Transactions of the ASME®

Technical Editor, **LEWIS T. WHEELER**

Department of Mechanical Engineering,
University of Houston,
Houston, TX 77204-4792

APPLIED MECHANICS DIVISION

Chairman, **C. T. HERAKOVICH**
Secretary, **T. HUGHES**
Associate Technical Editors,
J. R. BARBER (2000)
J. L. BASSANI (1999)
R. C. BENSON (2000)
M. M. CARROLL (2000)
I. M. DANIEL (1999)
W. J. DRUGAN (2000)
A. A. FERRI (2000)
J. T. JENKINS (1999)
J. W. JU (2001)
V. K. KINRA (1999)
S. KYRIAKIDES (2000)
W. K. LIU (1999)
A. K. MAL (2001)
A. NEEDLEMAN (2001)
M. ORTIZ (2001)
N. C. PERKINS (1999)
M.-J. PINDERA (2000)
K. T. RAMESH (2000)
D. A. SIGINER (2000)
C. G. SPEZIALE (2001)
M. TAYA (1999)

BOARD ON COMMUNICATIONS

Chairman and Vice-President
R. K. SHAH

OFFICERS OF THE ASME
President, **W. M. PHILLIPS**
Executive Director, **D. L. BELDEN**
Treasurer, **J. A. MASON**

PUBLISHING STAFF
Managing Director, Engineering
CHARLES W. BEARDSLEY
Director, Technical Publishing
PHILIP DI VIETRO
Managing Editor, Technical Publishing
CYNTHIA B. CLARK
Managing Editor, Transactions
CORNELIA MONAHAN
Production Coordinator
JUDITH SIERANT
Production Assistant
MARISOL ANDINO

Transactions of the ASME, Journal of Applied Mechanics (ISSN 0021-8936) is published quarterly (Mar., June, Sept., Dec.) for $240.00 per year by 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. **PRICES:** To members, $40.00; to nonmembers, $240.00; single issue, $96.00. Add $40.00 (airmail $60.00) for postage to countries outside the United States and Canada.

STATEMENT from By-Laws. The Society shall not be responsible for statements or opinions advanced in papers or . . . printed in its publications (B7.1, Para. 3), **COPYRIGHT © 1998** by The American Society of Mechanical Engineers. Authorization to photocopy material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by ASME to libraries and other users registered with the Copyright Clearance Center (CCC). Transactional Reporting Service provided that the base fee of $3.00 per article is paid directly to CCC, Inc., 222 Rosewood Drive, Danvers, MA 01923. Request for special permission or bulk copying should be addressed to Reprints/Permission Department.

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

Journal of Applied Mechanics

Published Quarterly by The American Society of Mechanical Engineers

VOLUME 65 • NUMBER 4 • DECEMBER 1998

TECHNICAL PAPERS

797 **Frequency Clusters in the Spectrum of Annular Cylinders**
K. I. Tzou, J. A. Wickert, and A. Akay

804 **Transient Discharge of Entrained Air From a Wound Roll**
M. B. Keshavan and J. A. Wickert

811 **An Iterative Approach to Dynamic Elastic-Plastic Analysis**
F. Giambanco, L. Palizzolo, and L. Cirone

820 **Layer-Wise Mixed Models for Accurate Vibrations Analysis of Multilayered Plates**
E. Carrera

829 **Stress Singularity Analysis of a Crack Terminating at the Interface of an Anisotropic Layered Composite**
P. Poonsawat, A. C. Wijeyewickrema, and P. Karasudhi

837 **Study of Residual Stress Distribution by a Combined Method of Moiré Interferometry and Incremental Hole Drilling, Part I: Theory**
Zhu Wu, Jian Lu, and Bongtae Han

844 **Study of Residual Stress Distribution by a Combined Method of Moiré Interferometry and Incremental Hole Drilling, Part II: Implementation**
Zhu Wu, Jian Lu, and Bongtae Han

851 **Distortion of a Harmonic Elastic Wave Reflected From a Dry Friction Support**
M. J. Leamy, J. R. Barber, and N. C. Perkins

858 **On the Stress Field in Sliding Ellipsoidal Inclusions With Shear Eigenstrain**
V. A. Lubarda and X. Markenscoff

863 **Inclusions of Uniform Eigenstrains and Constant or Other Stress Dependence**
X. Markenscoff

867 **A Nonuniform Cell Partition for the Analysis of Nonlinear Stochastic Systems**
J. Q. Sun

870 **Stiff Elastic Tape Wrapped Onto a Drum**
R. C. Benson

875 **Invariants of Chaotic Attractor in a Nonlinearly Damped System**
B. Ravindra and P. Hagedorn

880 **Solution of the Displacement Boundary Value Problem of an Interface Between Two Dissimilar Half-Planes and a Rigid Elliptic Inclusion at the Interface**
V. Boniface and N. Hasebe

889 **Effect of Material Properties in Certain Thermoelastic Contact Problems**
D. Joachim-Ajao and J. R. Barber

894 **Two Interpretations of Rigidity in Rigid-Body Collisions**
A. Chatterjee and A. Ruina

901 **Free and Forced Vibration of an Axially Moving String With an Arbitrary Velocity Profile**
W. D. Zhu and B. Z. Guo

908 **Flexibility-Based Finite Element Analysis for Structures With Random Material Properties**
Yongjian Ren and I. Elishakoff

914 **Application of Homotopy Analysis Method in Nonlinear Oscillations**
Shi-Jun Liao and A. T. Chwang

923 **Geometric Stiffening of Timoshenko Beams**
D. C. D. Oguamanam and G. R. Heppler

930 **Transient Bending Wave Propagation in Anisotropic Plates**
K.-E. Fällström and O. Lindblom

939 **A New Algebraic Rigid-Body Collision Law Based on Impulse Space Considerations**
A. Chatterjee and A. Ruina

(Contents continued on outside 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.

Downloaded 04 May 2010 to 171.66.16.52. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

CONTENTS (CONTINUED)

952 **An Energy-Based Coefficient of Restitution for Planar Impacts of Slender Bars With Massive External Surfaces**
Y. Hurmuzlu

963 **Adaptive Boundary Control of Out-of-Plane Cable Vibration**
H. Canbolat, D. Dawson, C. Rahn, and S. Nagarkatti

970 **Anisotropic Constraint Effects on Interactive Damage Progression of General Lay-up Composites**
J. Fan and J. Zhang

980 **On the Existence of Normal Modes of Damped Discrete-Continuous Systems**
H. T. Banks, Zheng-Hua Luo, L. A. Bergman, and D. J. Inman

990 **On Diverging Concertina Tearing**
T. Wierzbicki, K. A. Trauth, and A. G. Atkins

998 **Tangential Loading of General Three-Dimensional Contacts**
M. Ciavarella

1004 **Hybrid Stochastic Finite Elements and Generalized Monte Carlo Simulation**
R. Ghanem

1010 **Identification of a Dynamic System Using Ambient Vibration Measurements**
M. Q. Feng, J. M. Kim, and H. Xue

1022 **Almost Classically Damped Continuous Linear Systems**
S. Natsiavas and J. L. Beck

1032 **An Infrared Interferometer for Elastoplastic Crack-Tip Field Investigation of Homogeneous and Bimaterial Beams: A Study of Three-Dimensional Effects and *J*-Dominance**
J. K. Sinha and H. V. Tippur

1042 **Asymptotics of Extensional Waves in Isotropic Elastic Plates**
N. A. Losin

1048 **Three-Dimensional Interface Cracks in Anisotropic Bimaterials: The Non-Oscillatory Case**
Jianmin Qu and Yibin Xue

BRIEF NOTES

On the Effect of Loading Rate on the Compressive Strength of an AS4/PEEK Composite 1056
T. J. Vogler and S. Kyriakides

On the Decomposition of Deformation Into Elastic-Plastic Parts 1059
Li-jun Shen

Some Comments on Rocking Motion on Rotating Disk/Spindle Systems 1061
I. Y. Shen

1062 Stability of Gyroscopic Systems Near Vanishing Eigenvalues
A. A. Renshaw

1064 Linear Relationships Between the Stress Fields of ⟨111⟩ Line Defects
Y. T. Chou and Sanboh Lee

1066 Mohr-Coulomb Failure Criterion Expressed in Terms of Stress Invariants
G. S. Schajer

DISCUSSION

1069 **Time-Integral Variational Principles for Nonlinear Nonholonomic Systems, by J. G. Papastravridis—Discussion by B. Ravindra**

1070 **Enhanced Elastic Buckling Loads of Composite Plates With Tailored Thermal Residual Stresses, by S. F. Muller de Almeida and J. S. Hansen—Discussion by Ch. Sarath Babu and T. Kant**

1072 **Toroidal Vibrations of Anisotropic Spheres With Spherical Isotropy, by K. T. Chau—Discussion by W. Q. Chen**

BOOK REVIEW

1075 ***AMR Infobase of Journal Literature on CD-ROM*, by ASME—Reviewed by M. S. Qatu and C. R. Steele**

ERRATUM

1068 **A Volume Integral Equation Technique for Multiple Inclusion and Crack Interaction Problems, by Jungki Lee and Ajit Mal**

(Contents continued on inside back cover)

Downloaded 04 May 2010 to 171.66.16.52. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

CONTENTS (CONTINUED)

ANNOUNCEMENTS AND SPECIAL NOTICES

843	17th Biennial ASME Conference on Mechanical Vibration and Noise—Call for Papers
850	Call for Nominations
888	Symposium on Advances in the Continuum Mechanics and Thermodynamics of Material Behavior—Call for Papers
951	CIMS 2000—Announcement
979	Periodicals on ASMENET
989	Change of Address Form
1009	Fourth International Conference on Vibration Problems—First Announcement and Call for Papers
1077	Books Received by the Office of the Technical Editor
1079	Worldwide Mechanics Meetings List
1082	1999 ASME Mechanics and Materials Conference—Announcement
1083	Fifteenth ASME Freeman Scholar Program in Fluids Engineering—Announcement
1084	Information for Authors
1085	Annual Index

Downloaded 04 May 2010 to 171.66.16.52. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Frequency Clusters in the Spectrum of Annular Cylinders

K. I. Tzou
Mem. ASME

J. A. Wickert
Mem. ASME

A. Akay
Fellow ASME

Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213

As the length of a traction-free annular cylinder is increased, distinct members within any family of radial or longitudinal shear modes have natural frequencies that asymptotically approach a common nonzero value. Such modes, potentially having significantly different numbers of nodes along the cylinder's generator, can have natural frequencies that are indistinguishable from one another within the resolution of test equipment or numerical simulation. The three-dimensional vibration model discussed here predicts the formation of narrow "frequency clusters" with the cylinder's increasing length, the converged value of which bounds from below the frequencies of all modes within a particular family. In addition to these spectral characteristics, frequency clusters have implications for the forced response of annular cylinders. For the particular families of modes that are of interest here, the steady-state harmonic response at frequencies near a cluster can be spatially confined with displacements that decay rapidly away from the point of maximum response. At other driving frequencies, the response is distributed more uniformly along the length of the cylinder. The derived analytical model is compared with results from laboratory measurements, and from the predictions of wave propagation theory in the limit of infinite cylinder length.

1 Introduction

Annular cylinders are common structural elements, and the study of their vibration forms a classical area within the discipline of elastodynamics. The structural and acoustic responses of a variety of engineering systems—including submersibles, projectiles, waveguides, ducts, pipes, and fuselages—are treated to first order by using cylinder models. In such applications, knowledge of the mode shapes and natural frequencies is important in setting the structure's response, and, for instance, its potential interaction with a surrounding fluid.

Classical theories for the bending, torsion, and longitudinal vibration of the traction-free annular cylinder in Fig. 1(a) distinguish vibration modes according to the number m of nodes along the z-coordinate, the cylinder's generator. Similarly, the integer n is used conventionally in the shell vibration literature to identify the number of diametral nodes (Leissa, 1973). In annular cylinders, usage of the third index l for nodes in the radial direction is reserved for higher-order modes (McDaniel and Ginsberg, 1993) which are beyond the frequency range of interest here. For the radial and longitudinal shear modes examined in what follows, the radial u_r and circumferential u_θ displacement components have the same values of m as well as n.

As the diameter ratio $a^* = a/b$ approaches unity and the cylinder's wall becomes thin, u_r dominates u_θ for the "radial" class of modes. For example, Fig. 1(b) depicts the u_r nodal patterns for a radial mode with $m = 1$ and $n = 2$; hence, the notation ($m = 1$, $n = 2$)-R is adopted to identify this mode. "Longitudinal shear" modes, on the other hand, exhibit dominant u_z (Gazis, 1959) and are identified by the notation (m, n)-LS with $n > 0$. Unlike modes seen in classical, strength-of-materials, longitudinal vibration problems, the axial displacement in a longitudinal shear mode varies with θ over a given cross section. The (2, 1)-LS mode in Fig. 1(c) is one such example.

By way of motivation, Fig. 2 shows collocated point transfer functions that were measured with cylinders of four different lengths; the length-to-diameter ratio $L^* = L/b$ is indicated next to each transfer function. These measurements were obtained by exciting the cylinder with a piezoelectric force hammer, measuring the acceleration response, and capturing the resulting signals with a dual channel signal analyzer. Values m and n were determined through conventional methods of experimental modal analysis, with m being indicated next to each peak in Fig. 2. The cylinders were supported by foam rubber mounts which were sufficiently compliant, relative to the cylinders themselves, that their contributions to the natural frequency spectrum were judged to be negligible. The majority of the spectral peaks shown correspond to the family ($m \geq 0$, 2)-R of modes. At $L^* = 7.50$ and 11.4, the cylinder's beam-like bending modes also fall within the sampled frequency range. They are distinguished from the radial modes by having their values of m enclosed by parentheses in Fig. 2.

With increasing length, the cylinder's natural frequencies for the family (m, 2)-R of modes converged from above to a common value, approximately 2.14 kHz in Fig. 2. Once converged, the frequencies for all modes within that family remained sensibly constant at 2.14 kHz, even as L^* was further increased. The (1, 2)-R mode, for instance, had frequency 2.22 kHz at $L^* = 2.86$, and that value decreased with L^* until its frequency became indistinguishable near $L^* = 10$ from the 2.14 kHz value that was maintained by (0, 2)-R. Similarly, the frequency for the $m = 2$ mode approached 2.14 kHz from above, and merged near $L^* = 15$ with the "frequency cluster" already formed by the $m = 0$ and $m = 1$ modes. Although the cylinder's classical beam bending modes did sweep through the 2.0–3.6 kHz range as L^* was varied, they exhibited no sensitivity to the frequency cluster, and their frequencies decreased monotonically with length in the expected fashion. In Fig. 2, only the natural frequencies for the family (m, 2)-R of modes clustered at 2.14 kHz, although similar clusters existed at higher frequencies for the other radial families (m, $n \geq 3$)-R. What is most notable in Fig. 2 is that for a cylinder of sufficiently large length-to-diameter ratio, radial vibration modes within the same family, each having a different number of nodes in z, had finite-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Nov. 11, 1997; final revision, Mar. 2, 1998. Associate Technical Editor: N. C. Perkins.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

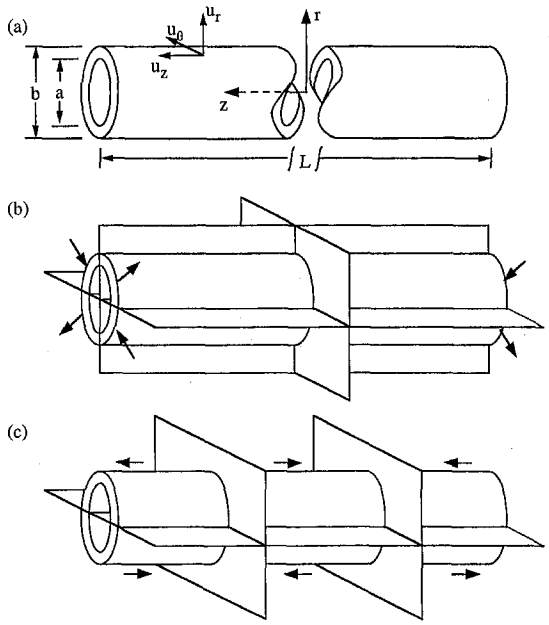


Fig. 1 (*a*) Schematic diagram of a traction-free annular cylinder; (*b*) motion and nodal planes of u_r for the (1, 2)-*R* mode; (c) motion and nodal planes of u_z for the (2, 1)-*LS* mode

valued natural frequencies that were virtually indistinguishable from one another.

Although an extensive literature on the vibration of cylinders is available (Thurston, 1978; 1992), few studies have investigated the effect of length-to-diameter ratio on the radial and longitudinal shear modes. For $L^* > 10$, frequencies for the $(0, n)$-*R* and *LS* modes in solid and annular cylinders have been shown experimentally to be insensitive to the cylinder's length (McMahon, 1964; Singal and Williams, 1988). Hutchinson (1980) and Wang and Williams (1996) recognized that over the limited range $1 \leq L^* \leq 10$, the frequencies for the families $(m > 0, n)$-*R* and *LS* of modes decrease steadily with L^*. The phenomenon whereby the frequencies for these modes cluster around a common value at large L^*, however, has evidently not been examined.

In this paper, the free and forced responses of an annular cylinder having arbitrary length are investigated with a view towards understanding frequency clusters that appear in the cylinder's spectrum. Experimental results are compared to those from a three-dimensional vibration model and to the limiting behavior that is predicted by wave propagation theory. As a cautionary note, the findings are applicable only to the cylinder's radial and longitudinal shear modes which, for most realistic geometries, are high in frequency compared to the bending modes.

2 Spectrum of a Finite Length Cylinder

Free vibration of an isotropic, homogeneous, and traction-free elastic cylinder is analyzed through Ritz discretization of the equations for three-dimensional elastodynamics. The dimensionless forms of the strain and kinetic energies evaluated over the cylinder shown in Fig. 1 are

$$U = \int_V (\tfrac{1}{2}\lambda\Delta^2 + \mu(\epsilon_{rr}^2 + \epsilon_{\theta\theta}^2 + \epsilon_{zz}^2 + 2(\epsilon_{r\theta}^2 + \epsilon_{\theta z}^2 + \epsilon_{rz}^2)))dV \quad (1)$$

$$T = \tfrac{1}{2}\int_V (u_{r,t}^2 + u_{\theta,t}^2 + u_{z,t}^2)dV \quad (2)$$

where λ and μ are Lamé constants; the dilatation is $\Delta = \epsilon_{rr} + \epsilon_{\theta\theta} + \epsilon_{zz}$; the comma subscript notation indicates partial differentiation; all lengths and displacements are nondimensionalized with respect to b; time is referenced to the dilatational wave speed $c = ((\lambda + 2\mu)/\rho)^{1/2}$; and ρ is the material's volumetric density. By using the appropriate strain displacement and constitutive laws, U and T are expressed in terms of u_r, u_θ, and u_z, which in turn are spatially discretized through admissible, finite sum, expansions (Tzou et al., 1998). For example, during free single-mode vibration,

$$u_r = \sin \omega^* t \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} q_r^{(ijk)} R_i(r)\Theta_j(\theta)Z_k(z), \quad (3)$$

where I, J, and K are the numbers of functions used in r, θ, and z, respectively; the $q_r^{(ijk)}$ are generalized coordinates; the R_i are polynomials generated by Gram-Schmidt orthonormalization over $(a^* < r < 1)$; the Θ_j are sine and cosine functions; the Z_k are Legendre polynomials over $(-L^*/2 < z < L^*/2)$; the dimensionless frequency is defined $\omega^* = b\omega/2c$; and ω is the dimensional circular frequency. Displacement components u_θ and u_z are treated similarly with the same number of expansion terms. Application of Lagrange's equation to the discretized Eqs. (1)–(2) yields the eigenvalue problem $\mathbf{Kq} = \omega^{*2}\mathbf{Mq}$ of dimension $6IJK + 3IK$, where the mass matrix $\mathbf{M}$ is identity; the stiffness matrix $\mathbf{K}$ is sparse and positive semi-definite; and the vector $\mathbf{q} = [\mathbf{q}_r^T\ \mathbf{q}_\theta^T\ \mathbf{q}_z^T]^T$ comprises the generalized coordinates for all three displacement components. Solutions of the eigenvalue problem are found using conventional *QR* techniques, with convergence realized to at least three significant digits with $I = J = 4$ and $K = 12$ for all modes discussed in the sequel.

In Fig. 3, the frequency spectrum is predicted as a function of the cylinder's length.[1] The natural frequencies for the bending,

[1] Illustratively, the values $\lambda = 50.3$ GPa, $\mu = 25.9$ GPa, and $\rho = 2860$ kg/m^3 were used for aluminum.

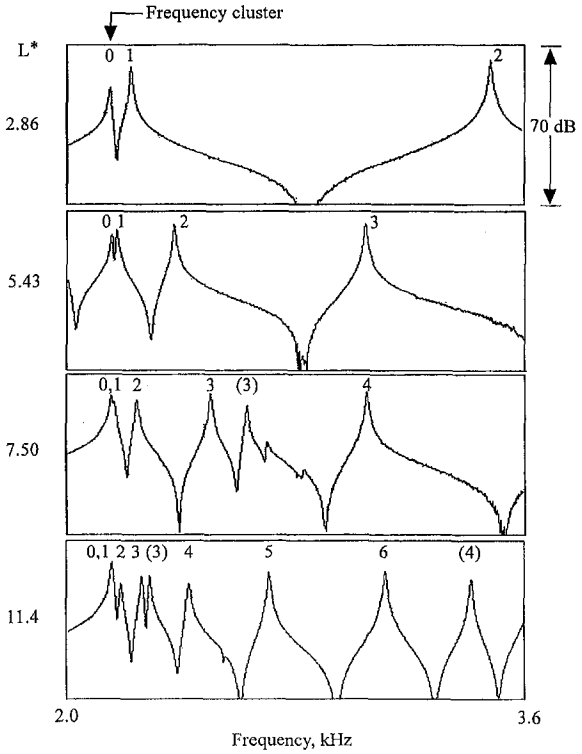


Fig. 2 Measured collocated point transfer functions of four different cylinders; aluminum, a = 77 mm, b = 89 mm, L = 254, 483, 667, and 1020 mm. The frequencies for the family $(m, 2)$ – *R* of modes are bounded from below at 2.14 kHz. The number m of nodal planes in z for u_r is indicated next to each peak. Values in parentheses denote the number of nodes for the bending modes, which are not bounded by the cluster.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

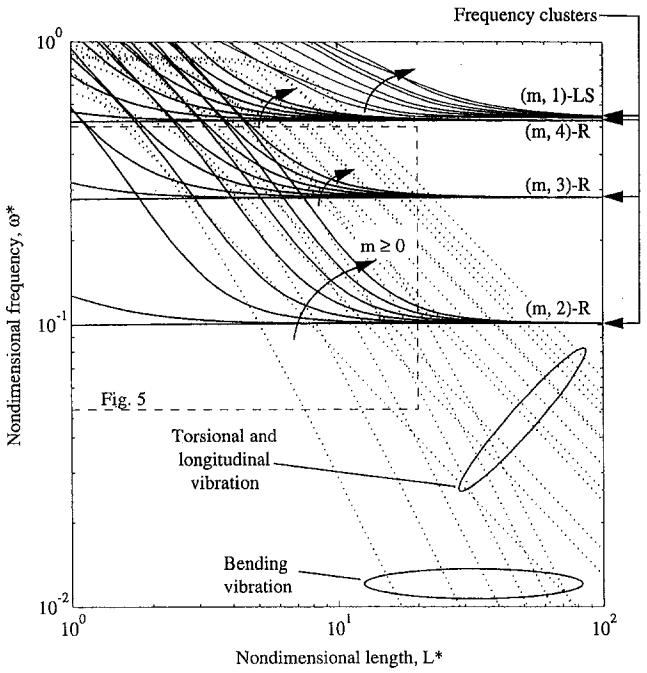


Fig. 3 Predicted natural frequencies as a function of cylinder length; $a^* = 0.868$. The line types denote classical bending, torsion, and longitudinal modes (· · ·), or *R* and *LS* modes (——). The inset region is magnified in Fig. 5.

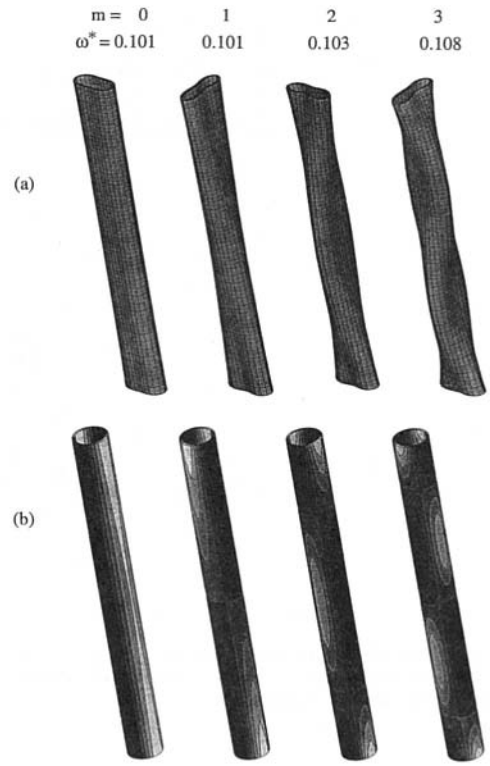


Fig. 4 Lowest four-mode shapes (m = 0, 1, 2, and 3) in the family (m, 2)-*R* of modes; $L^* = 10$, and $a^* = 0.868$; (*a*) isometric view, and (*b*) contours of u_r

torsion, and longitudinal vibration modes are shown as dotted lines, and those values agree with their respective classical strength-of-materials theories for appropriate values of L^* and m. Of particular interest in Fig. 3 are the vibration modes for which the natural frequencies converge to a constant value at large L^*. Their loci are drawn with solid lines, and the labels identifying the *R* or *LS* families are shown near each frequency cluster. Within each family of modes, the rate at which a particular locus coalesces with the frequency cluster depends on both L^* and m. Likewise, each family of modes is distinguished by a particular frequency to which its members converge. For instance, radial modes having any value of m and n = 2, 3, or 4 nodal planes in θ cluster near ω^* = 0.101, 0.282, or 0.528, respectively. Wang and Williams (1996) incorrectly concluded that the frequencies of the $(m, 2)$-*R* family continue to decrease with L^* below the discussed bound.

