**, pp. 269–277.

[7] Trinkle, J., Pang, J.-S., Sudarsky, S., and Lo, G., 1997, ''On Dynamic Multi-Rigid-Body Contact Problems With Coulomb Friction,'' Z. Angew. Math. Mech., **77**, No. 4, pp. 267–280.

[8] Lötstedt, P., 1981, ''Coulomb Friction in Two-Dimensional Rigid Body Systems,'' Z. Angew. Math. Mech., **61**, pp. 605–615.

[9] Mason, M. T., and Wang, Y., 1988, ''On the Inconsistency of Rigid-Body Frictional Planar Mechanics,'' *Proc. of the 1988 IEEE Int'l Conf. on Robotics and Automation*, pp. 524–528.

[10] Wang, Y.-T., and Kumar, V., 1994, ''Simulation of Mechanical Systems With Unilateral Constraints,'' ASME J. Mech. Des., **116**, No. 2, pp. 571–580.

[11] Howard, W. S., and Kumar, V., 1993, ''A Minimum Principle for the Dynamic Analysis of Systems With Frictional Contacts,'' *Proc. of the 1993 IEEE Int'l Conf. on Robotics and Automation*, Vol. 1, pp. 437–442.

[12] Armstrong-Hélouvry, B., Dupont, P., and de Wit, C. C., 1994, ''A Survey of Models, Analysis Tools and Compensation Methods for the Control of Machines With Friction,'' Automatica, **30**, No. 7, pp. 1083–1138.

[13] Kraus, P. R., Fredriksson, A., and Kumar, V., 1997, ''Modeling of Frictional Contacts for Dynamic Simulation,'' *Proceedings of IROS 1997 Workshop on Dynamic Simulation: Methods and Applications*, Sept.

[14] Hayward, V., and Armstrong, B., 2000, ''A New Computational Model of Friction Applied to Haptic Rendering,'' *Experimental Robotics VI* (Lecture Notes in Control and Information Sciences, Vol. 250), P. Corke and J. Trevelyan, eds., Springer-Verlag, Berlin, pp. 403–412.

[15] Song, P., Kraus, P., Kumar, V., and Dupont, P., 2000, ''Analysis of Rigid Body Dynamic Models for Simulation of Systems With Frictional Contacts,'' Technical Report MS-CIS-00-08, Department of Computer and Information Science, University of Pennsylvania, available at http://www.cis.upenn.edu/~techreports.

[16] Cottle, R. W., Pang, J. S., and Stone, R. E., 1992, *The Linear Complementarity Problem*, Academic Press, San Diego, CA.

[17] Johnson, K. L., 1985, *Contact Mechanics*, Cambridge University Press, New York.

[18] Hunt, K. H., and Crossley, F. R. E., 1975, ''Coefficient of Restitution Interpreted as Damping in Vibroimpact,'' ASME J. Appl. Mech., **42**, pp. 440–445.

[19] Khalil, H. K., 1996, *Nonlinear Systems*, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ.

[20] Dupont, P. E., and Yamajako, S. P., 1997, ''Stability of Frictional Contact in Constrained Rigid-Body Dynamics,'' IEEE Trans. Rob. Autom., **13**, No. 2, pp. 230–236.

[21] McClamroch, N. H., 1989, ''A Singular Perturbation Approach to Modeling and Control of Manipulators Constrained by a Stiff Environment,'' *Proceedings of the 28th Conference on Decision and Control*, pp. 2407–2411.

[22] Oden, J. T., and Pires, E. B., 1983, ''Nonlocal and Nonlinear Friction Laws and Variational Principles for Contact Problems in Elasticity,'' ASME J. Appl. Mech., **50**, pp. 67–76.

[23] Karnopp, D., 1985, ''Computer Simulation of Stick-Slip Friction in Mechanical Dynamic Systems,'' ASME J. Dyn. Syst., Meas., Control, **107**, pp. 100–103.

Journal of Applied Mechanics

Brief Notes

A Brief Note is a short paper that presents a specific solution of technical interest in mechanics but which does not necessarily contain new general methods or results. A Brief Note should not exceed 1500 words *or equivalent* (a typical one-column figure or table is equivalent to 250 words; a one line equation to 30 words). Brief Notes will be subject to the usual review procedures prior to publication. After approval such Notes will be published as soon as possible. The Notes should be submitted to the Technical Editor of the JOURNAL OF APPLIED MECHANICS. Discussions on the Brief Notes should be addressed to the Editorial Department, ASME, United Engineering Center, Three Park Avenue, New York, NY 10016-5990, or to the Technical Editor of the JOURNAL OF APPLIED MECHANICS. Discussions on Brief Notes appearing in this issue will be accepted until two months after publication. Readers who need more time to prepare a Discussion should request an extension of the deadline from the Editorial Department.

Correspondence Principle in Viscoelastic Functionally Graded Materials

G. H. Paulino[1]
Mem. ASME
e-mail: paulino@uiuc.edu

Z.-H. Jin
Mem. ASME

Department of Civil and Environmental Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801

This paper presents an extension of the correspondence principle (as applied to homogeneous viscoelastic solids) to nonhomogeneous viscoelastic solids under the assumption that the relaxation (or creep) moduli be separable functions in space and time. A few models for graded viscoelastic materials are presented and discussed. The revisited correspondence principle extends to specific instances of thermoviscoelasticity and fracture of functionally graded materials. [DOI: 10.1115/1.1331286]

1 Introduction

Functionally graded materials (FGMs) are special composites usually made from both ceramics and metals. The ceramic in an FGM offers thermal barrier effects and protects the metal from corrosion and oxidation. The FGM is toughened and strengthened by the metallic composition. *The composition and the volume fraction of the constituents vary gradually, giving a nonuniform microstructure with continuously graded macroproperties*. Various thermomechanical problems of FGMs have been studied, for example, constitutive modeling ([1]), fracture behavior ([2–4]), thermal stresses ([5,6]), strain gradient effects ([7]), plate bending problems ([8]), higher order theory ([9]), and so on. Comprehensive reviews of ongoing FGM research may be found in the article by Hirai [10] and the book by Suresh and Mortensen [11].