The natural frequencies for the longitudinal shear modes behave similarly with increasing L^*. The family $(m, 1)$-*LS* of modes converge near ω^* = 0.540 in Fig. 3. As with the $(m, 2)$-*R* family shown in Fig. 2, the natural frequencies of any two $(m, 1)$-*LS* modes can be rendered arbitrarily close by setting L^* to be sufficiently large. At L^* = 100, for instance, the (7, 1)-*LS* mode has a natural frequency 1.6 percent higher than that of the (0, 1)-*LS* mode despite the greater number of nodes along the cylinder's generator.

The isometric views and contour diagrams of u_r in Fig. 4 depict the number and location of nodal planes in z for the lowest four members of the $(m, 2)$-*R* family at L^* = 10. Due to axisymmetry, frequencies for all modes in this family are repeated, although only one member of each degenerate sine/cosine-oriented pair is shown. To three digits of accuracy, the nondimensional frequencies for these modes are 0.101, 0.101, 0.103, and 0.108. At this level of precision, the frequencies for the lowest two modes are "identical," notwithstanding the presence of the additional nodal plane in the second mode. Likewise, the frequency of the third mode is only some three percent higher than the frequency of the (0, 2)-*R* and (1, 2)-*R* cluster, but it has two additional nodal planes. As L^* grows further, the frequencies of (2, 2)-*R* and (3, 2)-*R* also tend toward ω^* = 0.101. At $L^* \geq 22$, for instance, the frequency of (2, 2)-*R* also becomes ω^* = 0.101, and at $L^* \geq 49$, all four modes shown in Fig. 4 have the same frequency at this level of precision. Such identical frequencies for "clustered" modes arise at large L^* and are not associated with axisymmetry. Consequently, frequency clusters are said to develop when two or more modes of the same family have natural frequencies that are indistinguishable at a specified, potentially high, level of resolution, yet have different numbers of nodes in z.

Comparisons between measured and predicted frequencies for the families $(m, 2)$-*R* and $(m, 3)$-*R* of modes are shown in Fig. 5. The measured values are identified from transfer function measurements as in Fig. 2. Over the range shown, frequencies for all modes agree within a relative root-mean-squared error of four percent.[2] As in Figs. 2 and 3, the frequency clusters in Fig. 5 bound from below the frequencies of all *R* modes within the same family, and for finite and increasing L^*, all of the $m \geq 0$ loci in the family coalesce with the cluster. As $L^* \to \infty$, a countably infinite number of modes within the same family will cluster and have frequencies arbitrarily close to one another, thus forming an accumulation point in the cylinder's frequency spectrum.

3 Calculation of Frequency Clusters

At large L^*, the frequency clusters identified in Figs. 2, 3, and 5 for radial and longitudinal shear modes are shown below to form at the cutoff frequencies ω_c^* as predicted by wave propagation calculations. Bending waves in an infinite cylinder have cutoff frequency $\omega_c^* = 0$, but radial and longitudinal shear waves have finite-valued nonzero ω_c^*. Indeed, the cutoff fre-

[2] At L^* = 2.86, 3.71, 5.43, and 7.50, the repeated natural frequencies of both the (1, 3)-*R* and (2, 3)-*R* pairs of modes split due to minor circumferential variation of the cylinder's wall thickness.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

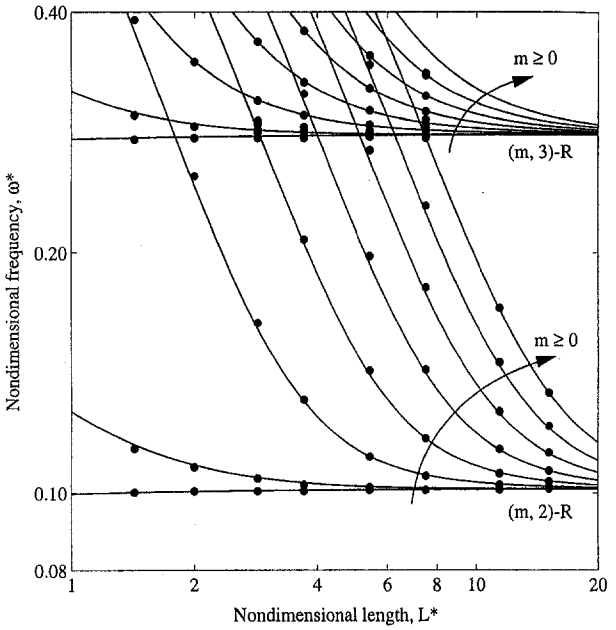


Fig. 5 Detail of the frequency spectrum shown in Fig. 3 in which measured (● ● ●) and predicted (——) frequencies for the families (m, 2)-R and (m, 3)-R of modes are compared

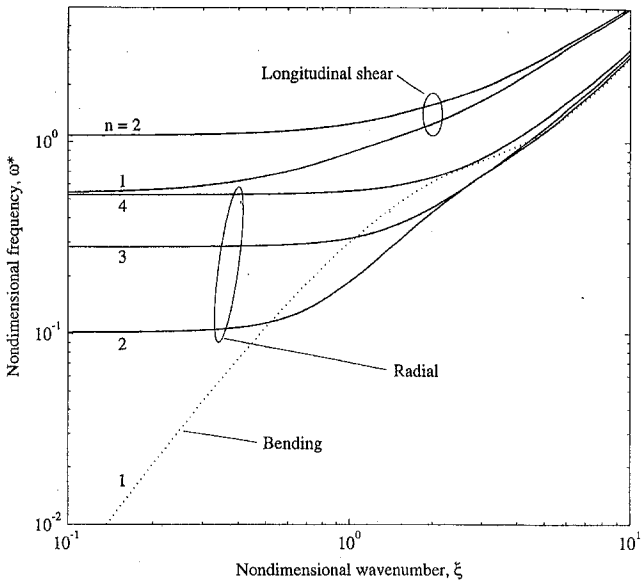


Fig. 6 Frequency spectrum of lower-order bending (· · ·), and radial or longitudinal shear (——), waves in an infinite annular cylinder; $a^* = 0.868$

quencies for those families of waves are the same as the asymptotes in L^* established by the frequency clusters.

In the case of an infinite cylinder, the cutoff frequency for a particular wave is determined through standard methods of elastodynamics (Graff, 1975). Each displacement component exhibits $\exp(-i(\xi z + \omega^* t))$ functional dependence along the cylinder's generator, where the wave number ξ in z is nondimensionalized with respect to b. The cutoff frequency is the lowest value at which a wave will propagate, and it is found from the condition $\xi \to 0$. The $\omega - \xi$ dependencies have been calculated previously for the bending, torsion, radial, longitudinal, and longitudinal shear waves in infinite solid (Pao and Mindlin, 1960) and annular (Armenàkas et al., 1969) cylinders, and those results are useful to predict frequency cluster values.

In Fig. 6, the spectrum is shown for the lower-order bending, longitudinal shear, and radial waves in an infinite cylinder with $a^* = 0.868$. The bending waves have frequencies proportional to ξ^2 for small wave numbers (roughly $\xi < 0.3$) where the effects of rotary inertia and shear deformation are negligible (Graff, 1975); consequently, the cutoff frequency is zero. On the other hand, loci for the radial and longitudinal shear classes of waves have vanishing slope in Fig. 6 as $\xi \to 0$; the cutoff frequencies in those cases are determined by the nonzero intercepts. Analogous behavior has been observed in duct acoustics where plane waves exhibit $\omega_c^* = 0$, but higher-order waves having multiple nodal planes across the duct's section do not (Morse and Ingard, 1968).

For the case of longitudinal shear waves, the cutoff frequency's dependence on a^* is determined from the characteristic equation

$$J_n'(\beta a^*)Y_n'(\beta) - J_n'(\beta)Y_n'(\beta a^*) = 0 \qquad (4)$$

for n diametral planes (Gazis, 1959), where J_n and Y_n are the nth-order Bessel functions of the first and second kind, and $\beta = \omega_c^*((\lambda + 2\mu)/\mu)^{1/2}$ is a dimensionless frequency parameter.[3] Equation (4) is obtained from the simplification $\xi = 0$ of the full dispersion relation and the stipulation that motions are independent of z. The characteristic equation for radial waves is similarly available to predict the frequency cluster values for the (m, n)-R modes (Gazis, 1958).

The dependence on a^* of the cutoff frequencies for radial and longitudinal shear waves is shown in Fig. 7. As the cylinder becomes more solid ($a^* \to 0$), the cutoff frequencies for both the R and LS families fall within the same range, with the $n = 1$ longitudinal shear wave having the lowest value for ω_c^*. As $a^* \to 1$, cutoff frequencies for longitudinal shear waves decrease only moderately with a^*, and those for the radial waves tend to zero. For very thin shells, then, the $n = 1$ longitudinal shear wave has the lowest nonzero cutoff frequency, and the first frequency cluster will develop for the (m, 1)-LS family of modes. For cylinders with moderate wall thicknesses ($0.30 < a^* < 0.95$), however, the $n = 2$ radial wave has the lowest

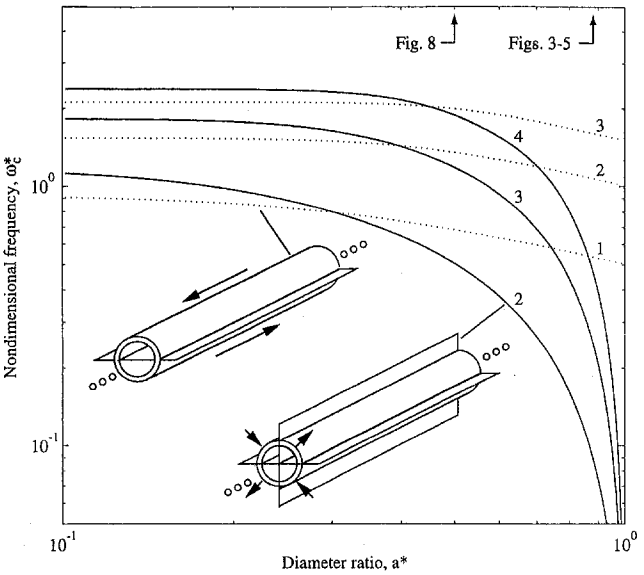


Fig. 7 Cutoff frequencies as a function of the cylinder's diameter ratio. The line types denote radial (——) or longitudinal shear (· · ·) waves. The insets depict the nodal planes and dominant directions of particle motion for representative waves.

[3] For consistency here, the nondimensionalization of β has been changed relative to that given by Gazis (1959).

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

nonzero cutoff frequency, and hence the first frequency cluster develops in that case for the $(m, 2)$-R family.

The predicted cutoff frequencies are consistent with the measured and predicted clusters discussed above for a finite cylinder. At $a^* = 0.868$ as in Fig. 3, $\omega_c^* = 0.101$, 0.282, and 0.528 for the $n = 2$, 3, and 4 radial waves; likewise $\omega_c^* = 0.540$ for the $n = 1$ longitudinal shear wave. A further comparison is made in Fig. 8 where the frequency spectrum for cylinders with $a^* = 0.5$ is shown. The number of expansion terms in Eq. (3) are adjusted here to $I = 6$, $J = 4$, $K = 12$. In this case, the frequency clusters appear at $\omega^* = 0.505$ and 1.18 for the $n = 2$ and 3 radial modes, and at $\omega^* = 0.682$ and 1.35 for the $n = 1$ and 2 longitudinal shear modes. Those values correlate to three digits of accuracy with the cutoff frequencies in Fig. 7, as predicted by Eq. (4).

This correlation between frequency clusters and cutoff frequencies reflects continuous transition of the natural frequencies and mode structure as the cylinder's length grows. As $L^* \to \infty$ and for m fixed, the axial separation between adjacent nodal points grows. With the corresponding wave number likewise decreasing, the natural frequencies approach the cutoff frequencies from above, as indicated by Fig. 6.

4 Response to a Harmonic Point Load

When a long but finite annular cylinder is excited harmonically at or near a frequency cluster, multiple modes within the same family will contribute to the response, more or less equally with respect to the frequency magnification factor, despite their having different values of m. Although each individual mode within a cluster has energy distributed along the cylinder's axis (as in Fig. 4), the superposition of many clustered modes need not. In this section, the response of a long cylinder to a harmonic point load near a frequency cluster is examined experimentally and through modal analysis, demonstrating that a spatially confined response can occur, much in the same vein as an evanescent wave.

For the test cylinder, the locations of the frequency clusters were determined through measured collocated transfer functions. In Fig. 9(a), beam bending modes dominated the low frequency range, while both bending and $(m, 2)$-R modes were present in the 1.2–3.2 kHz range. The spectrum was dense at the frequency cluster with the nine radial modes $m = 0, \ldots, 8$ being concentrated near 1.225 kHz. The transfer function's magnitude at that point was also roughly four times greater than the mean of the other $(m \geq 9, n = 2)$ radial modes. In the magnified view shown in Fig. 9(b), the frequencies for the first five $(m, 2)$-R modes remained indistinguishable at the higher resolution, and the transfer function's corresponding magnitude

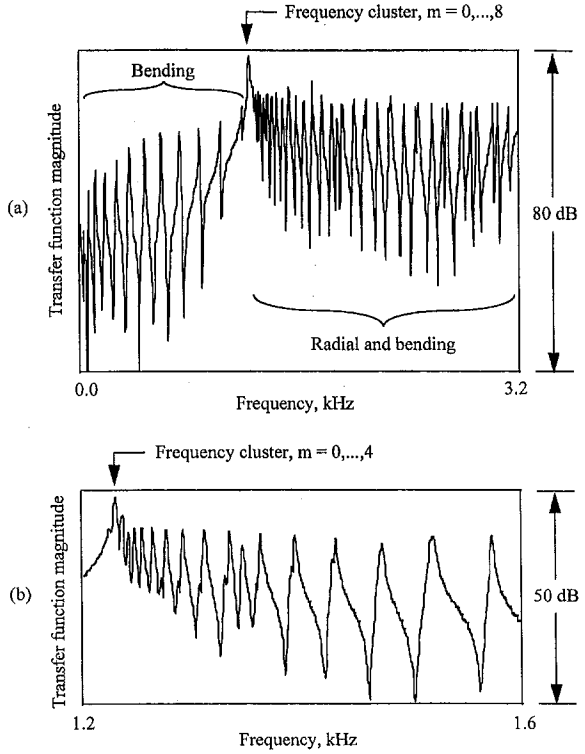


Fig. 9 (*a*) Measured collocated point transfer function for a long cylinder (steel, $a = 0.104$ m, $b = 0.115$ m, $L = 6.02$ m) is characterized by increased amplitude and modal density at the 1.225 kHz frequency cluster; $L^* = 52.4$, $a^* = 0.90$. Nine modes are clustered at this resolution. (*b*) Five modes remain concentrated in the cluster at higher resolution.

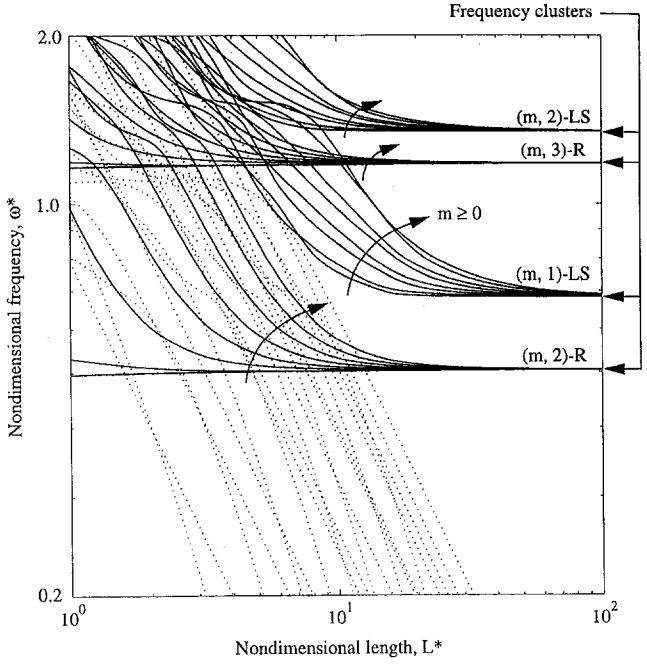


Fig. 8 Predicted natural frequencies as a function of cylinder length; $a^* = 0.5$. The line types denote bending, torsion, and longitudinal modes (· · ·), or *R* and *LS* modes (——).

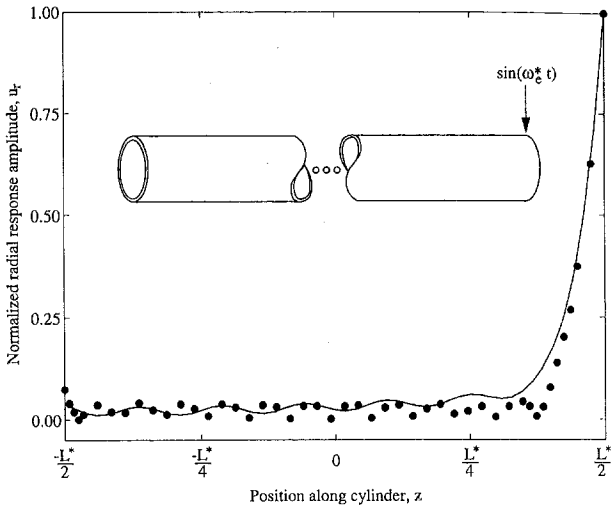


Fig. 10 Response of an annular cylinder driven at 1.208 kHz near, but slightly below, the 1.225 kHz frequency cluster; $L^* = 52.4$, $a^* = 0.90$; measured (● ● ●), predicted (——). The curves shown are the projection of the response onto the $r - z$ plane at $\theta = 0$.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

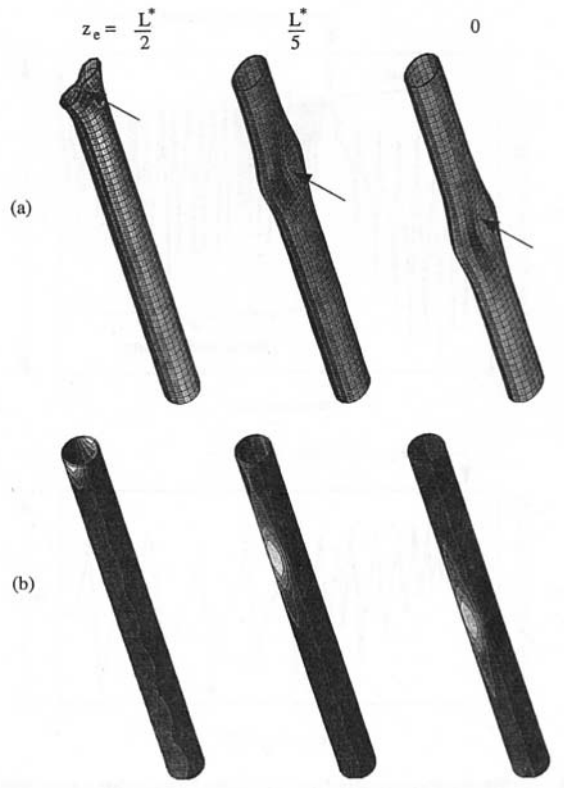


Fig. 11 Predicted point load response of an annular cylinder driven below the $\omega^* = 0.0771$ frequency cluster; $\omega_e^* = 0.0760$, $L^* = 52.4$, $a^* = 0.90$. The response is confined near the excitation point as indicated by the arrow. (*a*) Isometric view, and (*b*) contours of u_r.

at 1.225 kHz was halved relative to that in Fig. 9(*a*). Other frequency clusters at higher frequencies for the families $n = 3, 4, \ldots$ of radial modes also exhibited this amplitude dependence on resolution. Because the frequencies of individual modes within a frequency cluster were virtually identical, the finite width of the frequency bin used by the spectrum analyzer did not resolve each clustered mode individually for large L^*. Consequently, and by way of speculation, frequency clusters in experimentally measured transfer functions appeared to have increased magnitude due to the finite resolution of the test equipment.

Figure 10 shows the measured and predicted responses of the cylinder to a radially oriented point load at excitation frequency $\omega_e^* = 0.07600$ (1.208 kHz). In the tests, a constant amplitude harmonic force was applied using a controlled electromagnetic shaker located at one end. The amplitude of u_r was obtained from measured acceleration response. As damping is not present in the model, the displacements of both responses are normalized to their respective maximum amplitude at $z = L^*/2$ for the purpose of comparing the spatial response shape. The response in Fig. 10 shows that u_r decays rapidly with increasing distance from the excitation source. The resonant response is highly confined near the excitation point $z_e = L^*/2$, in contrast to the distributed displacements that are characteristic of each of the cylinder's vibration modes in this frequency range. The low-amplitude spatially periodic component between $-L^*/2 \leq z \leq L^*/4$ arises from leakage of the $m = 13$ bending mode which is nearest in frequency to ω_e^*.

The predicted response obtained through modal analysis is also shown in Fig. 10.[4] For the corresponding frequency cluster $\omega_e^* = 0.07705$ (1.225 kHz), the response sums over all modes having $m \leq 13$, with frequency convergence to five significant figures obtained with $I = 4$, $J = 2$, and $K = 70$. The predicted response decays at a rate nearly identical to that of the measured response. The full isometric views and u_r contours of the predicted response, shown in Fig. 11, demonstrate that this confined response occurs regardless of the excitation's location for those cases shown.

At excitation frequencies near the cluster, the spatial distribution of response is highly sensitive to even minor changes in ω_e^*. In Fig. 12, the modal coordinate amplitudes for point load excitation at $L^*/2$ are shown as a function of ω_e^*. At this resolution, the (0, 2)-R and (1, 2)-R modes have formed a cluster at $\omega^* = 0.07705$. Over $0.07650 < \omega_e^* < 0.07710$, the modal coordinates of these two modes are nearly an order of magnitude greater than those of all other modes. The isometric views in Fig. 12 show that over this frequency range, confined responses occur, similar to those shown in Figs. 10 and 11. When the cylinder is driven at higher ω_e^*, the modal coordinates for (0, 2)-R and (1, 2)-R each decrease in magnitude, and the resulting response is determined by a more equal superposition of the $m > 1$ radial modes. The response in that case can have displacements distributed over the length of the cylinder as a combination of modes ($\omega_e^* = 0.07718$), or near resonant response ($\omega_e^* = 0.07767$). The axial distributions of strain energy and their dependence on excitation frequency are reminiscent of evanescent and propagating waves in infinite media, reinforcing the duality between wave propagation and modal representations of vibration.

5 Summary

Experimental and analytical investigations of the vibration of finite cylinders demonstrate that individual modes within the radial and longitudinal shear families cluster near a common frequency as the cylinder's length is increased. For sufficiently

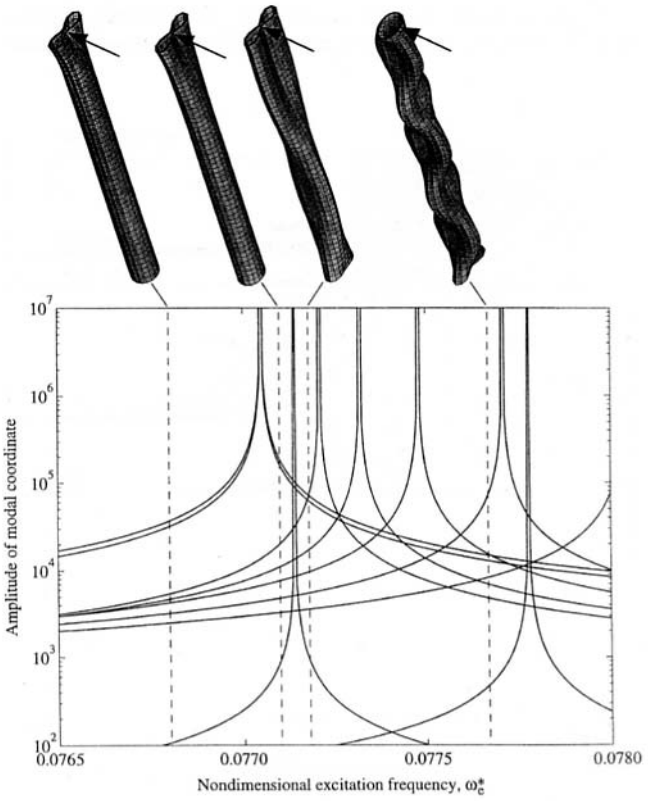


Fig. 12 Amplitude of modal coordinates as a function of ω_e^* in the point load response of an annular cylinder; $z_e = L^*/2$, $L^* = 52.4$, $a^* = 0.90$

[4] For steel, the values $\lambda = 112$ GPa, $\mu = 74.4$ GPa, and $\rho = 7900$ kg/m^3 were used.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

long cylinders, these frequencies match the cutoff frequencies as predicted by wave propagation solutions for an infinite cylinder. In the response of long finite cylinders to harmonic radial excitation, the overall response can be spatially confined when the driving frequency is near the cluster. At excitation frequencies well above the cluster, on the other hand, the response has displacements that are distributed along the cylinder's length, at least in the case studies examined here.

By way of application, frequency clusters and the forced response at frequencies near clusters can influence the testing of annular cylinders, and for instance, acoustic radiation from them. Without sufficient measurement or numerical resolution and knowledge of frequency clusters, damping measurements of a spectral peak containing multiple modes yield an apparent, equivalent damping for all modes within the cluster. Further, the high modal density complicates curve fitting of individual spectral peaks, hindering accurate evaluation of mode shapes.

References

Armenàkas, A. E., Gazis, D. C., Herrmann, G., 1969, *Free Vibrations of Circular Cylindrical Shells,* Pergamon Press, Tarrytown, NY.

Gazis, D. C., 1958, "Exact Analysis of the Plane-Strain Vibrations of Thick-Walled Hollow Cylinders," *Journal of the Acoustical Society of America,* Vol. 30, pp. 786–794.

Gazis, D. C., 1959, "Three-Dimensional Investigation of the Propagation of Waves of Hollow Circular Cylinders," *Journal of the Acoustical Society of America,* Vol. 30, pp. 786–794.

Graff, K. F., 1975, *Wave Motion in Elastic Solids,* Oxford University Press, New York.

Hutchinson, J. R., 1980, "Vibrations of Solid Cylinders," ASME JOURNAL OF APPLIED MECHANICS, Vol. 47, pp. 901–907.

Leissa, A. W., 1973, "Vibration of Shells," NASA SP-288.

McDaniel, J. G., and Ginsberg, J. H., 1993, "Thickness Expansions for Higher-Order Effects in Vibrating Circular Shells," ASME JOURNAL OF APPLIED MECHANICS, Vol. 60, pp. 463–469.

McMahon, G., 1964, "Experimental Study of the Vibrations of Solid, Isotropic, Elastic Cylinders," *Journal of the Acoustical Society of America,* Vol. 36, pp. 85–92.

Morse, P. M., and Ingard, K. U., 1968, *Theoretical Acoustics,* Princeton University Press.

Pao, Y. H., Mindlin, R. D., 1960, "Dispersion of Flexural Waves in an Elastic Circular Cylinder," ASME JOURNAL OF APPLIED MECHANICS, Vol. 27, pp. 513–520.

Singal, R. K., and Williams, K., 1988, "A Theoretical and Experimental Study of Vibrations of Thick Cylindrical Circular Shells and Rings," ASME *Journal of Vibration, Acoustics, Stress, and Reliability in Design,* Vol. 110, pp. 533–537.

Thurston, R. N., 1978, "Elastic Waves in Rods and Clad Rods," *Journal of the Acoustical Society of America,* Vol. 64, pp. 1–37.

Thurston, R. N., 1992, "Elastic Waves in Rods and Optical Fibers," *Journal of Sound and Vibration,* Vol. 159, pp. 441–467.

Tzou, K. I., Wickert, J. A., Akay, A., 1998, "In-Plane Vibration Modes of Arbitrarily Thick Disks," ASME *Journal of Vibration and Acoustics,* Vol. 120, pp. 384–391.

Wang, H., Williams, K., 1996, "Vibrational Modes of Thick Cylinders of Finite Length," *Journal of Sound and Vibration,* Vol. 191, pp. 955–971.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Transient Discharge of Entrained Air From a Wound Roll

M. B. Keshavan
Mem. ASME

J. A. Wickert
Mem. ASME

Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213

As magnetic tape or other web-like material is wound onto a roll, air moving with the incoming stream and the roll becomes forced into the converging wedge at the stream's point of tangency. The spiral-shaped air bearing so formed can extend many wraps into the roll's interior. When the roll is subsequently brought to rest, the entrained air discharges from it, and the roll's radius gradually decreases until all adjacent tape layers have come into direct contact. In the present paper, a model is developed for this transient discharge process, and for the rate at which the roll stabilizes following steady-state winding. Predictions of the model are compared with results from laboratory experiments in which the roll's radius is measured through laser interferometry during the steady entrainment, and transient discharge, stages of winding. Parametric effects of the tape's tension, speed, width, and surface roughness are specifically addressed with a view towards reducing the time required for the roll to stabilize.

1 Introduction

Magnetic tape systems span a broad range in data storage with applications including audio, video, instrumentation, and computer systems. Within the next decade, compact tape cartridges are projected to provide several terabytes of capacity, in part through a nearly fiftyfold growth of the areal storage density. As the technology migrates towards thinner media substrates (4–5 μm), lower tensions (0.28–0.56 N), and higher speeds (7–10 m/s) to improve capacity and access time, advances in the mechanics of tape transport and winding become increasingly important.

In particular, the process of air entrainment reduces the stability of the take-up roll as tape is wound onto it. Figure 1 depicts the prototypical winding process where the incoming tape stream of width b is transported onto the roll at speed V and tension T. During this process, air within the boundary layers of the incoming stream and the outermost layer on the roll becomes forced into the converging wedge formed at the stream's point of tangency with the roll. Some portion of this entrained air squeezes out directly, and some flows laterally across the tape, but a significant fraction flows into the interior of the layered roll and remains temporarily entrapped. For practical values of the tape's and roll's parameters, the outermost several dozen wraps are not in direct mechanical contact with their adjacent, neighboring layers. Instead, they remain separated from one another by a spiral-shaped self-pressurized air film. Such poor contact between layers of tape on the roll's periphery leads to staggered wrapping, damage to the tape's edges, lateral motion at the head/tape interface, and a longer effective tape length which in turn contributes to problems of tape vibration.

When the roll is eventually decelerated and brought to rest, the entrained air discharges from the nip region and from the upper and lower surfaces of the roll. The rate of outflow is a sensitive function of the tape's tension, width, surface roughness, and the velocity at which the roll was wound. To a large extent, the roll collapse process, and the mechanics of air discharging from it, determine the mechanical integrity and stability of the roll. "Scrambling," a particular type of roll instability, can occur as a result of spontaneous interlayer slippage when tape layers buckle as they are forced against a supporting flange. Uneven air discharge can also lead to a reduction of wound-in tension, which can cause the roll to collapse during deceleration or even after it has been brought to rest. Further, in the technology of interest here, several seconds can elapse between the times when the roll comes to rest and when the entrained air has fully discharged. Even after winding has stopped, complete stabilization of the roll does not occur immediately, a point having implications for the roll's winding, unwinding, cycling, and reversal times.

Existing techniques to minimize the problems of air entrainment include the use of packing or lay-on rollers, or drive belts that preload the tape against the roll. Those solutions are realized at the expense of deforming, and in some cases permanently warping, the tape. However, by many accounts, the fragility of the thinner media in next-generation systems will effectively preclude the use of such mechanical means to reduce the thickness of the entrained air film. An alternative technique that avoids such risk of damage involves the use of backcoats with nonisotropic surface roughness. Such structured roughness can include transverse striations and other variants that provide surface-scale channels to better discharge the entrained air.

Despite technological importance of the problem area, a paucity of published literature is available on air entrainment during high-speed winding. Smith and Von Behren (1989) developed a lumped parameter model which specifies that the film thickness remains constant throughout the entire air bearing region. That analysis did not include the effects of surface roughness, or of the interleaved structure of the tape and air layers. Experimental work is available on the subject of web-roller interactions (Jones, 1992; Ducotey and Good, 1995), particularly with regard to treating entrained air within the context of an apparent loss of traction between the web and roller. More recently, a predictive model and measurement method for the thickness of the entrained air film during steady state winding have been developed (Keshavan and Wickert, 1997).

With the objective of providing insight for the design of tape transport systems, a model for the transient process of air discharge from a wound roll is developed in what follows, and its predictions are compared with results from laboratory experiments. The material's speed, tension, width, and surface roughness are explicitly treated, and their parametric effects

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Nov. 26, 1997; final revision, Apr. 2, 1998. Associate Technical Editor: S. R. Barber.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

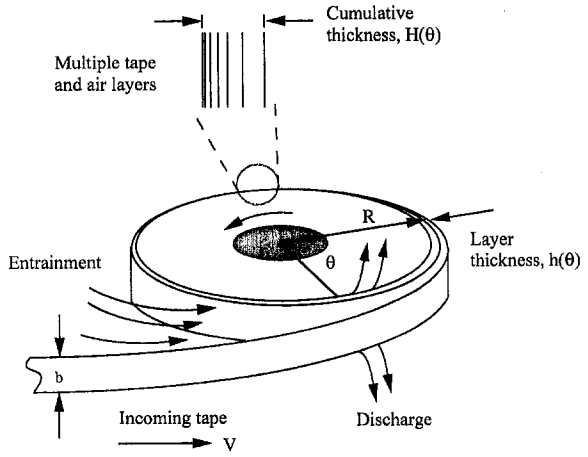


Fig. 1 Illustration of tape winding. Air trapped within boundary layers of the incoming stream and roll forms a spiral-shaped air bearing which can extend many wraps into the roll.

on the air discharge rate from a wound roll are investigated. Experimental results from tape samples having differing surface characteristics are compared to the model's predictions, which are shown to faithfully bound the measured rates of air discharge for the limiting cases of symmetric and asymmetric side flow.

Although the discussion focuses on air entrainment and discharge within the context of flexible media data storage, those processes are equally important in other "web" transport technologies, including the manufacturing of paper, pulp, polymer, and metal foils. To the degree that wound rolls provide the most convenient storage format in each case, air entrainment and discharge define a problem area that is generic across applications and industries dealing with the winding of flat, wide, continuous strips of thin and flexible material.

2 Measurement of Air Entrainment and Discharge

Air entrainment and discharge measurements were made on the tape transport test bed shown in Fig. 2. The radius of the take-up roll was measured by using a laser interferometer in which the reference light beam reflected off a retroreflective mirror, and the target beam was directed at the centerline of the tape on the take-up roll. In each measurement, the roll's radius was recorded as it changed from its value during steady winding (when the entrained air film was fully developed) to the smaller value reached after the roll came to rest (when the air film was fully collapsed). The film thickness was identified from the measured radius after correcting time records for the effects of roll eccentricity and the accretion (or depletion) of tape layers (Keshavan and Wickert, 1997). The test tape's tension and velocity were controlled during steady winding, and during the transient shut-down stage of operation which was triggered once the radius reached $R = 1.8$ cm. All tests were conducted on samples having a polyethylene-terephthalate substrate of width $b = 8$ mm and thickness $d = 9-10$ μm.

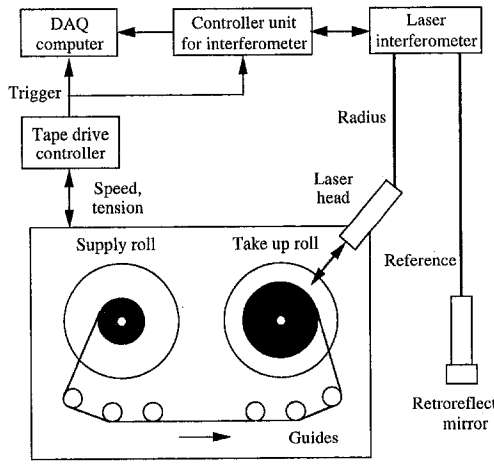


Fig. 2 Laboratory test stand for measuring the cumulative film thickness and discharge rate of entrained air

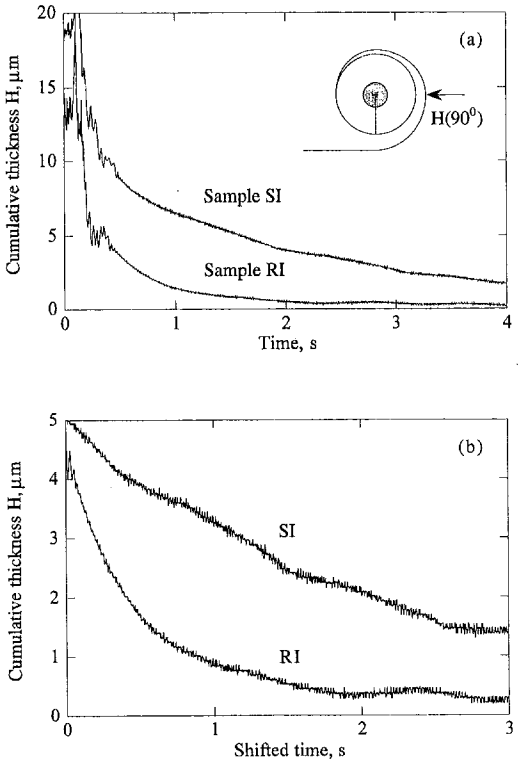


Fig. 3 Measured histories of air discharge for smooth (SI) and rough (RI) isotropic samples; $T = 0.28$ N, $V = 2.3$ m/s, and $\theta = 90$ deg. (*a*) The intercept values yield the cumulative thicknesses at shut-down. (*b*) Time-shifted histories indicate that air discharged more effectively with a rough backcoat than a smooth one.

The air film within the roll is characterized by the following quantities, as suggested in Fig. 1:

- $h(\theta, t)$ for $\theta \geq 0$ and arbitrarily large, which is the thickness of the air film between any two adjacent tape layers,
- $H(\theta, t)$ for $\theta \in [0, 2\pi)$, which is the combined or cumulative thickness of all layers interior to the roll at a particular circumferential location, and
- $\dot{H}(\theta, t)$, which is the rate at a specific circumferential location at which the radius decreases once the roll is brought to rest.

The discharge rate $\dot{H}$ provides a measure of both the settling time that is required for the roll to stabilize, and the efficiency at which air discharges from the lateral surfaces of the wound roll. Depending on the tension and speed at which the tape was initially wound, and its width and surface properties, several or tens of seconds can elapse between the times at which the roll came to rest and all air had discharged from it.

The time histories of Fig. 3(a) demonstrate discharge of entrained air from the take-up roll for two samples having normally distributed isotropic backcoat surfaces. Both samples were wound at identical tension (0.28 N; 1 oz) and speed (2.3 m/s; 90 in/s). However, with a root-mean-square roughness of 300 nm, sample RI (rough isotropic) was significantly rougher than SI (smooth isotropic; 25 nm). During steady winding and

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

prior to shut-down, H (90 deg) = 13.2 μm for RI when compared to the value 18.2 μm for SI. Some 27 percent less air was therefore entrained by sample RI under identical winding conditions. Similarly, as is evident from the discharge histories of Fig. 3(b), the roll of sample RI stabilized more rapidly to the degree that air discharged from it, and its radius diminished, at a rate greater than that of roll SI. The discharge histories shown in Fig. 3(b) are shifted in time relative to those in Fig. 3(a) such that the two samples have identical intercept values H = 5 μm, in order to better highlight the differences in their discharge rates. Note that over only 0.4 sec, H decayed from 5 to 2 μm for sample RI, but nearly five times longer was required for SI to decay to the same value. In the model that follows, the dependence of air discharge rate on surface topography is further investigated.

3 Transient Air Discharge Model

The model and measurements of transient air discharge are discussed in dimensionless form. The independent variables that govern $\dot{H}$ involve combinations of the roll's nominal radius; the tape's tension, width, and velocity; the viscosity μ of the surrounding fluid; the time t elapsed since steady winding; the equivalent standard deviation σ (Bhushan and Tonder, 1989a; b) for surface roughness of adjacent tape layers; and the Peklenik number γ which describes directionality (if any) of the surface's texture. For the technological applications and parameter values of interest here, the influence of air compressibility is negligible owing to the small layer-to-layer pressure rise, and the effects of inertia, centrifugal forces, and elasticity of the tape layers are likewise judged to be small.

Dimensional analysis provides the functional relationship $\dot{H}^*(\gamma, \sigma^*, b^*, \Pi^*, t^*)$ for the rate of collapse of the cumulative air film at specified θ, and an analogous expression for h^*. Here all length quantities denoted by the asterisk superscript are nondimensionalized with respect to R; the operating parameter is defined $\Pi^* = \mu Vb/T$; and $t^* = tV/R$. For samples of fixed width wound at identical values of Π^*, the dependence simplifies to $\dot{H}^*(\gamma, \sigma^*, t^*)$ as a function of surface characteristics and time. With this representation, experimental and model results are applicable in a consistent set of units to geometrically similar tape or web winding applications within the context of the modeled parameters.

3.1 Backcoat Surface. In the case of hydrodynamic lubrication, surface roughness effects are normally judged to be negligible when $h/\sigma \gg 1$, but they become relatively important when $3 < h/\sigma < 6$. Asperity-to-asperity contacts, on the other hand, begin to occur in the partial lubrication regime $h/\sigma < 3$. In several approaches that have been taken to date, the expected pressure within a rough bearing depends on the statistics of the roughness rather than the surface's pointwise topology. In one such method, stochastic effects are included by spatially averaging the lubrication equations over the bearing's domain prior to their solution (Tonder and Christensen, 1972; Christensen, 1969–1970; Tonder, 1986). A second technique invokes multiple dimensional scales, whereby the (short wavelength) roughness and (comparatively long wavelength) bearing scales are separated such that the film thickness is represented in a perturbation series (Elrod, 1973; Bayada and Faure, 1989). In the third approach, Reynold's equation is averaged over the surface roughness, and pressure and shear correction factors obtained independently through numerical simulation of specific, related, roughness patterns are defined (Patir and Cheng, 1978). In extensions, closed-form expressions for the factors have been obtained through perturbation methods and compared with the simulated values (Tripp, 1983), and the method has been applied to slider and foil squeeze film bearings (Bhushan and Tonder, 1989b). Mapping functions have also been employed to compute lubricant flow as functions of roughness orientation by transforming periodic anisotropic surfaces into isotropic ones (Lo, 1992).

The roughness is described by an $N \times M$ matrix of heights δ_{ij} with prescribed statistical properties (Patir, 1978) over the surface. Each δ_{ij} is the roughness amplitude at position $\theta = (i/R)\Delta l$ and $z = j\Delta l$ within the defined domain, where Δl is the chosen sampling interval. The amplitudes are normally distributed, and the correlation lengths λ_θ in the down and λ_z in the cross-tape directions are defined through the autocorrelation function

$$\mathcal{R}_{uv} = \begin{cases} \sigma^2(1 - u/n)(1 - v/m) & u \le n;\ v \le m \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

such that $\lambda_\theta = n\Delta l$ and $\lambda_z = m\Delta l$ specify the directionality (if any) of the roughness orientation through $\gamma = \lambda_\theta/\lambda_z$. Here $u\Delta l$ and $v\Delta l$ denote the delay lengths in the down and cross-tape directions, respectively. Roughness patterns for specified values of σ, λ_θ, and λ_z are then generated by the linear transformation

$$\delta_{ij} = \frac{\sigma}{\sqrt{mn}} \sum_{k=1}^{n} \sum_{l=1}^{m} \eta_{i+k,j+l} \qquad \begin{matrix} i = 1, 2, \ldots N \\ j = 1, 2, \ldots M \end{matrix} \tag{2}$$

where η_{ij} is normally distributed with zero mean and unit variance.

Simulated profiles of tapes having smooth isotropic and grooved surfaces are shown in Fig. 4. The roughness parameters were extracted directly from surface profiles measured for SI, and for sample G (σ = 100 nm) which had a pattern of grooves oriented in the cross-tape direction as shown in Fig. 4(b). The purposeful introduction of such grooves is expected to facilitate discharge of air entrained within the roll. Those measurements were obtained through white light interferometry over a 0.25 mm^2 microscope field of view with the tape samples placed under the nominal tension of 0.28 N. With that technique, the minimum measurable spatial wavelength was 8 μm as set by resolution and anti-aliasing requirements. Similarly, the selected field of view restricted the maximum measurable wavelength, thus eliminating contributions associated with cross-tape warpage or waviness (Majumdar and Bhushan, 1990). In each case, since the opposing (media) side of the tape was comparatively smooth with roughness below 10 nm, the backcoat values of σ approximate the equivalent roughness of the two bearing surfaces. The model simulates pure transverse striations in the limit $\gamma = 0$, and sensible correlation between the simulated and measured surface profiles was found at the values γ = 1 and 0.008 for the isotropic and transverse orientations, respectively. The generated patterns do not replicate the measured surfaces in a pointwise manner, nor are they intended to do so, but they do represent the roughness amplitude and directionality in the statistical and visual senses.

3.2 Transient Discharge. Surface texture is incorporated in the model through the use of pressure factors which modify the Poiseuille flow components in the down and cross-tape directions. Reynold's equation with flow terms so modified models the transient discharge of air from the tape roll once it is brought to rest from steady winding. For an isothermal incompressible fluid at pressure p (Patir and Cheng, 1978),

$$(\phi_\theta h^3 p_{,\theta})_{,\theta} + R^2(\phi_z h^3 p_{,z})_{,z} = 12\mu R^2 h_{,t} \tag{3}$$

where the factors ϕ_θ and ϕ_z are functions of h, σ, and γ. In the limiting cases of isotropic roughness (Bhushan and Tonder, 1989a),

$$\phi_\theta(h/\sigma, \gamma = 1) = \phi_z(h/\sigma, \gamma = 1) = 1 + (17/20)(\sigma/h)^2, \tag{4}$$

and for transverse roughness (Tripp, 1983) the factors become

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

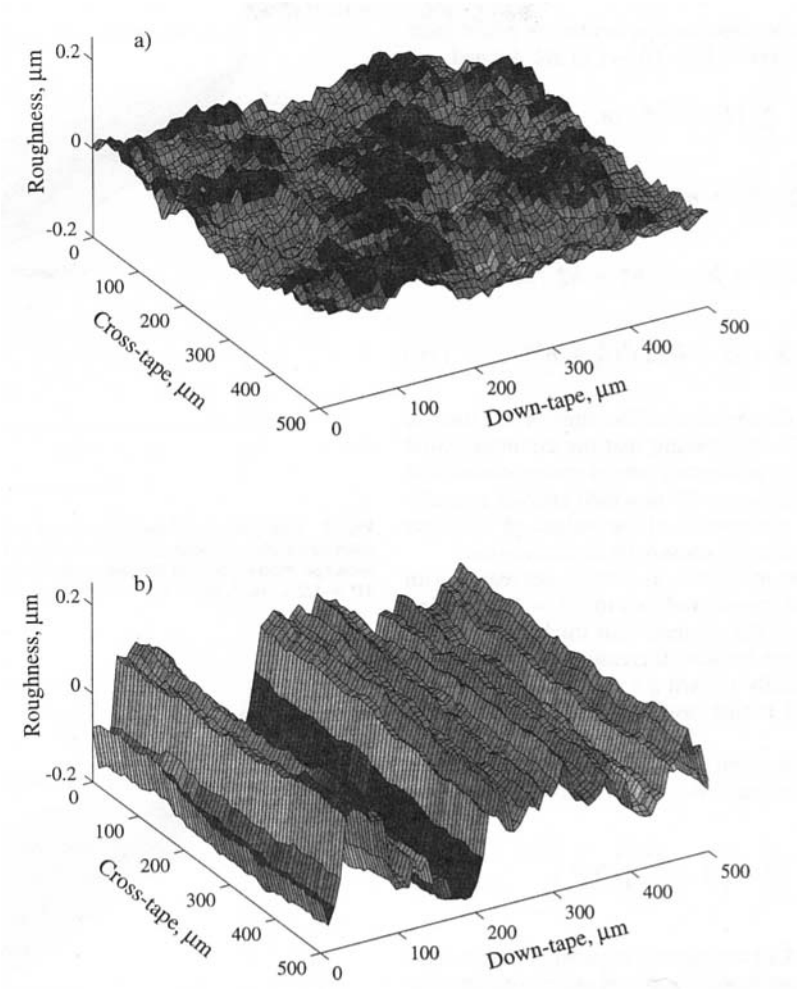


Fig. 4 Simulated backcoat roughness distributions. (*a*) Smooth isotropic surface; σ = 25 nm and γ = 1. (*b*) Grooved surface; σ = 100 nm and γ = 0.008.

$$\phi_\theta(h/\sigma,\ \gamma = 0) = 1 - 6(\sigma/h)^2$$
$$\phi_z(h/\sigma,\ \gamma = 0) = 1 + 3(\sigma/h)^2. \tag{5}$$

In short, transverse grooves enhance flow in the cross-tape direction while impeding down-tape flow, and greater isotropic roughness enhances flow equally in both directions, albeit by smaller amounts.

The pressure in Eq. (3) is expressed $p(\theta, z) = p_\theta(\theta)p_z(z)$ as the product of separable gauge distributions along and across the tape. In the winding direction, the pressure grows linearly by amount T/Rb over each interior wrap through $p_\theta = (\frac{1}{2} + \theta/2\pi)(T/Rb)$. The pressure in the cross-tape direction is approximated by a single-term expansion that satisfies boundary conditions for two approximate limiting models—symmetric or asymmetric side flow—of tape stacking. When the tape layers stack with even clearance between their edges and each guiding flange, the side flow is symmetric about $z = 0$, and the cross-tape pressure distribution is taken as parabolic. With ambient conditions along the two edges $z = \pm b/2$, $p_z = (\frac{3}{2})(1 - (2z/b)^2)$. When tape stacks preferentially against one flange, on the other hand, air discharges from only one edge of the tape layer. Here the distribution remains parabolic but satisfies the ambient condition along one edge and zero gradient along the other. For asymmetric leakage, $p_z = (\frac{9}{8})(1 - 4z/3b - \frac{4}{3}(z/b)^2)$. In each case, the leading coefficients of p_z are determined through the normalization requirement $(1/b)\int_{-b/2}^{b/2} p_z dz = 1$, which ensures that the average gauge pressure over the nth wrap is nT/Rb. Experimental tape transport systems fall between these two limiting cases, and in that sense, the symmetric and asymmetric leakage models are expected to bound measured responses.

The interlayer film thickness is expressed by the expansion $h^*(\theta, t^*) = \hat{h}^*(\theta)q^*(t^*)$ in terms of its piecewise constant representation $\hat{h}^*$ during steady-state winding (Keshavan and Wickert, 1997), and the coordinate q^* that characterizes temporal air discharge. Application of Galerkin's method to minimize the weighted residual error of Eq. (3) over the entire air bearing provides the differential equation $\dot{q}^* + C_1 q^* + C_2 q^{*3} = 0$ governing evolution of q^*, where

$$C_1 = \frac{\sigma^{*2}}{\Pi^* \sum_{n=1}^{N_c} (\hat{h}_n^* + \hat{h}_{n-1}^*)n\pi} \times \begin{cases} (3c_{11}/10\pi) + (3c_{12}/\kappa b^{*2}) & \text{grooved} \\ (-17c_{11}/400\pi) + (17c_{12}/20\kappa b^{*2}) & \text{isotropic} \end{cases} \tag{6}$$

and

$$C_2 = \frac{-3c_{21}/20\pi + c_{22}/\kappa b^{*2}}{\Pi^* \sum_{n=1}^{N_c} (\hat{h}_n^* + \hat{h}_{n-1}^*)n\pi}. \tag{7}$$

Roughness is represented by C_1 alone; the factor $\kappa = 1$ (or 4) for symmetric (or asymmetric) side flow; $\hat{h}_n^*$ is the nondimensional thickness of the nth layer; and N_c is the number of wraps

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

in steady-state before direct tape-to-tape contact is established. The intermediate parameters in Eqs. (6)–(7) are defined

$$c_{11} = \sum_{n=1}^{N_c} (\hat{h}_n^* - \hat{h}_{n-1}^*)n$$

$$c_{12} = \pi \sum_{n=1}^{N_c} (\hat{h}_n^* + \hat{h}_{n-1}^*)(\tfrac{1}{12} + n^2)$$

$$c_{21} = (\tfrac{1}{4}) \sum_{n=1}^{N_c} (\hat{h}_n^* + \hat{h}_{n-1}^*)^2(\hat{h}_n^* - \hat{h}_{n-1}^*)n$$

$$c_{22} = (\pi/4) \sum_{n=1}^{N_c} (\hat{h}_n^* + \hat{h}_{n-1}^*)^3(\tfrac{1}{12} + n^2). \qquad (8)$$

At each instant, H^* is determined as the sum of all interior air film thicknesses while recognizing that the circumferential location at which tape-to-tape contact occurs moves outward in the roll with time. The thickness h_c^* at which contact is established is determined by the greater of the values of backcoat roughness and long wavelength cross-tape deformations.