One of the primary application areas of FGMs is high-temperature technology. Materials will exhibit creep and stress relaxation behavior at high temperatures. Viscoelasticity offers a basis for the study of phenomenological behavior of creep and stress relaxation. The elastic-viscoelastic correspondence principle (or elastic-viscoelastic analogy) is probably one of the most useful tools in viscoelasticity because the Laplace transform of the viscoelastic solution can be directly obtained from the corresponding elastic solution. In the present work, the correspondence principle is revisited in the context of viscoelastic FGMs.

In this paper, the basic equations of viscoelasticity in FGMs are formulated. The correspondence principle is established for a class of FGMs where the relaxation moduli for shear and dilatation $\mu(\mathbf{x},t)$ and $K(\mathbf{x},t)$ take the forms $\mu(\mathbf{x},t)=\mu_0\tilde{\mu}(\mathbf{x})f(t)$ and $K(\mathbf{x},t)=K_0\tilde{K}(\mathbf{x})g(t)$, respectively, where μ_0 and K_0 are material constants, $\tilde{\mu}(\mathbf{x})$, $\tilde{K}(\mathbf{x})$, $f(t)$, and $g(t)$ are nondimensional functions, and $\mathbf{x}=(x_1,x_2,x_3)$. The correspondence principle states that the Laplace transforms of the nonhomogeneous viscoelastic variables can be obtained from the nonhomogeneous elastic variables by replacing μ_0 and K_0 with $\mu_0 p\overline{f}(p)$ and $K_0 p\overline{g}(p)$, respectively, where $\overline{f}(p)$ and $\overline{g}(p)$ are the Laplace transforms of $f(t)$ and $g(t)$, respectively, and p is the transform variable. The final nonhomogeneous viscoelastic solution is realized by inverting the transformed solution. The above correspondence principle can also be extended to specific instances of thermoviscoelasticity and fracture of FGMs.

2 Basic Equations

The basic equations of quasi-static viscoelasticity of FGMs are the equilibrium equation

$$\sigma_{ij,j}=0, \tag{1}$$

the strain-displacement relationship

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

and the viscoelastic constitutive law

$$s_{ij}=2\int_0^t \mu(\mathbf{x},t-\tau)\frac{de_{ij}}{d\tau}d\tau, \quad \sigma_{kk}=3\int_0^t K(\mathbf{x},t-\tau)\frac{d\epsilon_{kk}}{d\tau}d\tau, \tag{3}$$

in which σ_{ij} are stresses, ϵ_{ij} are strains, s_{ij} and e_{ij} are deviatoric components of stress and strain tensors given by

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, Jan. 18, 2000; final revision, June 14, 2000. Associate Technical Editor: M.-J. Pindera.

$$s_{ij}=\sigma_{ij}-\frac{1}{3}\sigma_{kk}\delta_{ij}, \quad e_{ij}=\epsilon_{ij}-\frac{1}{3}\epsilon_{kk}\delta_{ij}, \tag{4}$$

where u_i are displacements, δ_{ij} is the Kronecker delta, $\mu(\mathbf{x},t)$ and $K(\mathbf{x},t)$ are appropriate relaxation functions, t is time, and the Latin indices have the range 1, 2, 3 with repeated indices implying the summation convention. Note that the relaxation functions also depend on spatial positions, whereas in homogeneous viscoelasticity, they are only functions of time, i.e., $\mu\equiv\mu(t)$ and $K\equiv K(t)$ ([12]).

For a boundary value problem, the boundary conditions are given by

$$\sigma_{ij}n_j=S_i, \quad \text{on } B_\sigma, \tag{5}$$

$$u_i=\Delta_i, \quad \text{on } B_u, \tag{6}$$

where n_j are the components of the unit outward normal to the boundary of the body, S_i are the tractions prescribed on B_σ, and Δ_i are the prescribed displacements on B_u. The parts of the boundary B_σ and B_u are required to remain constant with time.

3 Correspondence Principle

In general, the correspondence principle of homogeneous viscoelasticity may not hold for FGMs. To circumvent this problem, we consider a class of FGMs in which the relaxation functions have the following general form:

$$\begin{aligned}\mu(\mathbf{x},t)&=\mu_0\tilde{\mu}(\mathbf{x})f(t),\\ K(\mathbf{x},t)&=K_0\tilde{K}(\mathbf{x})g(t),\end{aligned} \tag{7}$$

where μ_0 and K_0 are material constants, and $\tilde{\mu}(\mathbf{x})$, $\tilde{K}(\mathbf{x})$, $f(t)$, and $g(t)$ are nondimensional functions. The constitutive law (3) is then reduced to

$$\begin{aligned}s_{ij}&=2\mu_0\tilde{\mu}(\mathbf{x})\int_0^t f(t-\tau)\frac{de_{ij}}{d\tau}d\tau,\\ \sigma_{kk}&=3K_0\tilde{K}(\mathbf{x})\int_0^t g(t-\tau)\frac{d\epsilon_{kk}}{d\tau}d\tau.\end{aligned} \tag{8}$$