For a given initial film thickness, as tension decreases with $\Pi^* \to \infty$, the equation of motion reduces to $\dot{q}^* = 0$, implying that no air discharges and the original film thickness persists. If, on the other hand, the tension increases ($\Pi^* \to 0$), the solution tends asymptotically toward $q^* = 0$, except in a very small interval near $t^* = 0$. In that case, the air discharges almost instantaneously.

Subject to the initial condition $q^*(0) = 1$, which corresponds to the case of steady winding followed by shut-down, the response becomes

$$q^*(t^*) = \left(-\frac{C_2}{C_1} + \left(1 + \frac{C_2}{C_1}\right) e^{2C_1 t^*} \right)^{-1/2}. \qquad (9)$$

Typically $C_1/C_2 < 1$, and an asymptotic expansion in that ratio indicates that backcoat roughness at various orders perturbs the solution for smooth tape relative to the effects of such winding variables as speed, tension, and width[1]. In general, the magnitudes of the perturbation corrections are a function of the roughness amplitude and orientation, and they grow with time, becoming more important as the film gets thinner.

4 Discussion

The model's predictions are compared with air discharge rates measured over multiple test runs for the three samples of differing surface morphology described above. In each case, the rate of discharge is determined as the superposition of rates at which individual layers collapse from their respective values during steady winding. Measurements of H^* at steady-state correlated within 15 percent with the predictions of the symmetric side leakage model, and the $\hat{h}_n^*$ so determined establish the initial film thickness distribution for the transient discharge model. The predicted H^* histories for the cases of symmetric and asymmetric side flow, and the multiple measured traces, are set to the common ordinate value $H^* = 2.8 \times 10^{-4}$ ($H = 5\ \mu$m) at $t^* = 0$. Time-shifting in this manner ensures a consistent basis for comparing the discharge rates in the light of their nonlinear and non-self-similar nature.

Results for smooth isotropic and transversely grooved tape samples are shown in Figs. 5 and 6. In each case, the predictions of the symmetric and asymmetric side flow models bound the measured profiles (four runs in Fig. 5 and six in Fig. 6). The measured H^* values at steady-state for SI (16 μm) and G (18 μm) were within 12 percent of each other, so that $h_c^* \approx 1.1 \times 10^{-5}$ for each sample, with $\sigma^*/h_c^* = 0.12$ and 0.5 for SI and G, respectively. Sample G however, was substantially more effective at discharging air from the roll to the degree that H^* decreased by some 78 percent in Fig. 6 over the time span considered, whereas the smaller 55 percent decrease occurred for SI in Fig. 5. In fact, in Fig. 5, the abscissa must be extended to $t^* = 500$ before H^* falls to the value reached by G at $t^* = 250$, a factor of two difference in the settling time. Figure 7 presents analogous results for RI ($h_c^* = 3.3 \times 10^{-5}$, $\sigma^*/h_c^* = 0.5$). In this case, H^* decayed even faster than for G with a 90 percent decrease over $0 \le t^* \le 250$. Note, however, that surface RI was three times as rough as G, suggesting additional wear and reliability problems should it be used in practice.

Model predictions of the discharge rates for case studies in tension, width, and root-mean-square roughness amplitude are shown in Figs. 8–10 for the case of symmetric side leakage. Over four values of Π^* that are typical of tape transport systems, H^* histories are shown in Fig. 8. If V is held constant in this representation, tension increases by a factor of six. With increasing Π^*, H^* grows at $t^* = 0$ by some 53 percent as Π^* is increased from 1.2×10^{-6} to 2.4×10^{-6} and by a comparable value (58 percent) when Π^* grows from 0.6×10^{-6} to 1.2×10^{-6}. Owing to the greater thickness of the cumulative entrained

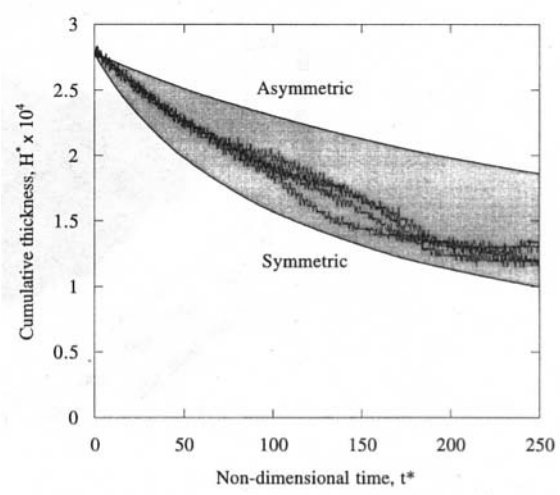


Fig. 5 Comparison of measured and predicted air discharge rates for samples with a smooth isotropic surface. The symmetric and asymmetric leakage models bound the measured traces (shaded region); $b^* = 0.44$, $\Pi^* = 1.2 \times 10^{-6}$, $h_c^* = 1.1 \times 10^{-5}$, and $\sigma^* = 1.4 \times 10^{-6}$.

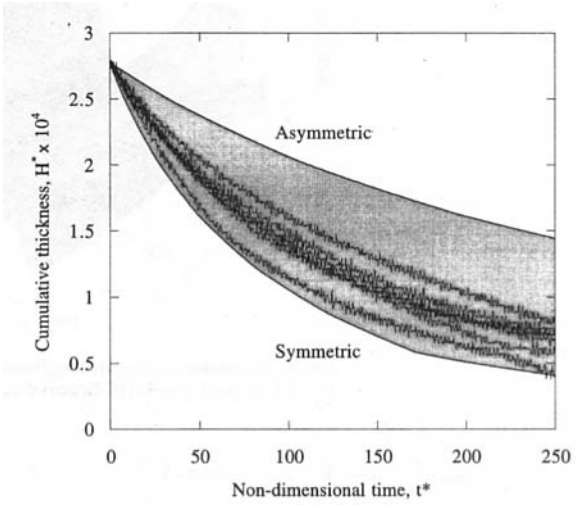


Fig. 6 Comparison between measured and predicted air discharge rates for samples with a grooved surface; $b^* = 0.44$, $\Pi^* = 1.2 \times 10^{-6}$, $h_c^* = 1.1 \times 10^{-5}$, and $\sigma^* = 5.6 \times 10^{-6}$

[1] For sample G of width $b = 8$ mm wound at $T = 0.28$ N and $V = 2.3$ m/s, for instance, $C_1/C_2 = 0.11$ with symmetric side flow.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

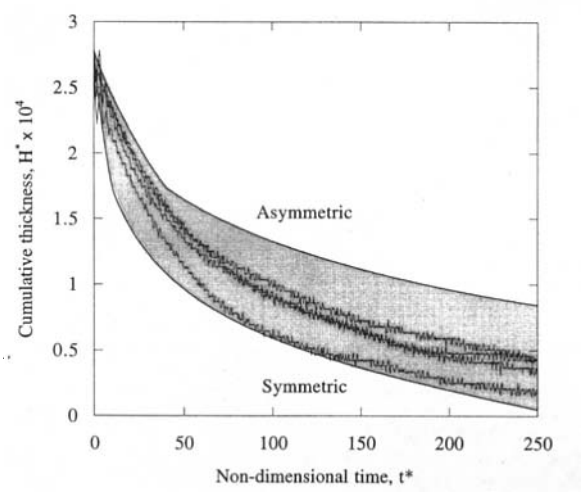


Fig. 7 Comparison between measured and predicted air discharge rates for samples with a rough isotropic surface; $b^* = 0.44$, $\Pi^* = 1.2 \times 10^{-6}$, $h_c^* = 3.3 \times 10^{-5}$, and $\sigma^* = 1.7 \times 10^{-5}$

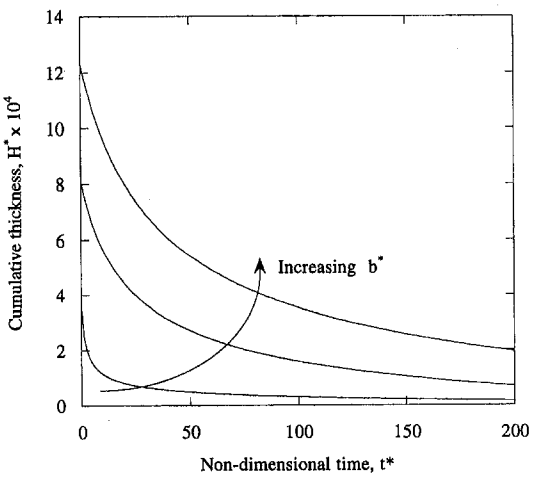


Fig. 9 Predicted contribution of b^* to the entrained air film and its discharge rate from steady-state for a grooved surface; $b^* = 0.22$, 0.44, and 0.66, $\Pi^* = 1.2 \times 10^{-6}$, $h_c^* = 1.1 \times 10^{-5}$, and $\sigma^* = 5.6 \times 10^{-6}$

air film in the steady-state, the discharge times likewise increase with Π^*.

Figure 9 shows the effect of width ratio on H^* for fixed Π^*. In this case, a tripling of b^* essentially triples H^* at $t = 0$, with the discharge rate being enhanced for smaller b^*. In short, the wider the roll of wound tape, the longer it will take for air to discharge from it and for the roll to stabilize. The effect of roughness amplitude on $\dot{H}^*$ is highlighted in Fig. 10. Results with isotropic and grooved surfaces are compared, and in each case for fixed h_c^*, the discharge rate grows as σ^* increases. In this case, grooves increase the discharge rate above that of an isotropic surface to the extent that at $\sigma^* = 0.28 \times 10^{-5}$, sample G discharged the entrained air at nearly the same rate as an isotropic surface having twice the root-mean-square surface roughness.

5 Summary

The derived transient model predicts the rate at which air discharges from a wound roll following steady winding. The rates of air discharge for three samples of differing roughness properties—smooth isotropic, rough isotropic, and grooved—were measured and compared with the model's predictions. Although the smooth isotropic and grooved samples had nearly the same cumulative air film thickness during steady winding, the rate at which air discharged from a roll of the textured material was typically some 20–25 percent greater. In short, the presence of grooves or channels enhances air flow in the cross-tape direction, tending to stabilize the roll more rapidly. The model and experimental technique establish a quantitative procedure to evaluate media having different roughness statistics from the perspective of providing candidate solutions to problems of air entrainment.

Tension, width, and velocity have the greatest effects in setting the cumulative film thickness at steady-state, namely the initial condition for transient air discharge. Each of those parameters has a first-order influence on the time required for the roll to stabilize. For the same root-mean-square roughness amplitude, however, a transversely grooved surface offers a measurable, but incremental, advantage over an isotropic one. For circumstances under which reductions in tension, width, or operating speed are not feasible, not to mention mechanical contact with a preload mechanism, texturing can offer a second-order contribution to enhancing the discharge rate. Of course, similar enhancement can be realized by a rougher isotropic surface, but only at the expense of heightened wear, imprinting, and reliability problems.

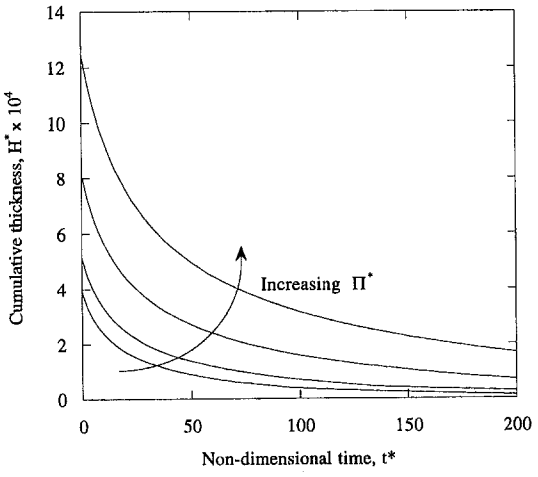


Fig. 8 Predicted contribution of Π^* to the entrained air film and its discharge rate from steady-state for a grooved surface; $b^* = 0.44$, $\Pi^* =$ (2.4, 1.2, 0.6, and 0.4) $\times 10^{-6}$, $h_c^* = 1.1 \times 10^{-5}$, and $\sigma^* = 5.6 \times 10^{-6}$

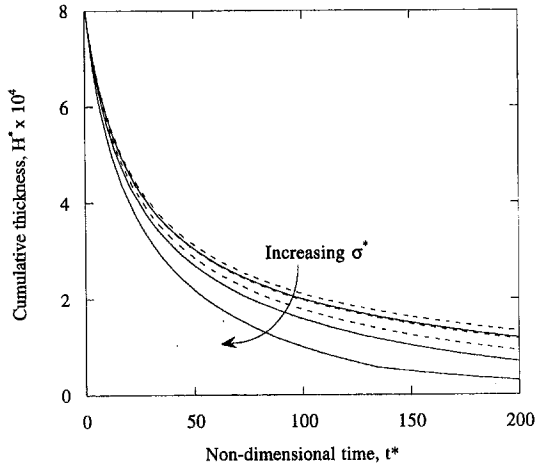


Fig. 10 Predicted contribution of root-mean-square roughness to the rate of air discharge from steady-state for grooved (—) and isotropic (– – –) surfaces; $b^* = 0.44$, $\Pi^* = 1.2 \times 10^{-6}$, and $h_c^* = 1.1 \times 10^{-5}$, $\sigma^* =$ (2.8, 5.6, and 8.3) $\times 10^{-6}$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Acknowledgment

This work is supported in part by the National Science Foundation and the Data Storage Systems Center under grant numbers ECD-8907068 and DMI-9622258, and by Imation Corporation. The authors acknowledge with gratitude the assistance of Dr. Mohammed Siddiq and Dr. Durkee Richards of Imation Corporation for providing experimental samples and for offering suggestions with this manuscript.

References

Bayada, G., and Faure, J. B., 1989, "A Double Scale Analysis Approach of the Reynold's Roughness: Comments and Application to the Journal Bearing," ASME *Journal of Tribology,* Vol. 111, pp. 323–330.

Bhushan, B., and Tonder, K., 1989a, "Roughness-Induced Shear- and Squeeze-Film Effects in Magnetic Recording—Part I: Analysis," ASME *Journal of Tribology,* Vol. 111, pp. 220–227.

Bhushan, B., and Tonder, K., 1989b, "Roughness-Induced Shear- and Squeeze-Film Effects in Magnetic Recording—Part II: Applications," ASME *Journal of Tribology,* Vol. 111, pp. 228–237.

Christensen, H., 1969–70, "Stochastic Models for Hydrodynamic Lubrication of Rough Surfaces," *Proceedings of the Institution of Mechanical Engineers,* Vol. 55, pp. 1013–1026.

Ducotey, K. S., and Good, J. K., 1995, "The Importance of Traction in Web Handling," ASME *Journal of Tribology,* Vol. 117, pp. 679–684.

Elrod, H. G., 1973, "Thin-Film Lubrication Theory for Newtonian Fluids With Surfaces Possessing Striated Roughness or Grooving," ASME *Journal of Lubrication Technology,* Vol. 95, pp. 484–489.

Jones, D. P., 1992, "Air Entrainment as a Mechanism for Low Traction on Rollers and Poor Stacking of Polyester Film Reels, and Its Reduction," *Web Handling,* AMD-Vol. 149, ASME, New York, pp. 123–131.

Keshavan, M. B., and Wickert, J. A., 1997, "Air Entrainment During Steady-State Web Winding," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 916–922.

Lo, Sy-Wei, 1992, "On the Effects of Roughness Orientation: A Mapping and Double Scale Analysis of Flow Factors," ASME *Journal of Tribology,* Vol. 114, pp. 747–754.

Majumdar, A., and Bhushan, B., 1990, "Role of Fractal Geometry in Roughness Characterization and Contact Mechanics of Surfaces," ASME *Journal of Tribology,* Vol. 112, pp. 205–216.

Patir, N., 1978, "A Numerical Procedure for Random Generation of Rough Surfaces," *Wear,* Vol. 47, pp. 263–277.

Patir, N., and Cheng, H. S., 1978, "An Average Flow Model for Determining Effects of Three-Dimensional Roughness on Partial Hydrodynamic Lubrication," ASME *Journal of Lubrication Technology,* Vol. 100, pp. 12–17.

Smith, D. P., and Von Behren, R. A., 1989, "Squeeze-Film Analysis of Tape Winding Effects in Data Cartridge," *Tribology and Mechanics of Magnetic Storage Systems,* ASME/STLE Conference Proceedings, Vol. 6, pp. 88–92.

Tonder, K., and Christensen, H., 1972, "Waviness and Roughness in Hydrodynamic Lubrication," *Proceedings of the Institution of Mechanical Engineers,* Vol. 186, pp. 807–812.

Tonder, K., 1986, "The Lubrication of Unidirectional Striated Roughness: Consequences for Some General Roughness Theories," ASME *Journal of Tribology,* Vol. 108, pp. 167–170.

Tripp, J. H., 1983, "Surface Roughness Effects in Hydrodynamic Lubrication: The Flow Factor Method," ASME *Journal of Lubrication Technology,* Vol. 105, pp. 458–465.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

F. Giambanco[1]
Associate Professor,
Mem. ASME

L. Palizzolo
Assistant Professor

L. Cirone
Graduate Student

Dipartimento di Ingegneria Strutturale
& Geotechnica, DISEG,
Università di Palermo,
Viale delle Scienze,
90128 Palermo, Italy

An Iterative Approach to Dynamic Elastic-Plastic Analysis

The step-by-step analysis of structures constituted by elastic-plastic finite elements, subjected to an assigned loading history, is here considered. The structure may possess dynamic and/or not dynamic degrees-of-freedom. As it is well-known, at each step of analysis the solution of a linear complementarity problem is required. An iterative method devoted to solving the relevant linear complementarity problem is presented. It is based on the recursive solution of a linear complementarity problem in which the constraint matrix is block-diagonal and deduced from the matrix of the original linear complementarity problem. The convergence of the procedure is also proved. Some particular cases are examined. Several numerical applications conclude the paper.

1 Introduction

Stepwise analysis is certainly the most complete and conclusive computational tool in the investigation of elastic-plastic behavior of structures subjected to a given loading history. It is founded on the space discretization of the structure and on the time discretization of the given loading history and of the unknown plastic deformation history. Such kind of analysis has been first utilized in studying structures subjected to quasi-static loading (see, e.g., Maier, 1968a, b; Capurso and Maier, 1970; Martin, 1975; De Donato and Maier, 1976; Maier, 1978; De Donato and Franchi, 1979; Martin, 1985) and it has been successively extended to structures subjected to dynamic loading (see, e.g., Wilson et al., 1973; Owen, 1980; Owen and Hilton, 1980; Polizzotto, 1986; Muscolino, 1989). In both cases, on the grounds of the adopted discretizations, at each analysis step, the solving set is represented by a linear complementarity algebraical problem (LCP) whose solution can be achieved by solving the same LCP or by solving a quadratic programming equivalent problem (QPP) (see, e.g., Cottle and Dantzig, 1968; Lemke, 1968; Cottle, 1979; Lemke, 1980; Cottle et al., 1992).

Several efficient algorithms devoted to the QPP solution already exist (see, e.g., Gill et al., 1984), but, unfortunately, in executing elastic-plastic analysis the dimensions of the QPPs are sometimes so considerable to create serious computational difficulties. The so-called element optimization technique (EOT) (see, e.g., Feijòo and Zouain, 1987; Borino et al., 1987; Borino and Polizzotto, 1989), which iteratively obtains the solution satisfying element by element the yield conditions, has been used to overcome these difficulties. Some approximate techniques (see, e.g., Zarka and Casier, 1981; Martin, 1985; Giambanco et al., 1994; Pyo et al., 1995; Dvorkin, 1996; Kojic, 1997; Wang et al., 1997) have been used, as well.

In the present paper a quite general formulation of the elastic-plastic analysis of classically damped discrete structures is first developed, taking into account the possible contemporaneous presence of dynamic and not dynamic degrees-of-freedom. Successively, a special iterative method aimed at solving even very large LCPs is presented. It is based on the recursive solution of a sequence of LCPs in which the constraint matrix is block-diagonal and it is deduced by the matrix of the original LCP. The technique shows some features of versatility and it can be applied in different ways: the available QPP routines can be recursively utilized solving a sequence of QPPs with orders smaller than the original one; on the grounds of very simple recursive formulae (identifying with a special case of the proposed technique) the original LCP can be directly solved. The convergence of the procedure is finally proved.

Several numerical experiences show the effectiveness of the method and provide useful information on the possible suitable choices of some check parameters as the prefixed tolerance and the maximum number of iterations.

Notation. Boldface symbols are vectors or matrices, symbols as $(\bullet)^T$ or $(\tilde{\bullet})$ denote the transpose of the relevant quantities, the symbol $I(m)$ indicates the set of integers $1, 2, \ldots, m$.

2 Finite Element Analysis of Elastic Plastic Structures

Let us consider a structure constituted by elastic linearly hardening material, discretized into n_e finite elements. We assume that for each element plastic deformations can occur just at the plastic nodes (i.e., Gauss points), which are conceived as sources from which plastic strains spread within the element volume, according with a fixed shape function (see, e.g., Corradi, 1978, 1983). Generally, plastic nodes and element nodes are not coincident. The elastic-plastic behavior of the typical finite element is described in terms of generalized variables (generalized stresses and strains). By hypothesis, the elastic domain of the finite element is an hyperpolyhedron and the hardening features are so characterized that it can expand or shrink homotetically and also translate; as a consequence, each face of the hyperpolyhedron can move, but remaining parallel to itself. The structure possesses n_d dynamic degrees-of-freedom (DDF) and n_s not dynamic degrees of freedom (SDF). In the hypothesis of node concentrated lumped masses, of classically damped structure and small displacements, the elastic-plastic structure behavior is described by the following equations:

$$\mathbf{u}_d = {}^0\mathbf{u}_d, \quad \dot{\mathbf{u}}_d = {}^0\dot{\mathbf{u}}_d, \quad \mathbf{p} = {}^0\mathbf{p} = \mathbf{N}^0\boldsymbol{\lambda}, \quad (t = 0), \quad (1)$$

$$\mathbf{M}_{11}\ddot{\mathbf{u}}_d + \mathbf{V}_{11}\dot{\mathbf{u}}_d + \mathbf{K}_{11}\mathbf{u}_d + \mathbf{K}_{12}\mathbf{u}_s - \mathbf{B}_1\mathbf{p} = \mathbf{F}_d^*, \quad (0 \le t \le {}^1t), \quad (2a)$$

$$\mathbf{K}_{21}\mathbf{u}_d + \mathbf{K}_{22}\mathbf{u}_s - \mathbf{B}_2\mathbf{p} = \mathbf{F}_s^*, \quad (0 \le t \le {}^1t), \quad (2b)$$

[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.

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.

Manuscript received by the ASME Applied Mechanics Division, Oct. 28, 1997; final revision, Feb. 25, 1998. Associate Technical Editor: M. Ortiz.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\dot{\mathbf{p}} = \mathbf{N}\dot{\boldsymbol{\lambda}}, \quad \mathbf{p} = {}^0\mathbf{p} + \int_0^t \dot{\mathbf{p}}(\bar{t})d\bar{t},$$

$$\boldsymbol{\lambda} = {}^0\boldsymbol{\lambda} + \int_0^t \dot{\boldsymbol{\lambda}}(\bar{t})d\bar{t}, \quad (0 \le t \le {}^lt), \quad (3a\text{–}c)$$

$$\mathbf{P} = \tilde{\mathbf{B}}_1\mathbf{u}_d + \tilde{\mathbf{B}}_2\mathbf{u}_s - \mathbf{D}_p\mathbf{p} + \mathbf{P}^*, \quad (0 \le t \le {}^lt), \quad (4)$$

$$-\boldsymbol{\varphi} = -\tilde{\mathbf{N}}\mathbf{P} + \mathbf{H}\boldsymbol{\lambda} + \mathbf{R}, \quad (0 \le t \le {}^lt), \quad (5a)$$

$$-\boldsymbol{\varphi} \ge \mathbf{0}, \quad \dot{\boldsymbol{\lambda}} \ge \mathbf{0}, \quad \tilde{\boldsymbol{\varphi}}\dot{\boldsymbol{\lambda}} = 0,$$

$$\tilde{\dot{\boldsymbol{\varphi}}}\dot{\boldsymbol{\lambda}} = 0, \quad (0 \le t \le {}^lt). \quad (5b\text{–}e)$$

In Eqs. (1)–(5) t is the time variable, defined in the interval $(0, {}^lt)$; vectors $\mathbf{u}_d(t)$ and $\mathbf{u}_s(t)$ are the node displacements corresponding to the n_d DDF and to the n_s SDF, respectively; vectors $\mathbf{F}_d^*(t)$ and $\mathbf{F}_s^*(t)$ are the equivalent nodal forces of the structure, sum of the concentrated load applied upon the structure nodes and the nodal forces equivalent to the actions (mechanical and/or kinematical actions) applied upon the elements, corresponding to the DDF and to the SDF, respectively; $\dot{\boldsymbol{\lambda}}(t)$ is the plastic activation vector; $\mathbf{p}(t)$ is the generalized plastic strain vector evaluated at the plastic nodes; $\mathbf{P}(t)$ is the generalized stress response vector evaluated at the plastic nodes (being $\mathbf{P}^*(t)$ the contribution to $\mathbf{P}(t)$ due to the mechanical and/or kinematical load acting upon the elements in absence of nodal displacements and plastic deformations); $\boldsymbol{\varphi}(t)$ is the yield function vector (having also the role of plastic potential vector); $\mathbf{R}$ is the time-independent plastic resistance vector; $\mathbf{M}_{11}$ and $\mathbf{V}_{11}$ are the mass and damping matrices, related to DDF; $\mathbf{K}_{\alpha\beta}$ ($\alpha = 1, 2$; $\beta = 1, 2$) are the blocks of the stiffness matrix $\mathbf{K} = \tilde{\mathbf{C}}\mathbf{D}\mathbf{C}$ corresponding to the DDF and SDF, being $\mathbf{C}$ the compatibility matrix and $\mathbf{D}$ the block diagonal internal stiffness matrix; $\mathbf{B}_1$ and $\mathbf{B}_2$ are the blocks corresponding to the DDF and SDF of the so-called pseudoforce matrix $\mathbf{B} = \tilde{\mathbf{C}}\mathbf{D}\mathbf{W}$, which transforms plastic strains into nodal equivalent forces, being $\mathbf{W}$ a matrix which, applied to plastic strains, provides nodal relative displacements of the elements; $\mathbf{D}_p = \tilde{\mathbf{W}}\mathbf{D}\mathbf{W}$ is the block diagonal stiffness matrix of the elements related to the plastic nodes; $\mathbf{N}$ is the block diagonal matrix of unit external normals to the yield surface; $\mathbf{H}$ is the work-hardening matrix, by hypothesis constant, symmetric, and positive-definite or positive-semi-definite. Let n be the order of vectors $\dot{\boldsymbol{\lambda}}$, $\boldsymbol{\lambda}$, $\dot{\boldsymbol{\varphi}}$, $\boldsymbol{\varphi}$, and $\mathbf{R}$.