By assuming the material initially at rest, the Laplace transforms of the basic Eqs. (1), (2), (8), and the boundary conditions (5) and (6) are obtained as

$$\bar{\sigma}_{ij,j}=0, \tag{9}$$

$$\bar{\epsilon}_{ij}=\frac{1}{2}(\bar{u}_{i,j}+\bar{u}_{j,i}), \tag{10}$$

$$\bar{s}_{ij}=2\mu_0\tilde{\mu}(\mathbf{x})p\bar{f}(p)\bar{e}_{ij}, \tag{11}$$

$$\bar{\sigma}_{kk}=3K_0\tilde{K}(\mathbf{x})p\bar{g}(p)\bar{\epsilon}_{kk}, \tag{12}$$

$$\bar{\sigma}_{ij}n_j=\bar{S}_i, \quad \text{on } B_\sigma, \tag{13}$$

$$\bar{u}_i=\bar{\Delta}_i, \quad \text{on } B_u, \tag{14}$$

where a bar over a variable represents its Laplace transform, and p is the transform variable. Thus

$$\begin{aligned}&\bar{\sigma}_{ij}=\int_0^\infty \sigma_{ij}\exp(-pt)dt, \quad \bar{\epsilon}_{ij}=\int_0^\infty \epsilon_{ij}\exp(-pt)dt,\\ &\bar{u}_i=\int_0^\infty u_i\exp(-pt)dt, \quad \bar{f}(p)=\int_0^\infty f(t)\exp(-pt)dt,\\ &\bar{g}(p)=\int_0^\infty g(t)\exp(-pt)dt.\end{aligned} \tag{15}$$

It is seen that the set of Eqs. (9)–(12), and conditions (13) and (14) have a form identical to those of nonhomogeneous elasticity with the shear modulus $\mu=\mu_0\tilde{\mu}(\mathbf{x})$ and the bulk modulus $K=K_0\tilde{K}(\mathbf{x})$ provided that the transformed viscoelastic variables are associated with the corresponding elastic variables and $\mu_0 p\bar{f}(p)$ and $K_0 p\bar{g}(p)$ are associated with μ_0 and K_0, respectively. Therefore, the *correspondence principle* in homogeneous viscoelasticity still holds for the FGM with the material properties given in Eq. (7), i.e., *the Laplace transformed nonhomogeneous viscoelastic solution can be obtained directly from the solution of the corresponding nonhomogeneous elastic problem by replacing μ_0 and K_0 with $\mu_0 p\bar{f}(p)$ and $K_0 p\bar{g}(p)$, respectively. The final solution is realized upon inverting the transformed solution.*

4 Some Models for Graded Viscoelastic Materials

Among the various models for graded viscoelastic materials are the *standard linear solid* defined by

$$\begin{aligned}\mu(\mathbf{x},t)&=\mu_\infty(\mathbf{x})+[\mu_e(\mathbf{x})-\mu_\infty(\mathbf{x})]\exp\left[-\frac{t}{t_\mu(\mathbf{x})}\right],\\ K(\mathbf{x},t)&=K_\infty(\mathbf{x})+[K_e(\mathbf{x})-K_\infty(\mathbf{x})]\exp\left[-\frac{t}{t_K(\mathbf{x})}\right],\end{aligned} \tag{16}$$

the *power-law model*

$$\mu(\mathbf{x},t)=\mu_e(\mathbf{x})\left[\frac{t_\mu(\mathbf{x})}{t}\right]^q, \quad K(\mathbf{x},t)=K_e(\mathbf{x})\left[\frac{t_K(\mathbf{x})}{t}\right]^q, \quad 0<q<1, \tag{17}$$

and the *Maxwell material*

$$\mu(\mathbf{x},t)=\mu_e(\mathbf{x})\exp\left[-\frac{t}{t_\mu(\mathbf{x})}\right], \quad K(\mathbf{x},t)=K_e(\mathbf{x})\exp\left[-\frac{t}{t_K(\mathbf{x})}\right], \tag{18}$$

where $t_\mu(\mathbf{x})$ and $t_K(\mathbf{x})$ are the relaxation times in shear and bulk moduli, respectively, and q is a material constant. The discussion below indicates the revisions needed in the general models so that the correspondence principle holds.

- Standard Linear Solid (16). If the relaxation times t_μ and t_K are constant, if $\mu_e(\mathbf{x})$ and $\mu_\infty(\mathbf{x})$ have the same functional form, and if $K_e(\mathbf{x})$ and $K_\infty(\mathbf{x})$ have the same functional form, then the standard linear solid satisfies assumption (7).
- Power Law Model (17). It is seen that if the relaxation times t_μ and t_K are independent of spatial position, then the assumption (7) is readily satisfied. Moreover, even if the relaxation times depend on the spatial position in (17), the correspondence principle may still be applied with some revision, which consists of taking the corresponding nonhomogeneous elastic material with the following properties:

$$\mu=\mu_e(\mathbf{x})[t_\mu(\mathbf{x})]^q, \quad K=K_e(\mathbf{x})[t_K(\mathbf{x})]^q, \tag{19}$$

instead of $\mu=\mu_e(\mathbf{x})$ and $K=K_e(\mathbf{x})$.
- Maxwell Material (18). If the relaxation times t_μ and t_K are independent of spatial position, the assumption (7) is promptly satisfied.

5 Thermoviscoelastic Problem

The basic equations of thermoviscoelasticity of FGMs are identical to those of viscoelasticity except the constitutive law. The constitutive relation for thermoviscoelastic FGMs is given by

$$\begin{aligned}s_{ij}&=2\int_0^t \mu(\mathbf{x},t-\tau)\frac{de_{ij}}{d\tau}d\tau,\\ \sigma_{kk}&=3\int_0^t K(\mathbf{x},t-\tau)\frac{d[\epsilon_{kk}-\alpha(\mathbf{x})T]}{d\tau}d\tau,\end{aligned} \tag{20}$$

where T is the temperature and $\alpha(\mathbf{x})$ is the coefficient of thermal expansion. Here α is assumed to be time-independent. By apply-

ing the Laplace transform to the above equation and adopting the form of the relaxation functions given in (7), we obtain

$$\bar{s}_{ij}=2\mu_0\tilde{\mu}(\mathbf{x})p\bar{f}(p)\bar{e}_{ij}, \quad \bar{\sigma}_{kk}=3K_0\tilde{K}(\mathbf{x})p\bar{g}(p)(\bar{\epsilon}_{kk}-\alpha\bar{T}), \tag{21}$$

while the constitutive relation of the nonhomogeneous thermoelasticity may be expressed as

$$s_{ij}=2\mu_0\tilde{\mu}(\mathbf{x})e_{ij}, \quad \sigma_{kk}=3K_0\tilde{K}(\mathbf{x})(\epsilon_{kk}-\alpha T). \tag{22}$$

Thus it can be seen that the correspondence principle still holds.

6 A Path-Independent Integral

The *J*-integral ([13]) has been extended to certain classes of elastic materials with varying Young's modulus in the crack-line direction by Honein and Herrmann [14]. Here, a *J*-like path-independent integral is presented for characterizing fracture in nonhomogeneous viscous materials.

Consider the shear modulus with the specific functional form

$$\mu(x_1,x_2,t)=\mu_0(x_2)\exp(\beta x_1)f(t) \tag{23}$$

where $\mu_0(x_2)$ is an arbitrary function of x_2 and β is an arbitrary material constant. Note that (23) has the form given in (7). Moreover, the Poisson's ratio is assumed to be independent of x_1. The proposed integral to characterize crack growth in such graded material undergoing creep is

$$C_e^*=\int_\Gamma\left[\left(\dot{W}n_1-\sigma_{ij}n_j\frac{\partial\dot{u}_i}{\partial x_1}\right)-\frac{\beta}{2}\sigma_{ij}n_j\dot{u}_i\right]ds \tag{24}$$

where Γ is a contour enclosing the crack tip, n_1 is the first component of the unit outward normal to Γ, $\sigma_{ij}n_j=S_i$ are the components of tractions along Γ, ds is an infinitesimal length element along the contour Γ, and $\dot{W}$ is the stress work rate (power) density defined as

$$\dot{W}=\int_0^{\dot{\epsilon}_{kl}}\sigma_{ij}d\dot{\epsilon}_{ij}. \tag{25}$$

The integral (24) has been obtained by replacing strain with strain rates, and displacement with displacement rates in the corresponding J_e-integral ([14]) for nonhomogeneous elastic materials.

The integral of the term within parentheses in (24) is the so-called C^* integral (e.g., [15]) which is valid for homogeneous viscous materials undergoing steady-state creep. The extra term in (24), which appears outside the parentheses, is due to the modulus variation. Equation (24) can be seen as an extension of the C^* integral for nonhomogeneous viscous media. The C^* integral is a special case of the J_v-integral derived by Schapery [16] by means of correspondence principle arguments. The latter integral accounts for a wide range of time-dependent material behavior, and includes viscous creep as special case.

7 A Simple Example

As an example of application, we consider an infinite strip of width h occupying the region $0\leqslant x_1\leqslant h$, $-\infty<x_2<\infty$, $-\infty<x_3<\infty$. It is assumed that the strip deforms in the x_1-x_2 plane under the plane-strain conditions. A ''fixed grip'' loading condition is considered, i.e., $\epsilon_{22}(x_1,\pm\infty)=\epsilon_0$, where ϵ_0 is a constant. The nonvanishing stress σ_{22} in a nonhomogeneous elastic material with the Young's modulus $E=E_e(x_1)$ and the Poisson's ratio $\nu=\nu_e(x_1)$ is given by ([4])

$$\sigma_{22}=\frac{E_e(x_1)\epsilon_0}{1-\nu_e^2(x_1)}=\frac{4\epsilon_0\mu_e(x_1)[3K_e(x_1)+\mu_e(x_1)]}{3K_e(x_1)+4\mu_e(x_1)}, \tag{26}$$

where the following relations are used:

$$E_e=\frac{9K_e\mu_e}{3K_e+\mu_e}, \quad \nu_e=\frac{3K_e-2\mu_e}{2(3K_e+\mu_e)}. \tag{27}$$

According to the correspondence principle, the Laplace transform of the stress in a viscoelastic FGM with the shearing and dilatational relaxation functions $\mu=\mu_e(x_1)f(t)$ and $K=K_e(x_1)g(t)$ is given by

$$\bar{\sigma}_{22}=\frac{4\epsilon_0\mu_e(x_1)\bar{f}(p)[3K_e(x_1)\bar{g}(p)+\mu_e(x_1)\bar{f}(p)]}{3K_e(x_1)\bar{g}(p)+4\mu_e(x_1)\bar{f}(p)}. \tag{28}$$

For the Maxwell material (18) with constant relaxation times t_μ and t_K, the above transformed stress becomes

$$\bar{\sigma}_{22}=\frac{[4\epsilon_0\mu_e(x_1)/(p+1/t_\mu)][3K_e(x_1)/(p+1/t_K)+\mu_e(x_1)/(p+1/t_\mu)]}{3K_e(x_1)/(p+1/t_K)+4\mu_e(x_1)/(p+1/t_\mu)}. \tag{29}$$

By inverting (29), we get the stress in the time domain as follows:

$$\sigma_{22}=\left\{\frac{9K_e(x_1)}{4\mu_e(x_1)+3K_e(x_1)}\exp\left[-\frac{4\mu_e(x_1)t_\mu/t_K+3K_e(x_1)}{4\mu_2(x_1)+3K_e(x_1)}\frac{t}{t_\mu}\right]+\exp\left(-\frac{t}{t_\mu}\right)\right\}\mu_e(x_1)\epsilon_0. \tag{30}$$

By letting $t\to0^+$, the nonhomogeneous elastic solution (26) is recovered.