Equations (1) are the initial conditions, to be satisfied for $t = 0$, while Eqs. (2)–(5) must be satisfied for every $t \in (0, {}^lt)$. In the particular case in which $n_d = 0$, the time variable t does not necessarily have the meaning of physical time: it can merely be a monotonically increasing variable for the specification of the loading sequence.

Utilizing (2b), displacements $\mathbf{u}_s$ can be easily expressed in terms of $\mathbf{u}_d$ and $\mathbf{p}$:

$$\mathbf{u}_s = \mathbf{K}_{22}^{-1}\mathbf{F}_s^* - \mathbf{K}_{22}^{-1}\mathbf{K}_{21}\mathbf{u}_d + \mathbf{K}_{22}^{-1}\mathbf{B}_2\mathbf{p}, \quad (0 \le t \le {}^lt). \quad (6)$$

Setting

$$\mathbf{M}_d = \mathbf{M}_{11}, \quad \mathbf{V}_d = \mathbf{V}_{11}, \quad \mathbf{K}_d = \mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}, \quad (7a\text{–}c)$$

$$\mathbf{B}_d = \mathbf{B}_1 - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{B}_2, \quad \mathbf{F}_d = \mathbf{F}_d^* - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{F}_s^*, \quad (7d\text{–}e)$$

and substituting (6) in (2a) and (4) one obtains

$$\mathbf{M}_d\ddot{\mathbf{u}}_d + \mathbf{V}_d\dot{\mathbf{u}}_d + \mathbf{K}_d\mathbf{u}_d - \mathbf{B}_d\mathbf{p} = \mathbf{F}_d, \quad (0 \le t \le {}^lt), \quad (8)$$

$$\mathbf{P} = \tilde{\mathbf{B}}_2\mathbf{K}_{22}^{-1}\mathbf{F}_s^* + \tilde{\mathbf{B}}_d\mathbf{u}_d + (\tilde{\mathbf{B}}_2\mathbf{K}_{22}^{-1}\mathbf{B}_2 - \mathbf{D}_p)\mathbf{p} + \mathbf{P}^*, \quad (0 \le t \le {}^lt). \quad (9)$$

Equation (8) allows us to express displacements $\mathbf{u}_d$ in terms of known forces $\mathbf{F}_d$ and unknown plastic deformations $\mathbf{p}$, utilizing the superimposition of natural vibration modes (see, e.g., Polizzotto, 1986; Muscolino, 1989). This method implies a previous modal analysis of the structure, necessary to obtain the essential dynamic features of the structure, as the displacement modes $\boldsymbol{\Phi}_i$ (normalized in such a way that $\tilde{\boldsymbol{\Phi}}_i\mathbf{M}_d\boldsymbol{\Phi}_i = 1$), the natural frequencies ω_i, the damping ratios ξ_i, and the reduced frequencies $\omega_i^D = \omega_i\sqrt{1 - \xi_i^2}$, for $i = 1, 2, \ldots, n_d$. The knowledge of these quantities allows us to construct some dynamic operators by which the elastic response of the structure can be determined $\forall t \in (0, {}^lt)$. These operators are (see, e.g., Polizzotto, 1986; Muscolino, 1989) the modal matrix $\boldsymbol{\Phi}$, the matrix $\boldsymbol{\Theta}$, the dynamic correction matrix $\mathbf{G}$ and the matrix $\mathbf{A}$, i.e.,

$$\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1 \quad \boldsymbol{\Phi}_2 \quad \ldots \quad \boldsymbol{\Phi}_i \quad \ldots \quad \boldsymbol{\Phi}_{n_d}], \quad (10a)$$

$$\Theta_i = \frac{1}{\omega_i^2}\exp(-\xi_i\omega_i t)\left[\cos\omega_i^D t + \frac{\xi_i}{\sqrt{1-\xi_i^2}}\sin\omega_i^D t\right], \quad i = 1, 2, \ldots, n_d, \quad (10b)$$

$$\boldsymbol{\Theta}(t) = \text{diag}\,(\Theta_1, \Theta_2, \ldots, \Theta_i, \ldots, \Theta_{n_d}), \quad (10c)$$

$$\mathbf{G}(t) = \boldsymbol{\Phi}\boldsymbol{\Theta}(t)\tilde{\boldsymbol{\Phi}}, \quad (10d)$$

$$\mathbf{A}(t) = \mathbf{K}_d^{-1} - \mathbf{G}(t). \quad (10e)$$

Setting ${}^0\mathbf{F}_d = \mathbf{F}_d(0)$, the elastic response of the structure, $\forall t \in (0, {}^lt)$, can be expressed as follows (see, e.g., Polizzotto, 1986; Muscolino, 1989):

$$\mathbf{u}_d = \mathbf{G}(t)\mathbf{K}_d^0\mathbf{u}_d - \dot{\mathbf{G}}(t)\mathbf{M}_d^0\dot{\mathbf{u}}_d + \mathbf{A}(t){}^0\mathbf{F}_d + \int_0^t \mathbf{A}(t-\bar{t})\dot{\mathbf{F}}_d(\bar{t})d\bar{t} + \mathbf{A}(t)\mathbf{B}_d^0\mathbf{p} + \int_0^t \mathbf{A}(t-\bar{t})\mathbf{B}_d\dot{\mathbf{p}}(\bar{t})d\bar{t}. \quad (11)$$

Substituting (11) in (9), substituting (9) so modified in (5a), and taking into account (3a) one obtains the solving set, to be satisfied $\forall t \in (0, {}^lt)$,

$$-\boldsymbol{\varphi} = \mathbf{R} - \tilde{\mathbf{N}}\mathbf{P}^*(t) - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_2\mathbf{K}_{22}^{-1}\mathbf{F}_s^*(t) - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_d\mathbf{G}(t)\mathbf{K}_d^0\mathbf{u}_d + \tilde{\mathbf{N}}\tilde{\mathbf{B}}_d\dot{\mathbf{G}}(t)\mathbf{M}_d^0\dot{\mathbf{u}}_d - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_d\mathbf{A}(t){}^0\mathbf{F}_d - \int_0^t \tilde{\mathbf{N}}\tilde{\mathbf{B}}_d\mathbf{A}(t-\bar{t})\dot{\mathbf{F}}_d(\bar{t})d\bar{t} + (\mathbf{H} + \tilde{\mathbf{N}}\mathbf{D}_p\mathbf{N} - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_2\mathbf{K}_{22}^{-1}\mathbf{B}_2\mathbf{N} - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_d\mathbf{A}(t)\mathbf{B}_d\mathbf{N}){}^0\boldsymbol{\lambda} + \int_0^t (\mathbf{H} + \tilde{\mathbf{N}}\mathbf{D}_p\mathbf{N} - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_2\mathbf{K}_{22}^{-1}\mathbf{B}_2\mathbf{N} - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_d\mathbf{A}(t-\bar{t})\mathbf{B}_d\mathbf{N})\dot{\boldsymbol{\lambda}}(\bar{t})d\bar{t}, \quad (12a)$$

$$-\boldsymbol{\varphi} \ge \mathbf{0}, \quad \dot{\boldsymbol{\lambda}} \ge \mathbf{0}, \quad \tilde{\boldsymbol{\varphi}}\dot{\boldsymbol{\lambda}} = 0, \quad \tilde{\dot{\boldsymbol{\varphi}}}\dot{\boldsymbol{\lambda}} = 0, \quad (12b\text{–}e)$$

where just the independent unknown $\boldsymbol{\varphi}$ and $\dot{\boldsymbol{\lambda}}$ appear. If all the degrees-of-freedom are DDF and/or $\mathbf{H}$ is positive-definite, the solution of (12) exists and is unique, otherwise neither the existence nor its uniqueness is guaranteed.

When the load history and the initial conditions are known, it is possible, at least in theory, to obtain the exact elastic plastic response of the structure (if it exists) integrating (12) and then substituting $\dot{\boldsymbol{\lambda}}(t)$ in (2)–(4). But this is a very hard task.

In practice, in order to obtain the solution, it is necessary to discretize the problem also with respect to the time, subdividing the time axis into l time subintervals, usually chosen with the same width δt (sufficiently small with respect to the smallest natural period of the structure), so that ${}^0t = 0$, ${}^1t = \delta t$, ${}^2t = 2\delta t, \ldots, {}^st = s\delta t, \ldots, {}^lt = l\delta t$ are the discretization instants.

During the typical sth subinterval, displacements ${}^{s-1}\mathbf{u} = \mathbf{u}({}^{s-1}t)$, velocities ${}^{s-1}\dot{\mathbf{u}} = \dot{\mathbf{u}}({}^{s-1}t)$ and plastic strains ${}^{s-1}\mathbf{p} = \mathbf{p}({}^{s-1}t)$, evaluated at the end of the $(s-1)$th subinterval, are assumed as initial conditions; the unknown time function $\dot{\boldsymbol{\lambda}}(\tau)$ is modeled so as to be expressed in terms of a time-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

independent unknown vector ${}^s\boldsymbol{\lambda}$, $\forall\tau = (t - {}^{s-1}t) \in (0, \delta t)$, i.e.:

$$\dot{\boldsymbol{\lambda}}(\tau) = \mathbf{g}(\tau){}^s\boldsymbol{\lambda}, \quad (0 \le \tau \le \delta t), \quad {}^s\boldsymbol{\lambda} \ge \mathbf{0}, \quad (s = 1, 2, \ldots, l), \quad (13a, b)$$

$$g_{ij}(\tau) \ge 0, \quad (i, j = 1, \ldots, n; 0 \le \tau \le \delta t), \quad \int_0^{\delta t} \mathbf{g}(\tau)d\tau = \mathbf{I}, \quad (13c, d)$$

being $\mathbf{g}(\tau)$ a suitably chosen square dimensional shape matrix and $\mathbf{I}$ the identity matrix.

Due to modeling (13), elastic unloading can occur just at the discretization instants, every element will remain either in the elastic regime or in the elastoplastic one during each step, and (12e) becomes meaningless; Eqs. (12b) and (12d) cannot be satisfied at some instant $\tau \in (0, \delta t)$; however, they must be satisfied in an integrated, holonomic form, i.e.,

$$-{}^s\boldsymbol{\varphi} = -\int_0^{\delta t} \tilde{\mathbf{g}}(\tau)\boldsymbol{\varphi}(\tau)d\tau = \mathbf{S}{}^s\boldsymbol{\lambda} + {}^s\mathbf{b}, \quad (14a)$$

$$-{}^s\boldsymbol{\varphi} \ge \mathbf{0}, \quad {}^s\boldsymbol{\lambda} \ge \mathbf{0}, \quad {}^s\tilde{\boldsymbol{\varphi}}{}^s\boldsymbol{\lambda} = 0, \quad (s = 1, 2, \ldots, l). \quad (14b\text{–}d)$$

The known term ${}^s\mathbf{b}$ depends on the initial conditions, on the forces of the current step, and on the plastic deformations accumulated at the end of step $s - 1$; it also depends on $\mathbf{g}$ and on the width δt of the time step. Matrix $\mathbf{S}$ depends on the material elastic-plastic property, on the structure geometry, on $\mathbf{g}$ and δt; if $\delta t = {}^lt/l$, $(s = 1, 2, \ldots, l)$, is assumed, matrix $\mathbf{S}$ is constant $(s = 1, 2, \ldots, l)$. $\mathbf{S}$ and ${}^s\mathbf{b}$ can be easily obtained substituting (13a) in (12a), substituting (12a) so modified in (14a) and taking into account that the total plastic activation intensity vector accumulated at the end of the $(s - 1)$th step takes on the form $\sum_{k=0}^{s-1} {}^k\boldsymbol{\lambda}$, so it results in

$${}^{s-1}\mathbf{p} = \mathbf{N}\sum_{k=0}^{s-1} {}^k\boldsymbol{\lambda}. \quad (15)$$

Although it is not strictly necessary, the load history is often modeled in the typical sth time subinterval utilizing a shape function $\mathbf{f}(\tau)$:

$$\dot{\mathbf{F}}_d = \mathbf{f}(\tau){}^s\delta\mathbf{F}_d, \quad \dot{\mathbf{F}}_s^* = \mathbf{f}(\tau){}^s\delta\mathbf{F}_s^*, \quad \dot{\mathbf{P}}^* = \mathbf{f}(\tau){}^s\delta\mathbf{P}^*, \quad (0 \le \tau \le \delta t), \quad (s = 1, 2, \ldots, l). \quad (16a\text{–}c)$$

A stepwise constant shape for $\dot{\boldsymbol{\lambda}}$ is very often chosen, i.e., $\mathbf{g}(\tau) = \mathbf{g} = \mathbf{I}(1/\delta t)$, and, in addition, $\mathbf{f}(\tau) = \mathbf{g}(\tau)$ is assumed. In such a case, it results in

$$\begin{aligned}{}^s\mathbf{b} = {} & \mathbf{R} - \tilde{\mathbf{N}}({}^{s-1}\mathbf{P}^* + \tilde{\mathbf{B}}_2\mathbf{K}_{22}^{-1}{}^{s-1}\mathbf{F}_s^*) \\ & - \frac{1}{2}\tilde{\mathbf{N}}({}^s\delta\mathbf{P}^* + \tilde{\mathbf{B}}_2\mathbf{K}_{22}^{-1}{}^s\delta\mathbf{F}_s^*) - [\tilde{\mathbf{N}}\tilde{\mathbf{B}}_d(\mathbf{K}_d^{-1} - \overset{+}{\mathbf{G}})]^{s-1}\mathbf{F}_d \\ & - \left[\tilde{\mathbf{N}}\tilde{\mathbf{B}}_d\left(\frac{1}{2}\mathbf{K}_d^{-1} - \overset{++}{\mathbf{G}}\right)\right]{}^s\delta\mathbf{F}_d - (\tilde{\mathbf{N}}\tilde{\mathbf{B}}_d\overset{+}{\mathbf{G}}\mathbf{K}_d)^{s-1}\mathbf{u}_d \\ & - \left[\tilde{\mathbf{N}}\tilde{\mathbf{B}}_d\frac{1}{\delta t}(\mathbf{K}_d^{-1} - \mathbf{G})\mathbf{M}_d\right]^{s-1}\dot{\mathbf{u}}_d \\ & + [\mathbf{H} + \tilde{\mathbf{N}}\mathbf{D}_p\mathbf{N} \\ & - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_2\mathbf{K}_{22}^{-1}\mathbf{B}_2\mathbf{N} - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_d(\mathbf{K}_d^{-1} - \overset{+}{\mathbf{G}})\mathbf{B}_d\mathbf{N}]\sum_{k=0}^{s-1} {}^k\boldsymbol{\lambda}, \quad (17a)\end{aligned}$$

$$\mathbf{S} = \tfrac{1}{2}\mathbf{H} + \tfrac{1}{2}\tilde{\mathbf{N}}\mathbf{D}_p\mathbf{N} - \tfrac{1}{2}\tilde{\mathbf{N}}\tilde{\mathbf{B}}_2\mathbf{K}_{22}^{-1}\mathbf{B}_2\mathbf{N} - \tilde{\mathbf{N}}\tilde{\mathbf{B}}_d(\tfrac{1}{2}\mathbf{K}_d^{-1} - \overset{++}{\mathbf{G}})\mathbf{B}_d\mathbf{N}, \quad (17b)$$

$$\mathbf{S}{}^s\boldsymbol{\lambda} + {}^s\mathbf{b} \ge \mathbf{0}, \quad {}^s\boldsymbol{\lambda} \ge \mathbf{0}, \quad {}^s\tilde{\boldsymbol{\lambda}}(\mathbf{S}{}^s\boldsymbol{\lambda} + {}^s\mathbf{b}) = 0, \quad (s = 1, 2, \ldots, l). \quad (17c\text{–}e)$$

In Eq. (17) it has been assumed that

$$\overset{+}{\mathbf{G}} = \frac{1}{\delta t}\boldsymbol{\Phi}\left(\int_0^{\delta t}\boldsymbol{\Theta}(\tau)d\tau\right)\tilde{\boldsymbol{\Phi}},$$

$$\overset{++}{\mathbf{G}} = \frac{1}{\delta t^2}\boldsymbol{\Phi}\left(\int_0^{\delta t}\int_0^{\tau}\boldsymbol{\Theta}(\bar{\tau})d\bar{\tau}d\tau\right)\tilde{\boldsymbol{\Phi}}. \quad (18a, b)$$

In order to obtain ${}^s\mathbf{b}$ in (17), it is necessary to know ${}^{s-1}\mathbf{u}_d$ and ${}^{s-1}\dot{\mathbf{u}}_d$; in order to obtain ${}^{s+1}\mathbf{b}$, it is necessary to know ${}^s\mathbf{u}_d$ and ${}^s\dot{\mathbf{u}}_d$; and so on. They can be calculated by the following formulae that can be easily obtained starting from (11):

$$\begin{aligned}{}^s\mathbf{u}_d = {} & \mathbf{G}\mathbf{K}_d^{s-1}\mathbf{u}_d - \dot{\mathbf{G}}\mathbf{M}_d^{s-1}\dot{\mathbf{u}}_d + (\mathbf{K}_d^{-1} - \mathbf{G})^{s-1}\mathbf{F}_d \\ & + (\mathbf{K}_d^{-1} - \overset{+}{\mathbf{G}})^s\delta\mathbf{F}_d + (\mathbf{K}_d^{-1} - \mathbf{G})\mathbf{B}_d\mathbf{N}\sum_{k=0}^{s-1} {}^k\boldsymbol{\lambda} \\ & + (\mathbf{K}_d^{-1} - \overset{+}{\mathbf{G}})\mathbf{B}_d\mathbf{N}{}^s\boldsymbol{\lambda}, \quad (19a)\end{aligned}$$

$$\begin{aligned}{}^s\dot{\mathbf{u}}_d = {} & \dot{\mathbf{G}}\mathbf{K}_d^{s-1}\mathbf{u}_d - \ddot{\mathbf{G}}\mathbf{M}_d^{s-1}\dot{\mathbf{u}}_d - \dot{\mathbf{G}}^{s-1}\mathbf{F}_d - \frac{1}{\delta t}(\mathbf{K}_d^{-1} - \mathbf{G})^s\delta\mathbf{F}_d \\ & - \dot{\mathbf{G}}\mathbf{B}_d\mathbf{N}\sum_{k=0}^{s-1} {}^k\boldsymbol{\lambda} - \frac{1}{\delta t}(\mathbf{K}_d^{-1} - \mathbf{G})\mathbf{B}_d\mathbf{N}{}^s\boldsymbol{\lambda}. \quad (19b)\end{aligned}$$

It is worth noticing that, being constant the width δt of the typical time subinterval, matrices $\ddot{\mathbf{G}}(\delta t)$, $\dot{\mathbf{G}}(\delta t)$, $\mathbf{G}(\delta t)$, $\overset{+}{\mathbf{G}}(\delta t)$, and $\overset{++}{\mathbf{G}}(\delta t)$ are also constant.

The solving set (17) is a generalization of analogous equations (see, e.g., Polizzotto, 1986; Muscolino, 1989) to the case of contemporaneous presence of SDF and DDF, including the particular cases of static and dynamic elastic-plastic analysis. With reference to matrix $\mathbf{S}$ in (17b), it is possible to state that: if all the degrees-of-freedom are not dynamic, the fourth term on the second side vanishes; if all the degrees-of-freedom are dynamic, the third term on the second side vanishes.

In this last case, matrix $\mathbf{S}$ is the sum of two block-diagonal matrices (the first and the second addend) plus a sparse matrix (the fourth addend). Considering that $\lim_{\delta t \to 0} \overset{++}{\mathbf{G}} = \frac{1}{2}\mathbf{K}_d^{-1}$, one recognizes that it becomes correct to disregard the fourth term with respect to the sum of the first two, if δt is sufficiently small with respect to the smallest natural period of the structure. The approximate elastic-plastic analysis that can be performed in such hypothesis implies the solution of relatively simple LCPs, with some consequent computational advantages (Giambanco et al., 1994).

Equations (17c–e) are a sequence of LCPs. If all the degrees-of-freedom are DDF (see, e.g., Polizzotto, 1986; Muscolino, 1989) and/or matrix $\mathbf{H}$ is positive-definite (see, e.g., Maier, 1968a), matrix $\mathbf{S}$ is positive-definite, otherwise it is positive-semi-definite (see, e.g., Maier, 1968a). If $\mathbf{S}$ is positive-definite the solution of problem (17) exists and is unique (see, e.g., Cottle and Dantzig, 1968; Lemke, 1968; Cottle, 1979; Cottle et al., 1992); if $\mathbf{S}$ is positive-semi-definite neither the existence of a bounded solution (i.e., instantaneous collapse may occur) nor its uniqueness is guaranteed (see, e.g., Cottle and Dantzig, 1968; Lemke, 1968; Cottle, 1979; Cottle et al., 1992); in this last case, every two solutions of the same problem

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

must differ by a stressless (i.e., compatible, corresponding to a mechanism) set of plastic deformations (Maier, 1968a).

Utilizing the Lagrange multiplier method it is easy to prove that Eqs. (17$c-e$) are the (necessary and sufficient) Kuhn-Tucker conditions related to the following equivalent sequence of QPPs (see, e.g., Cottle, 1979; Cottle et al., 1992):

$$ {}^s\Psi^* = \min_{({}^s\boldsymbol{\lambda})} {}^s\Psi = \min_{({}^s\boldsymbol{\lambda})} \tfrac{1}{2}\,{}^s\tilde{\boldsymbol{\lambda}}\mathbf{S}\,{}^s\boldsymbol{\lambda} + {}^s\tilde{\mathbf{b}}\,{}^s\boldsymbol{\lambda}, $$

$$ \text{subjected to} \quad {}^s\boldsymbol{\lambda} \geq \mathbf{0}, \quad \forall s \in I(l). \tag{20} $$

Functional ${}^s\Psi = {}^s\Psi({}^s\boldsymbol{\lambda})$ in (20) is strictly convex if $\mathbf{S}$ is positive-definite or convex if $\mathbf{S}$ is positive-semi-definite (see, e.g., Maier, 1968a).

It is well known that the sequence of LCPs (17) is also equivalent to the QPP sequence

$$ \min_{({}^s\boldsymbol{\lambda})} \tfrac{1}{2}\,{}^s\tilde{\boldsymbol{\lambda}}\mathbf{S}\,{}^s\boldsymbol{\lambda}, \quad \text{subjected to} $$

$$ \mathbf{S}\,{}^s\boldsymbol{\lambda} + {}^s\mathbf{b} \geq \mathbf{0}, \quad \forall s \in I(l). \tag{21} $$

One prefers to solve QPPs (20) rather than LCPs (17) very often. QPPs (and LCPs) may involve a high number of variables and this may cause the ineffectiveness of the available computational algorithms.

In the following section a special iterative method, suitably performed so as to quickly solve even large LCPs, is presented.

3 Linear Complementarity Problem Iterative Solution

3.1 The Proposed Method. Let us consider the typical sth step of the LCP sequence (17), i.e.,

$$ \mathbf{S}\,{}^s\boldsymbol{\lambda} + {}^s\mathbf{b} \geq \mathbf{0}, \quad {}^s\boldsymbol{\lambda} \geq \mathbf{0}, \quad {}^s\tilde{\boldsymbol{\lambda}}(\mathbf{S}\,{}^s\boldsymbol{\lambda} + {}^s\mathbf{b}) = 0. \tag{22} $$

The technique we are going to propose aims at obtaining iteratively the solution of problem (22).

Let us denote with $i = 1, 2, \ldots, f$ the index related to the ith iteration, f being related to the final one. Let us indicate with ${}^s\bar{\boldsymbol{\lambda}}^i$ the tentative solution of problem (22) which is determined during the ith iteration, starting from the known values of ${}^s\bar{\boldsymbol{\lambda}}^{i-1}$ (by definition ${}^s\bar{\boldsymbol{\lambda}}^0 = \mathbf{0}$).

Let us arbitrarily operate the following partitions:

$$ {}^s\tilde{\bar{\boldsymbol{\lambda}}}^i = [{}^s\tilde{\bar{\boldsymbol{\lambda}}}_1^i \quad {}^s\tilde{\bar{\boldsymbol{\lambda}}}_2^i \quad \ldots \quad {}^s\tilde{\bar{\boldsymbol{\lambda}}}_j^i \quad \ldots \quad {}^s\tilde{\bar{\boldsymbol{\lambda}}}_m^i], \tag{23a} $$

$$ {}^s\tilde{\mathbf{b}} = [{}^s\tilde{\mathbf{b}}_1 \quad {}^s\tilde{\mathbf{b}}_2 \quad \ldots \quad {}^s\tilde{\mathbf{b}}_j \quad \ldots \quad {}^s\tilde{\mathbf{b}}_m], \tag{23b} $$

such that

$$ \sum_{j=1}^{m} \rho_j = n, \tag{23c} $$

ρ_j being the order of the jth subvectors ${}^s\bar{\boldsymbol{\lambda}}_j^i$ and ${}^s\mathbf{b}_j$.

Let us define the $n \times n$ Jordan's canonical block-diagonal matrix

$$ \mathbf{S}_D = \text{diag}\,(\mathbf{S}_{jj}), \tag{24} $$

where $\mathbf{S}_{jj}(j \in I(m))$ is the $\rho_j \times \rho_j$ jth diagonal block of the matrix $\mathbf{S}$ corresponding to subvector ${}^s\bar{\boldsymbol{\lambda}}_j^i$. Furthermore, let us define the vector

$$ {}^s\bar{\mathbf{b}}^i = {}^s\mathbf{b} + (\mathbf{S} - \mathbf{S}_D)\,{}^s\bar{\boldsymbol{\lambda}}^{i-1} \quad (i = 1, 2, \ldots, f) \tag{25} $$

and also operate on it the same partition effected for ${}^s\mathbf{b}$:

$$ {}^s\tilde{\bar{\mathbf{b}}}^i = [{}^s\tilde{\bar{\mathbf{b}}}_1^i \quad {}^s\tilde{\bar{\mathbf{b}}}_2^i \quad \ldots \quad {}^s\tilde{\bar{\mathbf{b}}}_j^i \quad \ldots \quad {}^s\tilde{\bar{\mathbf{b}}}_m^i] \tag{26a} $$

where

$$ {}^s\bar{\mathbf{b}}_j^i = {}^s\mathbf{b}_j + \sum_{k=1}^{j-1} \mathbf{S}_{jk}\,{}^s\bar{\boldsymbol{\lambda}}_k^i + \sum_{k=j+1}^{m} \mathbf{S}_{jk}\,{}^s\bar{\boldsymbol{\lambda}}_k^{i-1}, $$

$$ (j = 1, 2, \ldots, m), \quad (i = 1, 2, \ldots, f). \tag{26b} $$

The vector ${}^s\bar{\boldsymbol{\lambda}}^i$ can be obtained by solving the recursive LCP

$$ \mathbf{S}_D\,{}^s\bar{\boldsymbol{\lambda}}^i + {}^s\bar{\mathbf{b}}^i \geq \mathbf{0}, \quad {}^s\bar{\boldsymbol{\lambda}}^i \geq \mathbf{0}, \quad {}^s\tilde{\bar{\boldsymbol{\lambda}}}^i(\mathbf{S}_D\,{}^s\bar{\boldsymbol{\lambda}}^i + {}^s\bar{\mathbf{b}}^i) = 0, $$

$$ (i = 1, 2, \ldots, f), \tag{27} $$

simpler than LCP (22) because matrix $\mathbf{S}_D$ is a block-diagonal one. Therefore, the search for vector ${}^s\bar{\boldsymbol{\lambda}}^i$ can be decoupled into the search for subvectors ${}^s\bar{\boldsymbol{\lambda}}_j^i$ $(j = 1, 2, \ldots, m)$ and it can be computed by solving the sequence of LCPs with order ρ_j:

$$ \mathbf{S}_{jj}\,{}^s\bar{\boldsymbol{\lambda}}_j^i + {}^s\bar{\mathbf{b}}_j^i \geq \mathbf{0}, \quad {}^s\bar{\boldsymbol{\lambda}}_j^i \geq \mathbf{0}, \quad {}^s\tilde{\bar{\boldsymbol{\lambda}}}_j^i(\mathbf{S}_{jj}\,{}^s\bar{\boldsymbol{\lambda}}_j^i + {}^s\bar{\mathbf{b}}_j^i) = 0, $$

$$ (j = 1, 2, \ldots, m), \quad (i = 1, 2, \ldots, f). \tag{28} $$

Making reference to formula (28), the index j assumes the meaning of index of the jth subiteration of the ith iteration of the sth analysis step.

Taking into account (25), problem (27) can be interpreted as a Jacobi family method.

According to (26b), during the current jth subiteration of the current ith iteration, the known terms ${}^s\bar{\mathbf{b}}_j^i$ in (28) depend on the value of the tentative solution obtained at the last $(i-1)$th iteration, ${}^s\bar{\boldsymbol{\lambda}}^{i-1}$. On the contrary, for a better effectiveness of the procedure, during the current jth subiteration of the current ith iteration it is suitable to consider the values of vectors ${}^s\boldsymbol{\lambda}_1^i, {}^s\boldsymbol{\lambda}_2^i, \ldots, {}^s\boldsymbol{\lambda}_{j-1}^i$ obtained during the effected $(j-1)$ subiterations of the same ith iteration. In order to achieve this goal, let us define the vector

$$ {}^s\tilde{\boldsymbol{\lambda}}^{ij} = [{}^s\tilde{\boldsymbol{\lambda}}_1^i\,{}^s\tilde{\boldsymbol{\lambda}}_2^i \ldots {}^s\tilde{\boldsymbol{\lambda}}_{j-1}^i\,{}^s\tilde{\boldsymbol{\lambda}}_j^i\,{}^s\tilde{\boldsymbol{\lambda}}_{j+1}^{i-1} \ldots {}^s\tilde{\boldsymbol{\lambda}}_{m-1}^{i-1}\,{}^s\tilde{\boldsymbol{\lambda}}_m^{i-1}] \tag{29} $$

whose entries are

(i) the known entries of vectors ${}^s\boldsymbol{\lambda}_1^i, {}^s\boldsymbol{\lambda}_2^i, \ldots, {}^s\boldsymbol{\lambda}_{j-1}^i$ evaluated during the effected $(j-1)$ subiterations of the current ith iteration;
(ii) the unknown entries of vector ${}^s\boldsymbol{\lambda}_j^i$ to be determined during the current jth subiteration of the current ith iteration;
(iii) the known entries of vectors ${}^s\boldsymbol{\lambda}_{j+1}^{i-1}, \ldots, {}^s\boldsymbol{\lambda}_{m-1}^{i-1}, {}^s\boldsymbol{\lambda}_m^{i-1}$ evaluated during the last $(m-j)$ subiterations of the last $(i-1)$th iteration.

Furthermore, vector ${}^s\boldsymbol{\lambda}^{im}$, obtained at the final subiteration m of iteration i, is assumed as tentative solution of the same iteration i, i.e.,

$$ {}^s\boldsymbol{\lambda}^i \equiv {}^s\boldsymbol{\lambda}^{im} \tag{30} $$

where, by definition, ${}^s\boldsymbol{\lambda}^0 \equiv {}^s\boldsymbol{\lambda}^{0m} = \mathbf{0}$.

Taking into account position (29), the sequence of LCPs (28) does not formally change, but the known terms ${}^s\bar{\mathbf{b}}_j^i$ assume a brought up-to-date value ${}^s\mathbf{b}_j^i$, i.e.,

$$ {}^s\mathbf{b}_j^i = {}^s\mathbf{b}_j + \sum_{k=1}^{j-1} \mathbf{S}_{jk}\,{}^s\boldsymbol{\lambda}_k^i + \sum_{k=j+1}^{m} \mathbf{S}_{jk}\,{}^s\boldsymbol{\lambda}_k^{i-1}, $$

$$ (j = 1, 2, \ldots, m), \quad (i = 1, 2, \ldots, f), \tag{31} $$

$$ \mathbf{S}_{jj}\,{}^s\boldsymbol{\lambda}_j^i + {}^s\mathbf{b}_j^i \geq \mathbf{0}, \quad {}^s\boldsymbol{\lambda}_j^i \geq \mathbf{0}, \quad {}^s\tilde{\boldsymbol{\lambda}}_j^i(\mathbf{S}_{jj}\,{}^s\boldsymbol{\lambda}_j^i + {}^s\mathbf{b}_j^i) = 0, $$

$$ (j = 1, 2, \ldots, m), \quad (i = 1, 2, \ldots, f). \tag{32} $$

Making reference to position (29), one recognizes that problem (32) represents a Gauss-Seidel family method.

By means of the Lagrange multiplier method it is easy to show that the LCP sequence (32) represents the Kuhn-Tucker equations related to the following QPP sequence:

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\min_{({}^s\boldsymbol{\lambda}_j^i)} {}^s\Psi_j^i = \min_{({}^s\boldsymbol{\lambda}_j^i)} \tfrac{1}{2}{}^s\tilde{\boldsymbol{\lambda}}_j^i \mathbf{S}_{jj}{}^s\boldsymbol{\lambda}_j^i + {}^s\tilde{\mathbf{b}}_j^i{}^s\boldsymbol{\lambda}_j^i \quad \text{subject to:} \quad (33a)$$

$${}^s\boldsymbol{\lambda}_j^i \geq \mathbf{0}, \quad (j = 1, 2, \ldots, m), \quad (i = 1, 2, \ldots, f), \quad (33b)$$

where ${}^s\Psi_j^i$ is a convex functional depending just on ${}^s\boldsymbol{\lambda}_j^i$.

The vector ${}^s\boldsymbol{\lambda}^i$ in (30) can thus be obtained, starting from ${}^s\boldsymbol{\lambda}^{i-1}$, by solving the sequence of problems (32) or (33) for all j.

3.2 Procedure Convergence. It is possible to prove the following convergence theorem: *during the iterative procedure (32) the functional* ${}^s\Psi^i = \frac{1}{2}{}^s\tilde{\boldsymbol{\lambda}}^i\mathbf{S}{}^s\boldsymbol{\lambda}^i + {}^s\tilde{\mathbf{b}}{}^s\boldsymbol{\lambda}^i$, *associate to the solution* ${}^s\boldsymbol{\lambda}^i$ *related to the typical ith iteration, monotonically converges towards its minimum* ${}^s\Psi^*$.

In order to prove the above theorem, let us suppose that matrix $\mathbf{S}$ is positive-definite and let us first make reference to the LCP sequence (27). At the fth (final) iteration, if ${}^s\bar{\boldsymbol{\lambda}}^f = {}^s\bar{\boldsymbol{\lambda}}^{f-1}$ results, then problem (27) takes on the form

$$\mathbf{S}{}^s\bar{\boldsymbol{\lambda}}^f + {}^s\mathbf{b} \geq \mathbf{0}, \quad {}^s\bar{\boldsymbol{\lambda}}^f \geq \mathbf{0}, \quad {}^s\tilde{\bar{\boldsymbol{\lambda}}}^f(\mathbf{S}{}^s\bar{\boldsymbol{\lambda}}^f + {}^s\mathbf{b}) = 0, \quad (34)$$

so that problem (27) coincides with problem (22) and thus

$${}^s\bar{\boldsymbol{\lambda}}^f = {}^s\boldsymbol{\lambda} \quad (35)$$

must also result. Relation (35) also holds for ${}^s\bar{\boldsymbol{\lambda}}^f$ obtained by solving the LCP sequence (28) equivalent to the LCP sequence (27).

Now, let us make reference to the LCP sequence (32). At the fth (final) iteration, if ${}^s\boldsymbol{\lambda}^{fm} = {}^s\boldsymbol{\lambda}^{(f-1)m}$ results, then ${}^s\mathbf{b}_j^f$ in (31) coincides with ${}^s\bar{\mathbf{b}}_j^f$ in (26b) ($j = 1, 2, \ldots, m$), LCPs (32) coincide with LCPs (28) and thus even ${}^s\boldsymbol{\lambda}^{fm} \equiv {}^s\boldsymbol{\lambda}^f$ fulfills relation (35).

As a consequence, we can state the following:

PROPERTY 1: *if the tentative solution of problem (32)* ${}^s\boldsymbol{\lambda}^{im} \equiv {}^s\boldsymbol{\lambda}^i$ *converges, then it converges to the solution of LCP (22)* ${}^s\boldsymbol{\lambda}$ *and functional* ${}^s\Psi^i$ *reaches its global minimum* ${}^s\Psi^*$.

Let us consider now the strictly convex functional

$${}^s\Psi^{ij} = {}^s\Psi^{ij}({}^s\boldsymbol{\lambda}^{ij}) = \tfrac{1}{2}{}^s\tilde{\boldsymbol{\lambda}}^{ij}\mathbf{S}{}^s\boldsymbol{\lambda}^{ij} + {}^s\tilde{\mathbf{b}}{}^s\boldsymbol{\lambda}^{ij}, \quad (36)$$

analogous to functional ${}^s\Psi$ in (20), but with ${}^s\boldsymbol{\lambda}^{ij}$ as defined in (29). Taking into account that the unknown entries of vector ${}^s\boldsymbol{\lambda}^{ij}$ are just the components of subvector ${}^s\boldsymbol{\lambda}_j^i$, Eq. (36) can be rewritten as follows:

$${}^s\Psi^{ij}({}^s\boldsymbol{\lambda}^{ij}) = \tfrac{1}{2}{}^s\tilde{\boldsymbol{\lambda}}_j^i\mathbf{S}_{jj}{}^s\boldsymbol{\lambda}_j^i + {}^s\tilde{\mathbf{b}}_j^i{}^s\boldsymbol{\lambda}_j^i + {}^sc_j^i = {}^s\Psi_j^i({}^s\boldsymbol{\lambda}_j^i) + {}^sc_j^i, \quad (37)$$

where the strictly convex functional ${}^s\Psi_j^i$ is the same as in (33) and

$$\begin{aligned} {}^sc_j^i = \tfrac{1}{2}\Bigg(&\sum_{k1=1}^{j-1}\sum_{k2=1}^{j-1}{}^s\tilde{\boldsymbol{\lambda}}_{k1}^i\mathbf{S}_{k1k2}{}^s\boldsymbol{\lambda}_{k2}^i + \sum_{k1=j+1}^{m}\sum_{k2=j+1}^{m}{}^s\tilde{\boldsymbol{\lambda}}_{k1}^{i-1}\mathbf{S}_{k1k2}{}^s\boldsymbol{\lambda}_{k2}^{i-1} \\ &+ \sum_{k1=j+1}^{m}\sum_{k2=1}^{j-1}{}^s\tilde{\boldsymbol{\lambda}}_{k1}^{i-1}\mathbf{S}_{k1k2}{}^s\boldsymbol{\lambda}_{k2}^i + \sum_{k1=1}^{j-1}\sum_{k2=j+1}^{m}{}^s\tilde{\boldsymbol{\lambda}}_{k1}^i\mathbf{S}_{k1k2}{}^s\boldsymbol{\lambda}_{k2}^{i-1}\Bigg) \\ &+ \sum_{k=1}^{j-1}{}^s\tilde{\mathbf{b}}_k^i{}^s\boldsymbol{\lambda}_k^i + \sum_{k=j+1}^{m}{}^s\tilde{\mathbf{b}}_k^i{}^s\boldsymbol{\lambda}_k^{i-1}, \end{aligned} \quad (38)$$

is a constant scalar quantity during the jth subiteration of the ith iteration.

As a consequence of (37) and (38), we can write

$$\{\min_{({}^s\boldsymbol{\lambda}^{ij})} {}^s\Psi^{ij} \text{ subject to: } {}^s\boldsymbol{\lambda}^{ij} \geq \mathbf{0}\} = \{\min_{({}^s\boldsymbol{\lambda}_j^i)} ({}^s\Psi_j^i + {}^sc_j^i) \text{ subject to: } {}^s\boldsymbol{\lambda}_j^i \geq \mathbf{0}\}. \quad (39)$$

Taking in mind that ${}^s\Psi^{ij}$ is a strictly convex functional, if the minimization process on the left-hand side of (39) provides ${}^s\boldsymbol{\lambda}^{ij} \neq {}^s\boldsymbol{\lambda}^{i(j-1)}$ (and, consequently, ${}^s\boldsymbol{\lambda}_j^i \neq {}^s\boldsymbol{\lambda}_j^{i-1}$), then ${}^s\Psi^{ij} < {}^s\Psi^{i(j-1)}$ results. Moreover, subvector ${}^s\boldsymbol{\lambda}_j^i$, obtained by the minimization process of functional ${}^s\Psi^{ij}$, also minimizes functional ${}^s\Psi_j^i$ on the right-hand side of (39).

Remembering the equivalence between problems (32) and (33), we deduce that, if the jth subiteration of the ith iteration of problem (32) provides ${}^s\boldsymbol{\lambda}_j^i \neq {}^s\boldsymbol{\lambda}_j^{i-1}$, then ${}^s\Psi^{ij} < {}^s\Psi^{i(j-1)}$ results. Furthermore, setting

$${}^s\Psi^i \equiv {}^s\Psi^{im}, \quad (40)$$

if, at the end of the ith iteration, the LCP sequence (32) provides ${}^s\boldsymbol{\lambda}^i \neq {}^s\boldsymbol{\lambda}^{i-1}$, then ${}^s\Psi^i < {}^s\Psi^{i-1}$ also results, and we can finally state the following

PROPERTY 2: *if at the ith iteration the tentative solution* ${}^s\boldsymbol{\lambda}^i$ *of problem (32) changes with respect to the solution obtained at the (i − 1)th iteration* ${}^s\boldsymbol{\lambda}^{i-1}$, *then functional* ${}^s\Psi^i$ *reduces with respect to* ${}^s\Psi^{i-1}$.

Coupling the above described properties 2 and 1 one realizes that:

- if during two subsequent iterations the tentative solution changes, i.e., ${}^s\boldsymbol{\lambda}^i \neq {}^s\boldsymbol{\lambda}^{i-1}$ results, then ${}^s\Psi^i < {}^s\Psi^{i-1}$ also results;
- if during two subsequent iterations the tentative solution does not change, i.e., ${}^s\boldsymbol{\lambda}^f = {}^s\boldsymbol{\lambda}^{f-1}$ results, then ${}^s\boldsymbol{\lambda}^f = {}^s\boldsymbol{\lambda}$ also results and the functional $\frac{1}{2}{}^s\tilde{\boldsymbol{\lambda}}\mathbf{S}{}^s\boldsymbol{\lambda} + {}^s\tilde{\mathbf{b}}{}^s\boldsymbol{\lambda}$ reaches its global minimum value ${}^s\Psi^f = {}^s\Psi^*$.

As a consequence, the following CONVERGENCE THEOREM holds: *during the iterative procedure (32) the functional* ${}^s\Psi^i$ *monotonically converges towards its minimum* ${}^s\Psi^*$.

At the end of the procedure, when ${}^s\Psi^f = {}^s\Psi^{f-1} = {}^s\Psi^*$ results, ${}^s\boldsymbol{\lambda}^f = {}^s\boldsymbol{\lambda}^{f-1} = {}^s\boldsymbol{\lambda}$ also results, but vector ${}^s\boldsymbol{\lambda}^i$ does not necessarily converge in a monotonic way during the whole process.

The final number f of iterations necessary to exactly obtain ${}^s\Psi^f = {}^s\Psi^{f-1}$ can be infinite; nevertheless, as usual, it can always be reduced to a finite number breaking off the procedure when ${}^s\Psi^f = {}^s\Psi^{f-1}$ results, within a suitably chosen tolerance.

If $\mathbf{S}$ is PSD, functional ${}^s\Psi^i$ is unbounded if instantaneous collapse occurs, otherwise it possesses a global minimum ${}^s\Psi^* \equiv {}^s\Psi^f$ which can be related to more than one vector ${}^s\boldsymbol{\lambda}^f$: ${}^s\boldsymbol{\lambda}^{f'}$, ${}^s\boldsymbol{\lambda}^{f''}$, etc. (see, e.g., Lemke, 1968). As a consequence, property 2 certainly holds during the initial phase of the minimization process, but it does not necessarily hold in optimality conditions. Nevertheless, when the global minimum ${}^s\Psi^f$ is reached, the recursive formulae (32), due to their special structure, guarantee that the obtained solution ${}^s\boldsymbol{\lambda}^f$ does no more change. In other words, taking into account (31), it is easy to recognize that if ${}^s\boldsymbol{\lambda}^f$ is a solution to (22), then (32) provide ${}^s\boldsymbol{\lambda}^{f+k} = {}^s\boldsymbol{\lambda}^f$, $\forall k$.

The convergence of problems (27) and (28) can be proved following a reasoning path similar to the one exposed for problem (32). Actually, the validity of property 1 for problem (32) has already been proved starting from its validity for problems (27) and (28). In addition, the definition of a suitable functional ${}^s\bar{\Psi}^i$, analogous to ${}^s\Psi$ in (20), but related to the full vector ${}^s\bar{\boldsymbol{\lambda}}^i$ defined as in (23a), makes also possible the proof of a specialized property 2 for problem (27) and, consequently, for the equivalent problem (28).

3.3 The Improved Element-by-Element Optimization Technique. Let us consider the special case $m = n_e$ and let us suppose that vector ${}^s\boldsymbol{\lambda}_j^i$ collects the increments of the plastic activation intensities related to the jth finite element ($j = 1, 2, \ldots, m$).

On the grounds of the above positions the proposed formulae (32) identifies with an improved EOT (IEOT). Actually, by utilizing (32) one searches for the plastic strains element by element (as by EOT), but at the typical jth subiteration one takes into account the influence of the first j analyzed elements (in terms of increments of plastic activation intensity) on the elastic response of the structure, while by utilizing EOT the elastic response is brought up to date only when all the elements

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

have been analyzed. On the other side, vector ${}^s\boldsymbol{\lambda}_j^i$ obtained by means of (32) is the vector which minimizes functional ${}^s\Psi_j^i$ in (33); consequently, any other vector ${}^s\boldsymbol{\lambda}_j^i$ differently computed (e.g., by means of EOT) is related to a lower reduction of ${}^s\Psi_j^i$, and then with a slower convengence.

3.4 The Asymptotic Pivoting Method. Let us consider the special case $m = n$, so that $\rho_j = 1\ \forall j \in I(m)$ and

$$ {}^s\tilde{\bar{\boldsymbol{\lambda}}}^i = [{}^s\bar{\lambda}_1^i \quad {}^s\bar{\lambda}_2^i \quad \ldots \quad {}^s\bar{\lambda}_j^i \quad \ldots \quad {}^s\bar{\lambda}_n^i], \tag{41} $$

$$ \mathbf{S}_D = \text{diag}\,(S_{11}, S_{22}, \ldots, S_{jj}, \ldots, S_{nn}). \tag{42} $$

In this case problem (27) can be written as

$$ \left.\begin{aligned} &{}^s\bar{\boldsymbol{\lambda}}^i + [\mathbf{S}_D^{-1}{}^s\mathbf{b} - (\mathbf{I} - \mathbf{S}_D^{-1}\mathbf{S}){}^s\bar{\boldsymbol{\lambda}}^{i-1}] \geq \mathbf{0},\ {}^s\bar{\boldsymbol{\lambda}}^i \geq \mathbf{0}, \\ &{}^s\tilde{\bar{\boldsymbol{\lambda}}}^i\{{}^s\bar{\boldsymbol{\lambda}}^i + [\mathbf{S}_D^{-1}{}^s\mathbf{b} - (\mathbf{I} - \mathbf{S}_D^{-1}\mathbf{S}){}^s\bar{\boldsymbol{\lambda}}^{i-1}]\} = 0, \end{aligned}\right\} $$

$$ (i = 1, 2, \ldots, f). \tag{43} $$

The solution of problem (43) can be reached by solving, component-wise, the following maximum problems:

$$ {}^s\bar{\boldsymbol{\lambda}}^i = \max\{[-\mathbf{S}_D^{-1}{}^s\mathbf{b} + (\mathbf{I} - \mathbf{S}_D^{-1}\mathbf{S}){}^s\bar{\boldsymbol{\lambda}}^{i-1}], [\mathbf{0}]\}, $$

$$ (i = 1, 2, \ldots, f). \tag{44} $$

Furthermore, with $m = n$, vector ${}^s\boldsymbol{\lambda}^{ij}$ in (29) assumes the form

$$ {}^s\tilde{\boldsymbol{\lambda}}^{ij} = [{}^s\lambda_1^i \quad {}^s\lambda_2^i \quad \ldots \quad {}^s\lambda_{j-1}^i \quad {}^s\lambda_j^i \quad {}^s\lambda_{j+1}^{i-1} \quad \ldots \quad {}^s\lambda_n^{i-1}], \tag{45} $$

so that the LCP sequence (32) can be written as

$$ \left.\begin{aligned} &{}^s\lambda_j^i + \frac{1}{S_{jj}}\left({}^sb_j + \sum_{k=1}^{j-1} S_{jk}{}^s\lambda_k^i + \sum_{k=j+1}^{n} S_{jk}{}^s\lambda_k^{i-1}\right) \geq 0,\ {}^s\lambda_j^i \geq 0, \\ &{}^s\lambda_j^i\left\{{}^s\lambda_j^i + \frac{1}{S_{ij}}\left({}^sb_j + \sum_{k=1}^{j-1} S_{jk}{}^s\lambda_k^i + \sum_{k=j+1}^{n} S_{jk}{}^s\lambda_k^{i-1}\right)\right\} = 0, \end{aligned}\right\} $$

$$ \forall j \in I(n),\ \forall i \in I(f). \tag{46} $$

The solution of problem (46) can be reached by solving the following maximum problems:

$$ {}^s\lambda_j^i = \max\left\{\left[-\frac{1}{S_{ij}}\left({}^sb_j + \sum_{k=1}^{j-1} S_{jk}{}^s\lambda_k^i + \sum_{k=j+1}^{n} S_{jk}{}^s\lambda_k^{i-1}\right)\right], [0]\right\} $$

$$ \forall j \in I(n),\ \forall i \in I(f). \tag{47} $$

The sequence of maximum problems (47) identifies with a pivoting method, named the Asymptotic Pivoting Method (APM) (Giambanco, 1998) because it theoretically may reach the solution of problem (22) in an infinite number of iterations.

4 Applications

As a first application the plane frame of Fig. 1(*a*) has been studied. The special feature of the frame allowed us to effect several simple computation in order to check the correctness of the utilized algorithm.