8 Conclusions

The correspondence principle is revisited and established for a class of FGMs where the relaxation functions for shear and dilatation take separable forms in space and time, i.e., $G_1(\mathbf{x},t)/2=\mu(\mathbf{x},t)=\mu_0\tilde{\mu}(\mathbf{x})f(t)$ and $G_2(\mathbf{x},t)/3=K(\mathbf{x},t)=K_0\tilde{K}(\mathbf{x})g(t)$, respectively. The correspondence principle states that the Laplace transforms of the nonhomogeneous viscoelastic variables can be obtained from the nonhomogeneous elastic variables by replacing μ_0 and K_0 with $\mu_0p\bar{f}(p)$ and $K_0p\bar{g}(p)$, respectively, where $\bar{f}(p)$ and $\bar{g}(p)$ are the Laplace transforms of $f(t)$ and $g(t)$, respectively, and p is the transform variable. The final nonhomogeneous viscoelastic solution is realized by inverting the transformed solution. Equivalently, if the creep functions $J_1(\mathbf{x},t)$ and $J_2(\mathbf{x},t)$ have separable forms in space and time, then the correspondence principle (as employed here) is also directly applicable.

Acknowledgments

We thank two anonymous reviewers for insightful comments. We also would like to acknowledge the support from the National Science Foundation (NSF) under grant No. CMS-9996378 (Mechanics & Materials Program).

References

[1] Reiter, T., Dvorak, G. J., and Tvergaard, V., 1997, ''Micromechanical Models for Graded Composite Materials,'' J. Mech. Phys. Solids, **45**, pp. 1281–1302.

[2] Cai, H., and Bao, G., 1998, ''Crack Bridging in Functionally Graded Coatings,'' Int. J. Solids Struct., **35**, pp. 701–717.

[3] Erdogan, F., 1995, ''Fracture Mechanics of Functionally Graded Materials,'' Composites Eng., **5**, pp. 753–770.

[4] Jin, Z.-H., and Batra, R. C., 1998, ''R-Curve and Strength Behavior of a Functionally Graded Material,'' Mater. Sci. Eng., A, **242**, pp. 70–76.
[5] Kawasaki, A., and Watanabe, R., 1987, ''Finite Element Analysis of Thermal Stress of the Metals/Ceramics Multi-Layer Composites with Controlled Compositional Gradients,'' J. Jpn. Inst. Met., **51**, pp. 525–529.
[6] Noda, N., 1999, ''Thermal Stresses in Functionally Graded Materials,'' J. Therm. Stresses, **22**, pp. 477–512.
[7] Paulino, G. H., Fannjiang, A. C., and Chan, Y. S., 1999, ''Gradient Elasticity Theory for a Mode III Crack in a Functionally Graded Material,'' Mater. Sci. Forum, **308–311**, pp. 971–976.
[8] Reddy, J. N., and Chin, C. D., 1998, ''Thermomechanical Analysis of Functionally Graded Cylinders and Plates,'' J. Therm. Stresses, **21**, pp. 593–626.
[9] Aboudi, J., Pindera, M. J., and Arnold, S. M., 1999, ''Higher-Order Theory for Functionally Graded Materials,'' Composites, Part B, **30B**, pp. 777–832.
[10] Hirai, T., 1996, ''Functionally Gradient Materials,'' *Processing of Ceramics, Part 2* (Materials Science and Technology, Vol. 17B), R. J. Brook, ed., VCH Verlagsgesellschaft mbH, Weinheim, Germany, pp. 292–341.
[11] Suresh, S., and Mortensen, A., 1998, *Fundamentals of Functionally Graded Materials*, The Institute of Materials, IOM Communications Ltd., London.
[12] Christensen, R. M., 1971, *Theory of Viscoelasticity*, Academic Press, New York.
[13] Rice, J. R., 1968, ''A Path Independent Integral and the Approximate Analysis of Strain Concentration by Notches and Cracks,'' ASME J. Appl. Mech., **35**, pp. 379–386.
[14] Honein, T., and Herrmann, G., 1997, ''Conservation Laws in Nonhomogeneous Plane Elastostatics,'' J. Mech. Phys. Solids, **45**, pp. 789–805.
[15] Landes, J. D., and Begley, J. A., 1976, ''A Fracture Mechanics Approach to Creep Crack Growth,'' *ASTM STP 590*, American Society for Testing and Materials, Philadelphia, PA, pp. 128–148.
[16] Schapery, R. A., 1984, ''Correspondence Principles and a Generalized J Integral for Large Deformation and Fracture Analysis of Viscoelastic Media,'' Int. J. Fract., **25**, pp. 195–223.

On Some Anomalies in Lamé's Solutions for Elastic Solids With Holes

G. B. Sinclair
Department of Mechanical Engineering, Louisiana State University, Baton Rouge, LA 70803-6413

G. Meda
Science and Technology Division, Corning, Inc., Corning, NY 14831-0001

Elastic solids with holes under remote tension are reconsidered. When hole dimensions are shrunk so that holes disappear, anomalies occur in the classical elasticity solutions of Lamé. By introducing cohesive laws on hole surfaces as they shrink, these anomalies may be removed. [DOI: 10.1115/1.1331285]

1 The Issue

Sketched in Fig. 1 is the Lamé problem of an infinite elastic plate, weakened by a circular hole of radius a, under a uniform remote tension σ_0. In cylindrical polar coordinates (Fig. 1), the stresses in its classical solution are given in Lamé [1] and are

$$\begin{Bmatrix} \sigma_r \\ \sigma_\theta \end{Bmatrix} = \sigma_0 \left(1 \begin{Bmatrix} - \\ + \end{Bmatrix} \frac{a^2}{r^2} \right), \tag{1}$$

for $a \leqslant r < \infty$, $0 \leqslant \theta < 2\pi$. The companion shear stress component is zero by virtue of the axisymmetry of the configuration. That such a stress field is indeed a valid solution within classical elasticity can be verified by direct substitution into the governing field equations and the boundary conditions.

Setting $r=a$ in σ_θ of (1) reveals a stress concentration factor (SCF) of 2 at the edge of the hole. Consider what happens to this concentration factor if $a \to 0$ and the hole disappears. The SCF is independent of a, so it remains equal to 2 even when $a \to 0$. This is inconsistent with what one would expect physically, namely that the limit $a \to 0$ should be the same as when the plate is whole without a hole and has no stress concentration.

The same sort of anomalous result occurs for an elastic solid with a spherical hole. Then Lamé [1] has that the SCF is 3/2 independent of the hole radius. Again, therefore, there is a stress concentration when the radius goes to zero, inconsistent with physical expectations.

These anomalous results are passed by without comment in Lamé [1]. While they have no doubt been noted by elasticans since, their existence may well not be as widely appreciated today as it could be. They are not mentioned in classical texts which include the Lamé solutions (e.g., Love [2], Art. 100, 98; Muskhelishvili [3], Art. 56a; Timoshenko and Goodier [4], Art. 28, 136). Further, we could not find them discussed in any other standard elasticity texts.[1] Nevertheless they bear explaining.

Mathematically, there is a clear distinction between solids with holes with radii tending to zero and solids without holes. When $a \to 0$ in either of Lamé's hole problems, the boundary condition $\sigma_r = 0$ holds at $r=0$. In contrast, for a plate without a hole, the field equations hold at $r=0$. Such mathematical distinctions, however, fall short of a fully satisfactory physical explanation.

We have been offered physical explanations of the following genre by a number of people: ''Physically speaking, one explains this nonuniform behavior as the presence of a stress concentration in an imperfect body such as at the boundary of a small entrained cavity in a casting.'' To examine the physical appropriateness of such explanations, we consider a further limit of (1) as $a \to 0$. Specifically, we take σ_θ of (1) at $\theta = 0$, denote it by $\hat{\sigma}_y$, and set $r = \lambda a$, $\lambda \geqslant 1$. Then

$$\hat{\sigma}_y = \sigma_0 (1 + \lambda^{-2}) \quad \text{as} \quad a \to 0. \tag{2}$$

Of course, as $a \to 0$, $r = \lambda a \to 0$ for all $\lambda \geqslant 1$. Hence with this model of an imperfection, as a tends to zero we can get any value

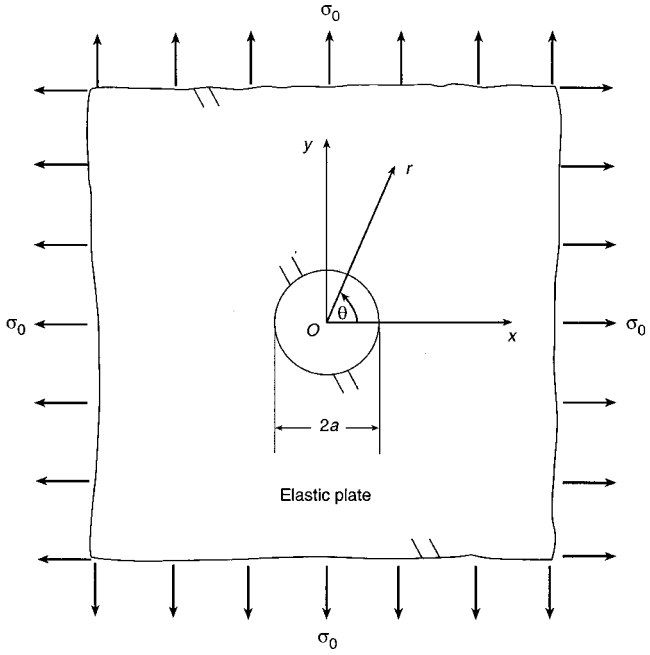


Fig. 1 Plate with hole under remote tension

[1] Swain [5], pp. 121,122, does note a similar anomalous result in the classical elasticity solution for an infinite plate with a circular hold under *uniaxial* far-field tension.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received and accepted by the ASME Applied Mechanics Division, Mar. 27, 2000; final revision, Aug. 8, 2000. Associate Technical Editor: J. R. Barber.

of $\hat{\sigma}_y$ at $r=0$ between σ_0 and $2\sigma_0$ as the imperfection's stress concentration.[2] This unsatisfactory situation is compounded by the ambiguity of which stress component σ_x or σ_y is what in the limit as $a\to 0$ for different θ. All told, such physical explanations are quite superficial. Here, then, we seek to furnish a physically sensible resolution of the differences between Lamé's solutions for plates with holes and responses for whole plates.

2 A Resolution

What is missing in the classical statement of Lamé's hole problems is the recognition that atoms or molecules on opposite sides of any hole must start to interact with each other as the hole closes. This interaction produces *cohesive stresses* on the hole walls. Here we model the action of these cohesive stresses.

There are three key elements in our simple models. First, we introduce cohesive stresses via cohesive stress-separation laws on hole boundaries. This simplifies the incorporation of the underlying solid-state physics and reduces the analysis of our models to involving just continuum mechanics. Such an approach was first introduced in Barenblatt [6] and has seen extensive use since (Sinclair [7] provides a recent bibliography). For the most part, it has been employed in the analysis of cracks, although Levy [8,9] treats a rigid inclusion without a crack. The implementation of cohesive stress-separation laws here could be viewed as the dual of their use in Levy [8,9].

Second, we only consider that portion of the cohesive stress-separation law near the equilibrium position. That is, we only track the action of cohesive stresses when the hole is extremely small. In this range, cohesive stress-separation laws can be taken as linear. Moreover, the constant of proportionality can be backed out by insisting that the insertion of such a cohesive law within the continuum without any hole leaves response there unaltered—a kind of cohesive-law patch test. For the present problem, this insertion is actually carried out on a circular ring of radius R in an elastic plate with the same moduli as the original plate. Then such a patch test in effect accounts for the action of all the atoms external to R on all those internal, and vice versa. Again, simplification is the intent. The so-simplified treatment does nonetheless serve to demonstrate the basic physics involved.[3]

Third, we take our cohesive stress-separation law as acting between the centers of the atoms or molecules comprising the hole surfaces: By symmetry, these atoms or molecules are diametrically opposed. The consequence of this assumption is that holes close when their radii reduce to half of the equilibrium center-to-center spacing of the atoms or molecules. This removes any ambiguity associated with $a\to 0$.