The frame is constituted by rectangular box section elements (Fig. 1(*b*)). The length of the columns is 350 cm, the length of the beams is 500 cm; for all the columns $B = H = 25$ cm has been adopted; for all the beams $B = 25$ cm and $H = 50$ cm have been adopted. The thickness b has been assumed:

$$ -\text{FE 1, 3, 4:}\quad b = 0.50 \text{ cm}; $$

$$ -\text{FE 2, 5:}\quad b = 1.00 \text{ cm}; $$

$$ -\text{FE 6 to 9:}\quad b = 0.40 \text{ cm}. $$

The typical FE (finite element) (column or beam element) is constituted by two purely elastic portions of equal length at

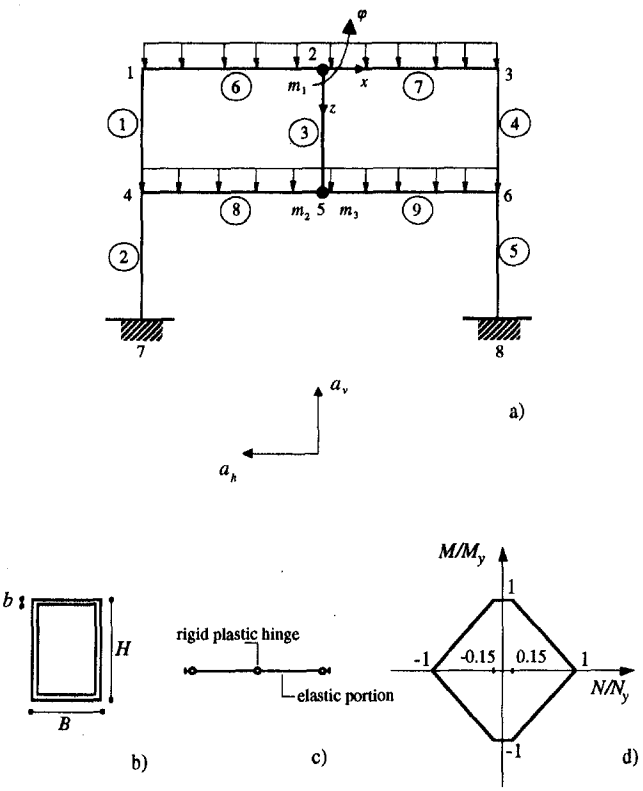


Fig. 1 Two floors elastic plastic plane frame dynamically loaded: (*a*) geometry and load condition; (*b*) typical cross section; (*c*) typical finite element; (*d*) dimensionless typical rigid plastic hinge domain

the extremes of which rigid-perfectly plastic hinges are located (Fig. 1(*c*)). The rigid plastic domain of the typical hinge is plotted in Fig. 1(*d*) where the interaction between N and M is taken into account. N_y and M_y represent the yield value for the axial force and the bending moment, respectively.

Three degrees-of-freedom are considered for each node. The horizontal displacements of nodes 2, 5 and the vertical displacement of node 5 (where the masses m_1, m_2, m_3 are, respectively, located) are dynamic, while the other ones are not dynamic. The other data are: $E^e = 21{,}000$ kN/cm^2; $\sigma_y = 40$ kN/cm^2; $m_1 = m_2 = m_3 = 0.2$ kNsec^2/cm; $\xi_1 = \xi_2 = \xi_3 =$ damping ratios $= 0.05$.

The natural frequencies and periods of the structure are: $\omega_1 = 11.069$ sec^{-1}, $\omega_2 = 21.500$ sec^{-1}, $\omega_3 = 32.071$ sec^{-1}; $T_1 = 0.56751$ sec, $T_2 = 0.29219$ sec, $T_3 = 0.19588$ sec. A time step $\delta t = 0.004$ sec, sufficiently small with respect to T_3, and a time interval ${}^lt = 2$ sec have been assumed.

The beams are loaded by a time-independent uniformly distributed vertical load $q = 0.15$ kN/cm, while the soil is subjected to a horizontal sinusoidal acceleration $a_h = \bar{a}_h \sin \omega_a t$ and to a vertical sinusoidal acceleration $a_v = \bar{a}_v \sin \omega_a t$ with amplitudes $\bar{a}_h$ and $\bar{a}_v$ and frequency ω_a.

The numerical computation has been effected solving the QPPs (20) utilizing the QPSOL Fortran routine (Gill et al., 1984), and solving the LCPs (17) by means of the proposed APM (Sect. 3.4) utilizing the own APMSOL Fortran routine (Giambanco, 1998).

At first it is assumed $\omega_a = \omega_1$; in Table 1 the CPU times (sec) required by APMSOL and QPSOL routines, for different values of amplitudes $\bar{a}_h$ and $\bar{a}_v$, are reported. Successively, it is assumed $\omega_a = \omega_2$; the CPU times are reported in Table 1 too. It is easy to observe that the CPU time necessary to perform the analysis by means of APMSOL has been definitely less than the analogous one related to QPSOL.

As a second application the plane frame of Fig. 2 has been studied. It is constituted by elements with rectangular box sec-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 1 CPU time (sec) required by QPSOL and by APMSOL

CPU time (sec)							
$\omega_a = \omega_1 = 11.069$; $^t t = 6$ sec; $\delta t = 0.004$				$\omega_a = \omega_2 = 21.5$; $^t t = 6$ sec; $\delta t = 0.004$			
$\bar{a}_h$	$\bar{a}_v$	QPSOL	APMSOL	$\bar{a}_h$	$\bar{a}_v$	QPSOL	APMSOL
50	0	54.383	1.886	0	200	70.984	5.674
75	0	54.367	1.887	20	200	68.582	5.180
100	0	62.977	5.926	40	200	67.600	5.200
125	0	77.129	10.503	60	200	66.947	5.053
150	0	93.305	14.471	80	200	66.892	5.232
50	50	54.422	1.891	100	200	67.112	5.535
75	50	54.385	1.892	120	200	66.323	5.250
100	50	63.262	5.965	140	200	65.432	5.174
125	50	67.916	10.675	160	200	65.168	5.241
150	50	93.724	14.509	180	200	65.772	5.459
50	100	54.387	1.885	200	200	66.010	5.813
75	100	54.373	1.887	220	200	65.871	5.701
100	100	63.415	6.026	240	200	66.271	5.916
125	100	72.283	10.955	260	200	66.069	5.916
150	100	93.712	14.501	280	200	65.705	6.089
subtotal time		1014.040	104.859	subtotal time		1000.738	82.433
total time		2014.778	187.292	p.c. time		100.0	9.29

tions as in Fig. 1(b). The length of the columns is 300 cm, the length of the beams is 600 cm; except beams 24 and 25 for which $B = 25$ cm and $H = 50$ cm, for all the other finite element (FE) $B = H = 25$ cm has been adopted, while thickness b has been assumed:

$$-\text{FE 1 to 15:} \quad b = 0.50 \text{ cm};$$

$$-\text{FE 16 to 19:} \quad b = 0.30 \text{ cm};$$

$$-\text{FE 20 to 25:} \quad b = 0.40 \text{ cm}.$$

The elastic-plastic behavior of the elements is assumed as previously described for frame in Fig. 1.

Three degrees-of-freedom are considered for each node. The horizontal displacements of nodes 2, 5, 8, 11, 14 (where the masses $m_1, m_2, \ldots, m_5$ are, respectively, located) are dynamic, while the other ones are not dynamic. The other data are: $E^e = 20{,}000$ kN/cm^2; $\sigma_y = 40$ kN/cm^2; $m_1 = m_2 = m_3 = m_4 = m_5 = 0.2$ kNsec^2/cm; $\xi_1 = \xi_2 = \xi_3 = \xi_4 = \xi_5 =$ damping ratios $= 0.05$.

The natural frequencies and periods of the structure are: $\omega_1 = 4.1467$ sec^{-1}, $\omega_2 = 12.700$ sec^{-1}, $\omega_3 = 22.858$ sec^{-1}; $\omega_4 = 34.514$ sec^{-1}, $\omega_5 = 45.322$ sec^{-1}; $T_1 = 1.5150$ sec, $T_2 = 0.49464$ sec, $T_3 = 0.27482$ sec, $T_4 = 0.18201$ sec, $T_5 = 0.13861$ sec. A time step $\delta t = 0.01$ sec, sufficiently small with respect to T_5, and a time interval $^t t = 4$ sec, have been assumed.

The beams are loaded by a time-independent uniformly distributed vertical load $q = 0.15$ kN/cm, while the soil is subjected to a horizontal sinusoidal acceleration $a = \bar{a} \sin \omega_a t$ with amplitude $\bar{a}$ and frequency ω_a. At first, it is assumed $\omega_a = \omega_1$ and $\bar{a} = 60$ cm/sec^2. In Fig. 3 the dimensionless acceleration history, the total and the residual displacements of floors 1, 3, and 5 are reported. In Table 2 the CPU times required by APMSOL and QPSOL routines, for different values of amplitude $\bar{a}$, are reported. Successively, it is assumed $\omega_a = \omega_2$; the CPU times are reported in Table 2, as well.

The structure has also been studied by operating different choices for ρ_j (see 23c) and, consequently, for $\mathbf{S}_D$ (see 24), in particular the IEOT (Sect. 3.3) has been utilized. In all the effected computations the procedure always showed very good convergence features, but the CPU time required did not result competitive with respect to the analogous time related to APM, as shown in Table 2.

It is worth noting that, although the frames studied represent relatively small size structures, the LCPs which must be solved in order to obtain the elastic plastic solution involve a relatively high number of variables: 162 plastic activations for the frame of Fig. 1 and 450 plastic activations for the frame of Fig. 2.

As stated in the previous sections, when matrix $\mathbf{S}$ is PSD neither the existence of a bounded solution nor its uniqueness is guaranteed. The computational experiences effected in the present paper always involved matrices $\mathbf{S}$ PSD, but, although this occurrence, we always obtained unique solutions, as verified by comparison among the results deduced from different approaches. This result probably depends on the particular adopted loading histories. Anyway, the authors verified in some applications worked out in the quasi-static field (Giambanco et al., 1998) that sometimes it is possible to obtain nonunique solution of the analysis problem.

In order to check the convergence of the proposed procedure, in all the computations performed by the proposed method the relative imprecision of functional $^s\Psi$ (defined as $(^s\Psi^f - {}^s\Psi^{f-1})/{}^s\Psi^{f-1}$) has been computed and it has been assumed that the convergence was reached when it was not greater than 10^{-15}. Satisfying this condition a maximum relative imprecision less than 10^{-7} has been always generated (and frequently less than 10^{-10}) with respect to the components of vector $^s\boldsymbol{\lambda}$ and this value seems to be certainly encouraging.

The check of the convergence involves high computational effort and, as a consequence, it has not been effected for each iteration but just after a suitably fixed number of iterations (the choice of one check every ten iterations has always given good results).

The number of iterations necessary to reach the fixed relative imprecision of functional $^s\Psi$ ranged between few tens and few thousands. In all the effected computations the maximum number of iterations has been fixed in 3000.

5 Conclusions

The present paper has been devoted to the study of a special iterative method devoted to solve LCPs, in the framework of the analysis of elastic plastic structures. For this purpose, a quite general formulation of discrete structure elastic-plastic analysis was first developed, taking into account work-hardening and the possible contemporaneous presence of dynamic and not dynamic degrees-of-freedom.

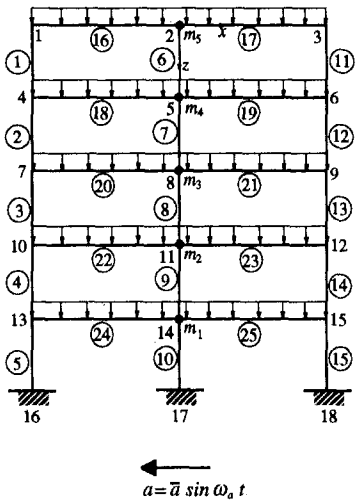


Fig. 2 Five floors elastic-plastic plane frame dynamically loaded: geometry and load condition

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

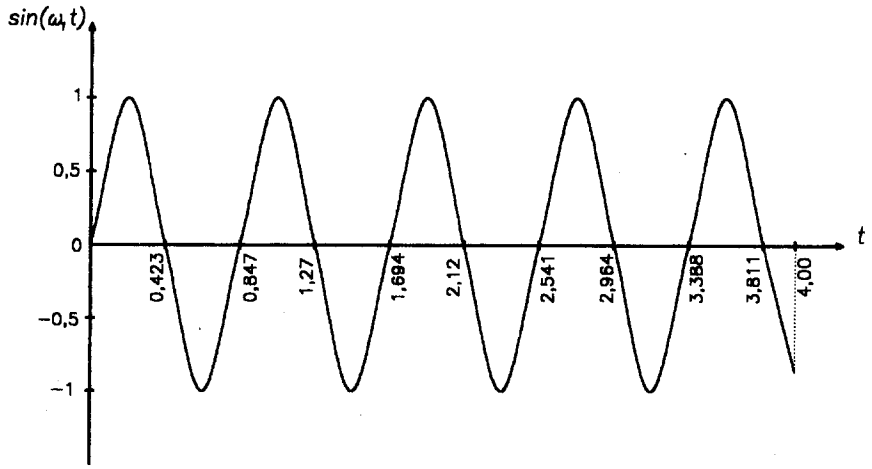


Fig. 3(*a*)

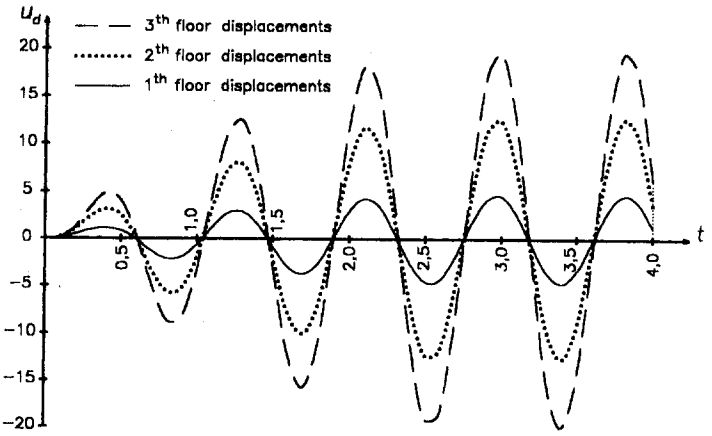


Fig. 3(*b*)

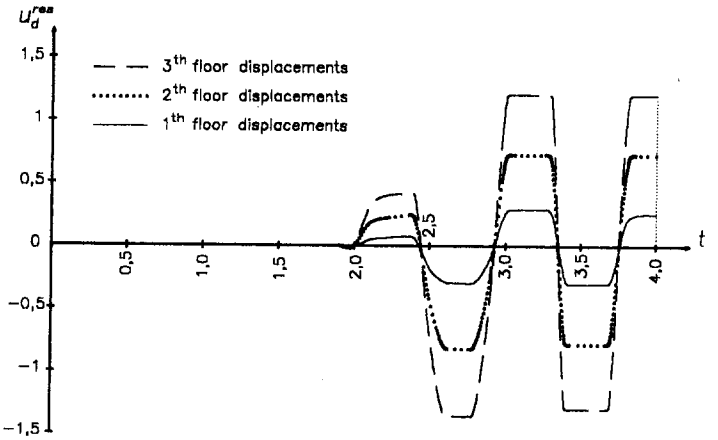


Fig. 3(*c*)

Fig. 3 Dynamic behavior of the frame in Fig. 2 subjected to a sinusoidal soil acceleration with frequency equal to the first natural frequency ($\omega_a = \omega_1$): (*a*) dimensionless acceleration history; (*b*) total displacements of floors 1, 3, and 5; (*c*) residual displacements of floors 1, 3, and 5

The proposed method is based on the recursive solution of a sequence of LCPs in which the constraint matrix is block-diagonal and is deduced by the matrix of the original LCP. It is worth noticing that the orders of the blocks of the constraint matrix can be arbitrarily chosen. This last feature makes possible the solution of even very large LCPs (they could be even exceeding the RAM of the available computer) through a sharing of the relevant problem in blocks with suitably reduced order which can sequentially be solved.

Two particular cases have been discussed. The first one assumes the order of the blocks of the block-diagonal constraint matrix equal to the number of plastic activations related to the typical finite element and has been recognized to be an improved EOT; the second one assumes order one for all the blocks and has been recognized to be a special pivoting method.

A convergence theorem was worked out. It was proved that a suitable functional of the variables monotonically converges

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 2 CPU time (sec) required by QPSOL, by IEOT, and by APMSOL

CPU time (sec)							
$\omega_a = \omega_1 = 4.1467$; $^1t = 4$ sec; $\delta t = 0.01$				$\omega_a = \omega_2 = 12.7$; $^1t = 4$ sec; $\delta t = 0.01$			
$\bar{a}$	QPSOL	IEOT	APMSOL	$\bar{a}$	QPSOL	IEOT	APMSOL
40	243.484	22.863	1.829	200	120.986	11.362	0.606
50	275.060	61.435	14.221	220	124.177	20.749	1.846
60	382.835	109.986	29.726	240	136.107	36.477	4.327
70	518.662	151.744	42.151	260	151.043	52.836	8.042
80	662.318	167.532	53.354	280	173.175	56.281	11.146
90	813.066	233.454	63.960	300	195.665	64.128	15.175
100	934.473	291.694	71.145	320	233.629	70.056	21.070
110	1096.414	301.519	77.115	340	257.665	74.723	24.486
120	1183.273	320.743	83.964	360	285.359	86.319	28.209
130	1306.832	346.578	94.953	380	312.997	101.819	29.771
140	1389.746	374.976	96.893	400	334.253	106.840	33.495
subtotal	8799.038	2382.524	629.311	subtotal	2325.056	681.590	178.173
total time	11124.094	3064.114	807.484	p.c. time	100.0	27.54	7.25

to its minimum value. This property has been useful in the computational stage.

The effected computational experiences have shown the effectiveness of the proposed method: suitably implemented, it provides extremely accurate results requiring very short CPU time. A further advantage of the method consists in the small required workspace storage: just the lower triangular part of the constraint matrix must be stored, while practically no other workspace storage is required.

The possible suitable choices of some check parameters, as the prefixed tolerance and the maximum number of iterations, have also been discussed.

The extension of the analysis to the presence of second-order geometrical effects and the refinement of the computational aspects will be object of future papers.

References

Borino, G., Caddemi, S., and Polizzotto, C., 1987, "Mathematical programming methods for evaluating dynamic deformations of elastoplastic structures," *Computational Plasticity,* Part II, D. R. J. Owen, E. Hinton, E. Oñate, eds., Pineridge Press, Swansea, UK, pp. 1231–1245.

Borino, G., and Polizzotto, C., 1989, "A finite element elastic-plastic dynamic analysis by mathematical programming," *Proc. FEMSA,* Feb., 8–10, Stellenboschy, South Africa.

Capurso, M., and Maier, G., 1970, "Incremental elastoplastic analysis and quadratic optimization," *Meccanica,* Vol. 5, pp. 107–116.

Corradi, L., 1978, "On compatible finite element models for elastic plastic analysis," *Meccanica,* Vol. 13, pp. 133–150.

Corradi, L., 1983, "A displacement formulation for the finite element elastic-plastic problem," *Meccanica,* Vol. 18, pp. 77–91.

Cottle, R. W., 1979, "Fundamentals of quadratic programming and linear complementarity," *Engineering Plasticity by Mathematical Programming,* M. Z. Cohn and G. Maier, eds., Pergamon Press, New York, pp. 293–323.

Cottle, R. W., and Dantzig, G. B., 1968, "Complementarity pivot theory of mathematical programming," *Mathematics of Decision Sciences,* Part I, G. B. Dantzig and A. F. Veinott, eds., American Mathematical Society, Providence, RI, pp. 115–132.

Cottle, R. W., Spang, J. S., and Stone, R. E., 1992, *The Linear Complementarity Problem,* Academic Press, San Diego, CA.

De Donato, O., and Franchi, A., 1979, "Elastic plastic analysis by finite elements," *Engineering Plasticity by Mathematical Programming,* M. Z. Cohn and G. Maier, eds., Pergamon Press, New York, pp. 413–435.

De Donato, O., and Maier, G., 1976, "Historical deformations analysis of elasto-plastic structures as a parametric linear complementarity," *Meccanica,* Vol. 11, pp. 166–171.

Dvorkin, E. N., 1996, "Finite strain elasto-plastic formulation using the method of mixed interpolation of tensorial components," *Computational Mechanics,* Vol. 18, pp. 290–301.

Feijoo, R. A., and Zouain, N., 1987, "Variational formulations for rates and increments in plasticity," *Computational Plasticity,* Part II, D. R. J. Owen, E. Hinton, and E. Oñate, eds., Pineridge Press, Swansea, UK, pp. 33–57.

Giambanco, F., 1998, "Elastic plastic analysis by the asymptotic pivoting method," *Computers & Structures,* in press.

Giambanco, F., Lo Bianco, M., and Palizzolo, L., 1994, "An Approximate Technique for Dynamic Elastic-Plastic Analysis," ASME JOURNAL OF APPLIED MECHANICS, Vol. 61, pp. 907–913.

Giambanco, F., Palizzolo, L., and Cirone, L., 1998, "Elastic plastic analysis iterative solution," *Computational Mechanics,* Vol. 21, pp. 149–160.

Gill, P. E., Murray, W., Saunders, M. A., and Wright, M. H., 1984, "QPSOL: A fortran package for quadratic programming," Technical Report SOL 84-6, Department of Operations Research, Stanford, CA.

Kojic, M., 1997, "The governing parameter method for implicit integration of viscoplastic constitutive relations for isotropic and orthotropic metals," *Computational Mechanics,* Vol. 19, pp. 49–57.

Lemke, C. E., 1968, "On complementarity pivot theory," *Mathematics of Decision Sciences,* Part I, G. B. Dantzig and A. F. Veinott, eds., American Mathematical Society, Providence, RI, pp. 95–114.

Lemke, C. E., 1980, "A survey of complementarity theory," *Variational Inequalities and Complementarity Problems,* R. W. Cottle, F. Giannessi, and J. L. Lions, eds., Wiley and Sons, Chichester, UK, pp. 213–239.

Maier, G., 1968a, "A quadratic programming approach for certain classes of nonlinear structural problems," *Meccanica,* Vol. 3, pp. 121–130.

Maier, G., 1968b, "Quadratic programming theory of elastic perfectly plastic structures," *Meccanica,* Vol. 3, pp. 31–39.

Maier, G., 1978, "Mathematical programming methods in structural analysis," *Proc. Int. Conf. on Variational Methods in Engineering,* Vol. II, Southampton University Press, Southampton, UK.

Martin, J. B., 1975, *Plasticity: Fundamentals and General Results,* The MIT Press, Cambridge, MA.

Martin, J. B., 1985, "Bounding principles and discrete time integration for statically loaded elastic plastic bodies," *Simplified Analysis of Inelastic Structures Subjected to Statical or Dynamical Loadings,* J. Zarka, ed., Laboratoire de Mécanique des Solides, Ecole Polytechnique, Palaiseau, France.

Muscolino, G., 1989, "Mode-superposition methods for elastoplastic systems," *Jour. Eng. Mech. Div., ASCE,* Vol. 115, No. 10, pp. 2199–2215.

Owen, D. R. J., 1980, "Implicity finite element methods for dynamic transient analysis of solids with particular reference to nonlinear situations," *Advanced Structural Analysis,* J. Donea, ed., Applied Science Publications, London, pp. 123–152.

Owen, D. R. J., and Hilton, E., 1980, *Finite Elements in Plasticity,* Pineridge Press, Swansea, UK.

Polizzotto, C., 1986, "Elastoplastic analysis method for dynamic agencies," *J. Eng. Mech. Div., ASCE,* Vol. 112, pp. 293–310.

Pyo, C. R., Okada, H., and Atluri, S. N., 1995, "An elastic-plastic finite element alternating method for analyzing wide-spread fatigue damage in aircraft structures," *Computational Mechanics,* Vol. 16, pp. 62–68.

Wang, L., Brust, F. W., and Atluri, S. N., 1997, "The elastic-plastic finite element alternating method (EPFEAM) and the prediction of fracture under WFD conditions in aircraft structures, Part I," *Computational Mechanics,* Vol. 19, pp. 356–369.

Wilson, E. L., Farhoomand, I., and Bathe, K. J., 1973, "Nonlinear dynamic analysis of complex structures," *Earth. Eng. and Struct. Dyn.,* Vol. 1, pp. 241–252.

Zarka, J., and Casier, J., 1981, "Elastic plastic response of a structure to cyclic loading: Practical rules," *Mechanics Today,* Vol. 6, S. Nemat-Nasser, ed., Pergamon Press, New York, pp. 93–198.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Layer-Wise Mixed Models for Accurate Vibrations Analysis of Multilayered Plates

E. Carrera
Research Professor,
Department of Aeronautics and Aerospace Engineering,
Politecnico di Torino,
Corso Duca degli Abruzzi, 24,
10129 Torino, Italy
e-mail: carrera@polito.it

This paper presents the dynamic analysis of multilayered plates using layer-wise mixed theories. With respect to existing two-dimensional theories at the displacement formulated, the proposed models a priori fulfill the continuity of transverse shear and normal stress components at each interface between two adjacent layers. A Reissner's mixed variational equation is employed to derive the differential equations, in terms of the introduced stress and displacement variables, that govern the dynamic equilibrium and compatibility of each layer. The continuity conditions at the interfaces are used to write corresponding equations at multilayered level. Related standard displacement formulations, based on the principle of virtual displacements, are given for comparison purposes. Numerical results are presented for the free-vibration response (fundamental and higher order frequencies are calculated) of symmetrically and unsymmetrically laminated cross-ply plates. Several comparisons to three-dimensional elasticity analysis and to some available results, related to both layer-wise and equivalent single-layer theories, have shown that the presented mixed models: (1) match the exact three-dimensional results very well and (2) lead to a better description in comparison to results related to other available analysis.

Introduction

Classical plate theories originally developed for traditional isotropic one-layer plates and based on Poisson-Kirchooff assumptions (Poisson, 1829; Kirchhoff, 1850) (CLT–Classical lamination Theory), have been proven to be inadequate in predicting the behavior of anisotropic plates (Yang et al., 1966). Furthermore, Reissner-Mindlin (Reissner, 1945; Mindlin, 1951) type models (FSDT—First Shear Deformation Theory) can lead to large errors in the prediction of local response of moderately thick laminated plates. In fact, three-dimensional solutions (Pagano, 1969; Srinivas et al., 1970; Noor, 1973; Noor and Burton, 1989a) have shown that both the displacement and transverse stress fields, for compatibility and equilibrium reasons, have a piecewise continuous distribution in the thickness plate direction of the multilayered structures: They reveal discontinuous derivatives in correspondence to each interface which cannot be described by CLT or FSDT type analyses. The so-called *zig-zag* form of displacement fields and the *interlaminar continuity* for the transverse stresses were summarized by Carrera (1995) with the acronym C_z^0-requirements: Displacement and transverse stress fields are C^0-continuous functions in the plate thickness direction z.