The corresponding reformulation of Lamé's problem for the plate with a hole then is as follows. Throughout the plate of Fig. 1 when a is small, we seek the axisymmetric planar stresses σ_r, σ_θ, and their companion displacement u_r, as functions of r, satisfying the following requirements: the stress equation of equilibrium in the absence of body forces,

$$r\sigma_{r,r}+\sigma_r-\sigma_\theta=0, \tag{3}$$

for $a<r<\infty$, $0\leqslant\theta<2\pi$; the stress-displacement relations for a homogeneous and isotropic, linear elastic solid,

$$\begin{Bmatrix}\sigma_r\\ \sigma_\theta\end{Bmatrix}=\mu\left[\frac{3-\kappa}{\kappa-1}\Theta+2\begin{Bmatrix}u_{r,r}\\ r^{-1}u_r\end{Bmatrix}\right],\quad \Theta=u_{r,r}+r^{-1}u_r, \tag{4}$$

for $a<r<\infty$, $0\leqslant\theta<2\pi$, wherein Θ is the dilation, μ is the shear modulus and κ is $3-4\nu$ for plane strain, $(3-\nu)/(1+\nu)$ for plane stress, ν being Poisson's ratio; the cohesive stress-separation law on the hole boundary,

$$\sigma_r=k(2u_r+2a-\delta)\ \text{ at }\ r=a, \tag{5}$$

for $0\leqslant\theta<2\pi$, wherein k is the law stiffness and δ is the equilibrium separation of the atoms or molecules comprising the plate; and the condition applying the tension at infinity,

$$\sigma_r=\sigma_0\ \text{ as }\ r\to\infty, \tag{6}$$

for $0\leqslant\theta<2\pi$. In addition, from our cohesive-law patch test at $r=R$, we have $\sigma_r=k[u_r(r=R+\delta/2)-u_r(r=R-\delta/2)]$, leading to

$$k=4\mu/\delta(\kappa-1). \tag{7}$$

This is the value of the stiffness to be used in (5) when a is sufficiently small.

Solution of the problem in (3)–(6) is elementary and gives

$$\begin{Bmatrix}\sigma_r\\ \sigma_\theta\end{Bmatrix}=\sigma_0\begin{Bmatrix}-\\ +\end{Bmatrix}\sigma_0'\frac{a^2}{r^2},\quad u_r=\frac{1}{4\mu}\left[\sigma_0 r(\kappa-1)+2\sigma_0'\frac{a^2}{r}\right], \tag{8}$$

where

$$\sigma_0'=\sigma_0-k\,\frac{2\mu(2a-\delta)+(\kappa+1)\sigma_0 a}{2(\mu+ka)}. \tag{9}$$

Observe that (8) and (9) recover Lamé's solution (1) when $k=0$, as they should.

Now consider what happens if the hole disappears. Introducing k of (7) into (9), and taking $a\to\delta/2$ to close the hole, gives $\sigma_0'=0$. Thus from (8),

$$\sigma_r=\sigma_\theta=\sigma_0\ \text{ as }\ a\to\delta/2. \tag{10}$$

Equation (10) is the physically sensible result for a plate without a hole.

A similar reformulation and analysis for the spherical hole problem leads to

$$\begin{Bmatrix}\sigma_r\\ \sigma_\theta\end{Bmatrix}=\sigma_0\begin{Bmatrix}-\\ +1/2\end{Bmatrix}\sigma_0''\frac{a^3}{r^3}, \tag{11}$$

where

$$\sigma_0''=\sigma_0-k\,\frac{4\mu(2a-\delta)+3(\kappa-1)\sigma_0 a}{2(2\mu+ka)}, \tag{12}$$

with κ being as for plane stress. Again Lamé's solution is recovered when $k=0$, and a state of uniform all-round tension obtains when $a\to\delta/2$ provided k is taken so that it passes the cohesive-law patch test in spherical polar coordinates ($k=8\mu/\delta(3\kappa-5)$).

Implicit in both the circular and spherical hole problems treated here is the existence of a length scale which is considerably larger than the initial radii, and which remains fixed as radii go to zero. This additional length scale can be made explicit by instead considering an annular plate and a hollow ball. The same anomalies result when internal holes are shrunk to zero: They can be remedied by a parallel introduction of cohesive laws.

It is also possible to adapt the foregoing if one actually wanted to model an imperfection. Then the fact that material on opposite sides of the holes had once been separated can be reflected in the choice of the cohesive law as material gets back together if indeed there is some impediment which modifies this law. To be truly physically appropriate, this choice needs to be founded in solid-state physics. Such an analysis is beyond the scope of the present note.

In sum, the boundary conditions in Lamé's classical solutions for elastic solids with holes are not physically appropriate when

[2]If instead r is not fixed in terms of a before taking the limit $a\to 0$, then a state of all-round tension obtains (see (1)). This is a different limit, however, since under it one is moving to infinity rather than to the center of the hole.

[3]Insertion of an entire, nonlinear, cohesive, stress-separation law is tractable within linear elasticity because the present problems are one-dimensional. It is not appropriate, though, because the large strains incurred near the peak stresses in cohesive laws really require a finite strain analysis.

hole surfaces come into extremely close proximity with one another. Cohesive stresses act under these circumstances. Without such proximities though, classical solutions are applicable.

References

[1] Lamé, M. G., 1852, *Leçons sur la Théorie Mathématique de l'Élasticité des Corps Solids*, Bachelier, Paris.

[2] Love, A. E. H., 1944, *A Treatise on the Mathematical Theory of Elasticity*, Dover, New York.

[3] Muskhelishvili, N. I., 1963, *Some Basic Problems of the Mathematical Theory of Elasticity*, Noordhoff, Groningen, The Netherlands.

[4] Timoshenko, S. P., and Goodier, J. N., 1970, *Theory of Elasticity*, McGraw-Hill, New York.

[5] Swain, G. F., 1924, *Structural Engineering—Strength of Materials*, McGraw-Hill, New York.

[6] Barenblatt, G. I., 1959, ''On the Equilibrium of Cracks Due to Brittle Fracture,'' Dokl. Akad. Nauk, **127**, pp. 47–50.

[7] Sinclair, G. B., 1999, ''A Bibliography on the Use of Cohesive Laws in Solid Mechanics,'' Report SM99-8, Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA.

[8] Levy, A. J., 1991, ''The Debonding of Elastic Inclusions and Inhomogeneities,'' J. Mech. Phys. Solids, **39**, pp. 477–505.

[9] Levy, A. J., 1994, ''Separation of a Circular Interface Under Biaxial Load,'' J. Mech. Phys. Solids, **42**, pp. 1087–1104.

Journal of Applied Mechanics

Book Review

Thermal Stresses, by Naotake Noda, Richard B. Hetnarski, and Yoshinobu Tanigawa. Lastran Corporation, Rochester, NY, 2000. 455 pages. Price: $70.00.

REVIEWED BY J. R. BARBER[1]

There are several excellent books on the subject of thermal stresses, including the classical texts by Boley and Weiner, Nowacki, and Nowinski, but none of them is entirely satisfactory for classroom use because they make few concessions to the reader and in particular do not include worked examples and end of chapter problems. *Thermal Stresses* by Noda, Hetnarski, and Tanigawa aims to fill this gap. It starts from the most elementary concepts of one-dimensional thermal expansion and stress, which most students will have encountered in a first course in Mechanics of Materials. Problems and methods of gradually increasing complexity are then introduced, each being illustrated by several text examples and supported by suitable student assignments.

The first two chapters cover linear thermoelasticity of one-dimensional bars and Euler beams, respectively, with various temperature distributions and mechanical loading conditions. The fundamental concepts are reinforced by the discussion of a wide range of applications including composite and inhomogeneous beams. Although the concepts involved in these applications are straightforward, the solutions can involve rather formidable looking algebra, which may lead the less insightful student to overlook the essential simplicity of the underlying concepts. I was also disappointed to see no reference to elastic-plastic problems in this section, since thermally induced residual stresses are an important branch of the subject and these one-dimensional examples provide a convenient vehicle for their introduction to the student at an elementary level.

Chapter 3 introduces the heat conduction equation and thermal boundary conditions. Problems in one dimension are solved first by separated variable methods, leading to series solutions, and then by Laplace transform methods. This methodology is first introduced in Cartesian coordinates and then extended to one-dimensional problems in cylindrical and spherical coordinates.

Chapter 4 presents an overview of the complete development of the general equations of linear thermoelasticity, starting with equilibrium and coordinate transformation of stress components, the definition of strain components, compatibility, and statements of the governing equations in terms of stress or displacement. The particular solution of these equations is defined in terms of the thermoelastic displacement potential, and the Papkovitch-Neuber solution is introduced as the homogeneous solution, using separated variable solutions of the Laplace equation in Cartesian coordinates. Appropriate results are then repeated in the cylindrical and spherical coordinate systems. The chapter ends with a discussion of the importance of multiply connected bodies in thermoelastic problems and a derivation of the Cesaro line integrals. My impression is that this chapter will appear very indigestible to any student without a previous grounding in linear elasticity. Apart from a single worked example, the derivations are presented without a break and without much indication of how the material will be used. Of course, these equations will later be reduced to simpler forms for specific examples and it would obviously be inefficient to derive them separately in each case. However, the instructor will probably find it necessary to supplement the material in this chapter with more examples, if only to allow the students time to catch their collective breath during the ascent.

Similar criticisms could be leveled at later chapters in the book, where the general results are specialized to two-dimensional (plane) problems (Chapter 5), problems for the cylinder (Chapter 6) and the sphere (Chapter 7), and for thin plates (Chapter 8). Most of these chapters contain only one or two text examples and these are concentrated near the beginning of the chapters. It is as though the authors gradually forgot their stated mission in the middle of each chapter, moving to a ''monograph'' style of writing. It is true that many of the later derivations in each chapter present essentially a general solution for a class of problems (for example, a built-in rectangular plate with a prescribed temperature distribution), so that the solution of a particular technical problem would merely involve the substitution of one or more given functions and the evaluation of some integrals. However, in my experience, these particular examples provide crucial motivation to students, particularly those with a more practical engineering perspective. They also provide an opportunity to the author and the instructor to draw important technical conclusions. For example, how big a temperature variation in the body might be sufficient to generate dangerously large thermal stresses in various practical engineering components? Where does the maximum stress occur and what is the physical character of the deformed shape of the body?

Chapter 9 generalizes the theory of beam-columns to include thermal effects and the final Chapter 10 discusses thermoelasticity from the perspective of thermodynamics. In particular, the coupled heat conduction equation is derived from thermodynamic considerations and the student is also introduced to variational theorems, the uniqueness theorem and the reciprocal theorem. Overall, the technical and mathematical level of the material makes this book most suitable for a first year graduate level course, though the more mathematically talented senior undergraduates may also be able to assimilate it with appropriate guidance.

Engineering designers routinely make use of ''everyday intuition'' to identify structurally sound methods of transmitting purely mechanical loads, but the effects of thermal loading are much harder to predict without detailed analysis. Also, what appear to be relatively modest temperature variations can cause serious states of stress. There is therefore an excellent case for a greater representation of thermal stresses in engineering degree curricula and the authors are to be complimented on providing us with a suitable text for this purpose.

[1]Professor of Mechanical Engineering and Applied Mechanics, University of Michigan, Ann Arbor, MI 48109-2125. Mem. ASME