Attempts to partially introduce the C_z^0-requirements have been made both in the field of the Equivalent Single Layer Model (ESLM) and Layer-Wise Models (LWM). In comparison to LWM analyses, the ESLM ones preserve the independence of the number of the unknowns from the numbers of the layers, permitting advantageous extension to computational mechanics. Many review papers of exhaustive literature have appeared on that topic; among these one can mention those by Sun and Whitney (1973), Librescu and Reddy (1986), Reddy (1987), Kapania and Raciti (1989), and Noor and Burton (1989b). A more comprehensive and detailed overview has been provided in a recent book by Reddy (1997). A few contributions that have been considered useful by the author and which are related overall, to dynamic analyses, are discussed in the following text.

ELSM analyses were considered by Reddy and Phan (1985) who extended an original plate model by Vlasov (1957) to the dynamics of laminate structures. The used cubic in-plane displacement fields permitted the fulfillment of top and bottom plate homogeneous conditions. Further discussion and analyses were provided by Librescu and Kheider (1988). Di Sciuva (1987) introduced a linear zig-zag in-plane displacement field which accounted for interlaminar continuity. Top and bottom plate conditions were introduced in this last work by Bhaskar and Varadan (1989) while the generalization to arbitrarily laminated structures was presented by Cho and Parmerter (1993). These two last articles were restricted to static analysis. An almost equivalent approach, based on early works by Lekhnitskii (1935), was presented by Ren and Owen (1989). Due to the mechanical coupling between transverse and in-plane normal stresses, all the proposed ESLM studies could not accurately describe the transverse stress and related effects (Carrera, 1997). As a consequence the analyses have shown severe limitations in the study of the vibration of thick plates with arbitrary layouts (Nosier et al., 1992; Cho et al., 1992). The so-called predictor-corrector methods should be mentioned to complete the picture. These methods improve the ESLM results by implementing a post-processing technique to the three-dimensional equations (Noor and Burton, 1989a).

A much better description can be obtained by the use of layer-wise models. The first attempts to consider each layer in a sandwich structures as a separate beam were made by Kao and Ross (1968) and Shift and Heller (1974). Seide (1980) expanded the idea to laminated plates by considering each layer as an individual Reissner-Mindlin plate. Reddy (1987) and Nosier et al., (1992) treated each layer separately using in-plane displacements linear in the thickness direction z. Valisetty and Rehfield (1988) treated each layer separately, by employing a higher-order displacement field for flexural wave propagation

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Sept. 23, 1997; final revision, Dec. 8, 1997. Associate Technical Editor: J. N. Reddy.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

analysis in laminated plates. Cho et al. (1991) suggested using the Hildebrand et al. (1949) (applied to laminate plates by Lo et al., 1974) higher-order displacement field in each individual layer. All the mentioned LW models, in particular those by Cho et al. (1991) and Nosier et al. (1992), showed good agreement with respect to exact three-dimensional analysis to predict free-vibration response. However, all these layer-wise models did not a priori fulfill the transverse shear stress continuity: The fulfillment of the C_z^0-requirement was restricted to the displacement variables.

The limitations of the analyses based on the displacement formulation motivated the works by Reissner (1984, 1986) who proposed a mixed variational theorem for the displacements and transverse stresses. This equation was employed by Murakami (1985) and Toledano and Murakami (1987a) in the field of ESLM. Even though the transverse normal stress was considered by Toledano and Murakami (1987a), the obtained results showed severe limitations in the analysis of thick generally laminated plates. Such a conclusion motivated another work by the same authors (Toledano and Murakami, 1987b) in which a layer-wise description was introduced. The normal stress was unfortunately discarded in this work.

The author (Carrera 1995, 1997) has recently proposed mixed layer-wise models based on Reissner's equation that are able to completely fulfill the C_z^0-requirements. This means that with respect to available models, Carrera's theory a priori fulfills the interlaminar equilibrium of the transverse shear and normal stresses and permits their evaluation without requiring any post-processing process as those used in other LWM analysis or in the predictor-corrector procedures by Noor and Burton (1989a). This theory was subsequently applied to several problems related to the static analysis of multilayered thick arbitrarily laminated plates (Carrera 1998a, 1998b). The obtained results showed the superiority of the mixed models, with respect to available layer-wise analyses, to evaluate both global and local characteristics. Based on these positive responses this paper extends the mixed layer-wise models to linear dynamics. Numerical results are given for the free-vibration response of cross-ply laminated plates; it has been anticipated that the good performance of the new models will be confirmed for the evaluation of fundamental and higher-order frequencies. Bold letters are here used for arrays. Indicial notations, subscripts and superscripts, have been used extensively in order to handle, in a concise manner, the presented derivations. The Appendix quotes the explicit forms of the introduced basic arrays.

Basic Assumptions

Geometry and Materials. The geometry and Cartesian coordinate system x, y, z of the multilayered plates made of N_l layers are shown in Fig. 1. The lamina are considered homogeneous and to operate in the linear elastic range. Stiffness coefficients of Hooke's law for the anisotropic k-lamina are employed in standard form. This reads $\sigma_i^k = \tilde{C}_{ij}^k \epsilon_j^k$ where subindices i and j, ranging from 1 to 6, stand for the index couples 11, 22, 33, 13, 23, and 12, respectively. The material is assumed to be orthotropic, as specified by: $\tilde{C}_{14} = \tilde{C}_{24} = \tilde{C}_{34} = \tilde{C}_{64} = \tilde{C}_{15} = \tilde{C}_{25} = \tilde{C}_{35} = \tilde{C}_{65} = 0$. This implies that σ_{xz}^k and σ_{yz}^k depend only on ϵ_{xz}^k and ϵ_{yz}^k. In matrix form,

$$\boldsymbol{\sigma}_{pH_d}^k = \tilde{\mathbf{C}}_{pp}^k \boldsymbol{\epsilon}_{pG}^k + \tilde{\mathbf{C}}_{pn}^k \boldsymbol{\epsilon}_{nG}^k$$
$$\boldsymbol{\sigma}_{nH_d}^k = \tilde{\mathbf{C}}_{np}^k \boldsymbol{\epsilon}_{pG}^k + \tilde{\mathbf{C}}_{nn}^k \boldsymbol{\epsilon}_{nG}^k \quad (1)$$

in which subscripts p, n denote in-plane and out-of-plane (normal) components ($\boldsymbol{\sigma}_p^k = \{\sigma_{xx}^k, \sigma_{yy}^k, \sigma_{xy}^k\}$, $\boldsymbol{\sigma}_n^k = \{\sigma_{xz}^k, \sigma_{yz}^k, \sigma_{zz}^k\}$); H and G indicate values from Hooke's law and from geometrical Eq. (3) relations, respectively; and the further subscript d denotes values related to the standard displacement formulation.

For the adopted mixed solution procedure, the stress-strain relationships are conveniently put in the following mixed form:

$$\boldsymbol{\sigma}_{pH}^k = \mathbf{C}_{pp}^k \boldsymbol{\epsilon}_{pG}^k + \mathbf{C}_{pn}^k \boldsymbol{\sigma}_{nM}^k$$
$$\boldsymbol{\epsilon}_{nH}^k = \mathbf{C}_{np}^k \boldsymbol{\epsilon}_{pG}^k + \mathbf{C}_{nn}^k \boldsymbol{\sigma}_{nM}^k \quad (2)$$

where both stiffness and compliance coefficients are employed. M denotes stress from the assumed model Eq. (4). The relations between the arrays of coefficients at the two forms of Hooke's law are easily found,

$$\mathbf{C}_{pp}^k = \tilde{\mathbf{C}}_{pp}^k - \tilde{\mathbf{C}}_{pn}^k (\tilde{\mathbf{C}}_{nn}^k)^{-1} \tilde{\mathbf{C}}_{np}^k, \quad \mathbf{C}_{pn}^k = \tilde{\mathbf{C}}_{pn}^k (\tilde{\mathbf{C}}_{nn}^k)^{-1}$$
$$\mathbf{C}_{np}^k = -(\tilde{\mathbf{C}}_{nn}^k)^{-1} \tilde{\mathbf{C}}_{np}^k, \quad \mathbf{C}_{nn}^k = (\tilde{\mathbf{C}}_{nn}^k).$$

Superscripts -1 indicates the inverse of a square array.

The strain components $\boldsymbol{\epsilon}_p^k = \{\epsilon_{xx}^k, \epsilon_{yy}^k, \epsilon_{xy}^k\}$, $\boldsymbol{\epsilon}_n^k = \{\epsilon_{xz}^k, \epsilon_{yz}^k, \epsilon_{zz}^k\}$ are linearly related to the displacements $\mathbf{u}^k = \{u_x^k, u_y^k, u_z^k\}$ according to the following geometrical relations:

$$\boldsymbol{\epsilon}_{pG}^k = \mathbf{D}_p \mathbf{u}^k, \quad \boldsymbol{\epsilon}_{nG}^k = \mathbf{D}_n \mathbf{u}^k. \quad (3)$$

$\mathbf{D}_p$ and $\mathbf{D}_n$ denote in-plane and out-of-plane differential operators (see the Appendix), respectively.

Displacement and Transverse Stress Models. In order to *completely* and a priori fulfill the C_z^0-requirements a layer-wise model is assumed for both displacement $\mathbf{u}^k$ and transverse stress fields $\boldsymbol{\sigma}_n^k$. In comparison on previous works by the author, the developments are here restricted to parabolic expansion. The following expansions, in the thickness direction of each k-layer ($k = 1, \ldots, N_l$) are used:

$$\mathbf{u}^k = F_t \mathbf{u}_t^k + F_b \mathbf{u}_b^k + F_2 \mathbf{u}_2^k = F_\tau \mathbf{u}_\tau^k$$
$$\boldsymbol{\sigma}_{nM}^k = F_t \boldsymbol{\sigma}_{nt}^k + F_b \boldsymbol{\sigma}_{nb}^k + F_2 \boldsymbol{\sigma}_{n2}^k = F_\tau \boldsymbol{\sigma}_{n\tau}^k. \quad (4)$$

Subscripts t and b denote values related to the top and bottom surface of the layer, respectively. They consist of the linear part of the expansion while the F_2 polynomials permit quadratic displacement and stress fields. Contracted forms have been written by means of τ-index ($\tau = t, b, 2$); repeated indexes are summed over their ranges.

The thickness functions $F_\tau(\zeta_k)$ ($\zeta = 2z_k/h_k$ not dimensional k-layer thickness coordinates, z_k local thickness coordinate, and h_k the layer thickness, see Fig. 1) are defined by

$$F_t = \frac{P_0 + P_1}{2}, \quad F_b = \frac{P_0 - P_1}{2}, \quad F_2 = P_2 - P_0,$$

in which $P_j = P_j(\zeta_k)$ is the Legendre polynomial of the j-order ($j = 0, 1, 2$). The chosen functions have the following properties:

$$\zeta_k = \begin{cases} 1 : F_t = 1; & F_b = 0; \quad F_2 = 0 \\ -1 : F_t = 0; & F_b = 1; \quad F_2 = 0. \end{cases}$$

That is, the top and bottom values have been used as unknowns. Based on this fact, the C_z^0-requirements can be easily linked. In fact, the compatibility of the displacement and the equilibrium for the transverse stress components are written simply as

$$\mathbf{u}_t^k = \mathbf{u}_b^{(k+1)}, \quad k = 1, N_l - 1$$
$$\boldsymbol{\sigma}_{nt}^k = \boldsymbol{\sigma}_{nb}^{(k+1)}, \quad k = 1, N_l - 1. \quad (5)$$

In those cases in which the top/bottom-plate stress values are prescribed (homogeneous or nonhomogeneous conditions), the following additional C_z^0-requirements must be accounted for

$$\boldsymbol{\sigma}_{nb}^1 = \bar{\boldsymbol{\sigma}}_{nb}, \quad \boldsymbol{\sigma}_{nt}^{N_l} = \bar{\boldsymbol{\sigma}}_{nt}. \quad (6)$$

The bar denotes imposed values.

Equations (4) refer to the same order of expansion of the three components for both displacements and transverse

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

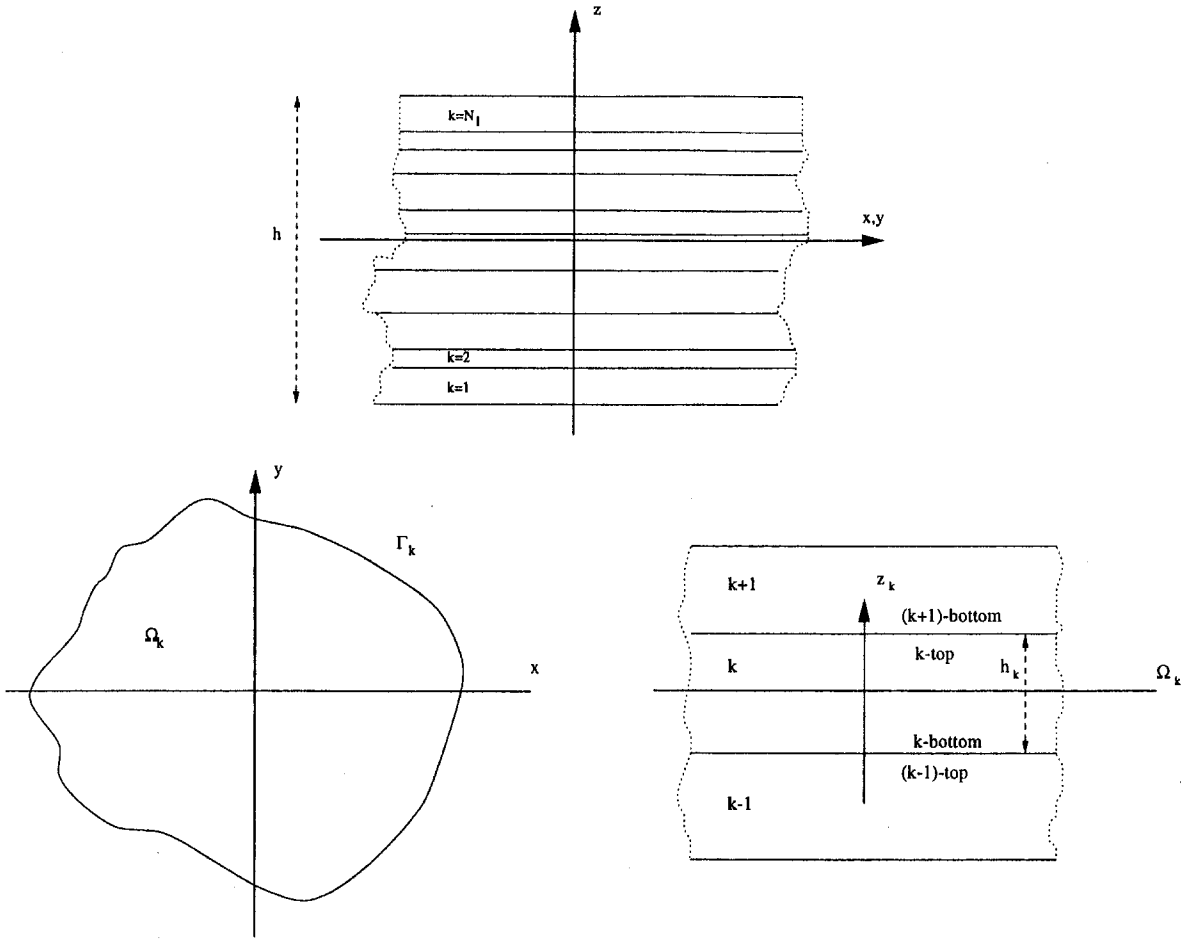


Fig. 1 Geometry and notation of multilayered plates

stresses. In order to encounter the well-known results from the literature, see, for example, Reissner (1985) or in the author's discussion reported in Carrera (1997)), different polynomial orders should be used in developments that will be presented in the subsequent sections. Trace operators that would lead to a shortening of the resulting arrays at Eqs. (15), (18), (26), and (29), could be introduced for this aims. For the sake of brevity, the results related to these aspects have not been discussed in the numerical parts as they do not change the obtained conclusions. These will be reported in future works.

Governing Equations for the k-Layer

Displacement Formulation. The classical displacement approach is formulated in terms of $\mathbf{u}^k$ via the principle of virtual displacements. In the dynamic case this states

$$\sum_{k=1}^{N_l} \int_{\Omega^k} \int_{A_k} (\delta \boldsymbol{\epsilon}_{p_G}^{k^T} \boldsymbol{\sigma}_{pH_d}^k + \delta \boldsymbol{\epsilon}_{nG}^{k^T} \boldsymbol{\sigma}_{nH_d}^k) d\Omega^k dz = \sum_{k=1}^{N_l} \int_{\Omega^k} \int_{A_k} \rho^k \delta \mathbf{u}^k \ddot{\mathbf{u}}^k dV + \delta L^e. \quad (7)$$

δ is the variational symbol and subscript T denotes transposition of arrays. A_k and V denote the layer-thickness domain and volume; Ω^k is the layer middle surface bounded by Γ^k (Γ_g^k, Γ_m^k denotes those parts of Γ^k on which the geometrical and mechanical boundary conditions are prescribed, respectively). ρ is the mass density and double dots denote acceleration. The variation of the internal work has been split into in-plane and out-of-plane parts and involves stress from Hooke's Law and strain from geometrical relations. δL_e is the virtual variation of the work made by the external layer-forces $\mathbf{p}^k = \{p_x^k, p_y^k, p_z^k\}$.

In order to handle the next developments in a concise manners, it is convenient to do as follows:

1 The displacement model Eq. (4) is substituted in the strains

$$\boldsymbol{\epsilon}_{pG}^k = \mathbf{D}_p F_\tau \mathbf{u}_\tau^k,$$
$$\boldsymbol{\epsilon}_{nG}^k = (\mathbf{D}_{n\Omega} + \mathbf{D}_{nz}) \mathbf{u}^k = \mathbf{D}_{n\Omega} F_\tau \mathbf{u}_\tau^k + F_{\tau_z} \mathbf{u}_\tau^k. \quad (8)$$

The subscript z denotes differentiation with respect to z.

2 The following stress layer-resultants are defined:

$$(\mathbf{R}_{pH_d}^{k\tau}, \mathbf{R}_{nH_d}^{k\tau}, \mathbf{R}_{nH_d}^{kt_z}) = \int_{A_k} (F_\tau \boldsymbol{\sigma}_{pH_d}^k, F_\tau \boldsymbol{\sigma}_{nH_d}^k, F_{\tau_z} \boldsymbol{\sigma}_{nH_d}^k) dz. \quad (9)$$

3 The following array formulas for the integration by parts is introduced:

$$\int_{\Omega^k} (\mathbf{D}_\Omega \phi)^T \varphi d\Omega^k = -\int_{\Omega^k} \phi^T \mathbf{D}_\Omega^T \varphi d\Omega^k + \int_{\Gamma^k} \phi^T \mathbf{I}_\Omega^T \varphi d\Gamma^k. \quad (10)$$

For the sake of simplicity, it is intended that the boundary Γ^k is parallel to the direction x, y; $\boldsymbol{\phi}$, and $\boldsymbol{\varphi}$ are two generic columns of displacements or stresses; $\mathbf{D}_\Omega$ denotes a generic array including only first-order partial differential operators with respect to the in-plane coordinates x, y. The algebraic array $\mathbf{I}_\Omega^T$ is built as follows: the unit number 1 is set in correspondence to those elements of $\mathbf{D}_\Omega$ which are different from zero (see the Appendix).

Based on this, Eq. (7) takes on the following form:

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\sum_{k=1}^{N_l}\left(\int_{\Omega^k}\delta\mathbf{u}_\tau^{kT}(-\mathbf{D}_p^T\mathbf{R}_{pH_d}^{k\tau}+\mathbf{R}_{nH_d}^{k\tau_z}-\mathbf{D}_{n\Omega}^T\mathbf{R}_{nH_d}^{k\tau})d\Omega^k\right.$$

$$\left.+\int_{\Gamma^k}\delta\mathbf{u}_\tau^{kT}(\mathbf{I}_p^T\mathbf{R}_{pH_d}^{k\tau}+\mathbf{I}_{n\Omega}^T\mathbf{R}_{nM}^{k\tau})d\Gamma^k\right)$$

$$=\sum_{k=1}^{N_l}\int_{\Omega^k}\delta\mathbf{u}_\tau^{kT}(E_{\tau s}\mathbf{I}\ddot{\mathbf{u}}^k+\mathbf{p}_\tau^k)d\Omega^k \quad (11)$$

where $\mathbf{p}_\tau^k=\{p_{x\tau}^k, p_{y\tau}^k, p_{z\tau}^k\}$ are the variationally consistent load vectors due to $\mathbf{p}^k$. By imposing the definition of virtual variations for the unknown displacements, the differential system of governing equations and related boundary conditions are derived in terms of the introduced stress resultants. For the k-layer, the equilibrium equations on Ω^k are

$$\delta\mathbf{u}_\tau^k:\ -\mathbf{D}_p^T\mathbf{R}_{pH_d}^{k\tau}+\mathbf{R}_{nH_d}^{k\tau_z}-\mathbf{D}_{n\Omega}^T\mathbf{R}_{nH_d}^{k\tau}=E_{\tau s}\mathbf{I}\ddot{\mathbf{u}}^k+\mathbf{p}_\tau^k \quad (12)$$

while the boundary conditions on Γ^k are

$$\begin{array}{cc}\text{geometrical on }\Gamma_g^k & \text{mechanical on }\Gamma_m^k\\ \mathbf{u}_\tau^k=\bar{\mathbf{u}}_\tau^k & \text{or}\quad \mathbf{I}_p^T\mathbf{R}_{pH_d}^{k\tau}+\mathbf{I}_{n\Omega}^T\mathbf{R}_{nH_d}^{k\tau}=\mathbf{I}_p^T\bar{\mathbf{R}}_{pH_d}^{k\tau}+\mathbf{I}_{n\Omega}^T\bar{\mathbf{R}}_{nH_d}^{k\tau}.\end{array} \quad (13)$$

The bar denotes imposed values at the boundary. The complete set of equations for the N_l-layers could be written simply by expanding over their ranges the introduced subscripts and superscripts.

In order to express the governing equations in terms of the displacement variables the stress resultants of Eq. (9) are written in terms of unknown variables $\mathbf{u}^k$ by substituting the Eqs. (2), (4), and (8) as follows:

$$\begin{aligned}\mathbf{R}_{pH_d}^{k\tau}&=\tilde{\mathbf{Z}}_{pp}^{k\tau s}\mathbf{D}_p\mathbf{u}_s^k+\tilde{\mathbf{Z}}_{pn}^{k\tau s}\mathbf{D}_{n\Omega}\mathbf{u}_s^k+\tilde{\mathbf{Z}}_{pn^z}^{k\tau s_z}\mathbf{D}_{n\Omega}\mathbf{u}_s^k,\\ \mathbf{R}_{nH_d}^{k\tau}&=\tilde{\mathbf{Z}}_{np}^{k\tau s}\mathbf{D}_p\mathbf{u}_s^k+\tilde{\mathbf{Z}}_{nn}^{k\tau s}\mathbf{D}_{n\Omega}\mathbf{u}_s^k+\tilde{\mathbf{Z}}_{nn^z}^{k\tau s_z}\mathbf{u}_s^k,\\ \mathbf{R}_{nH_d}^{k\tau_z}&=\tilde{\mathbf{Z}}_{np^z}^{k\tau_z s}\mathbf{D}_p\mathbf{u}_s^k+\tilde{\mathbf{Z}}_{nn^z}^{k\tau_z s}\mathbf{D}_{n\Omega}\mathbf{u}_s^k+\tilde{\mathbf{Z}}_{nn^z}^{k\tau_z s_z}\mathbf{u}_s^k.\end{aligned} \quad (14)$$

The further subscript/superscript $s=t, b, 2$ and the following layer stiffnesses and integrals have been introduced:

$$(\tilde{\mathbf{Z}}_{pp}^{k\tau s},\tilde{\mathbf{Z}}_{pn}^{k\tau s},\tilde{\mathbf{Z}}_{np}^{k\tau s},\tilde{\mathbf{Z}}_{nn}^{k\tau s})=(\tilde{\mathbf{C}}_{pp}^k,\tilde{\mathbf{C}}_{pn}^k,\tilde{\mathbf{C}}_{np}^k,\tilde{\mathbf{C}}_{nn}^k)E_{\tau s}$$

$$(\tilde{\mathbf{Z}}_{pn^z}^{k\tau s_z},\tilde{\mathbf{Z}}_{np^z}^{k\tau_z s},\tilde{\mathbf{Z}}_{nn^z}^{k\tau_z s},\tilde{\mathbf{Z}}_{nn^z}^{k\tau_z s_z})=(\tilde{\mathbf{C}}_{pn}^k E_{\tau s_z},\tilde{\mathbf{C}}_{np}^k E_{\tau_z s},\tilde{\mathbf{C}}_{nn}^k E_{\tau s_z},\tilde{\mathbf{C}}_{nn}^k E_{\tau_z s_z})$$

$$(E_{\tau s}^k,E_{\tau_z s}^k,E_{\tau s_z}^k,E_{\tau_z s_z}^k)=\int_{\Lambda_k}(F_\tau F_s,F_{\tau_z}F_s,F_\tau F_{s_z},F_{\tau_z}F_{s_z})dz \quad (15)$$

The explicit values of the integrals are given in the Appendix.

The obtained strain and stress resultants Eq. (14) can be introduced in Eqs. (12), (22), and (23) leading to equilibrium equations expressed in terms of the displacement variables

$$\delta\mathbf{u}_\tau^k:\ \mathbf{K}_d^{k\tau s}\mathbf{u}_s^k=\mathbf{M}^{k\tau s}\ddot{\mathbf{u}}_s^k+\mathbf{p}_\tau^k \quad (16)$$

the related boundary conditions are

$$\mathbf{u}_\tau^k=\bar{\mathbf{u}}_\tau^k\quad\text{or}\quad\mathbf{\Pi}_d^{k\tau s}\mathbf{u}_s^k=\mathbf{\Pi}_d^{k\tau s}\bar{\mathbf{u}}_s^k \quad (17)$$

where

$$\begin{aligned}\mathbf{K}_d^{k\tau s}&=-\mathbf{D}_p^T(\tilde{\mathbf{Z}}_{pp}^{k\tau s}\mathbf{D}_p+\tilde{\mathbf{Z}}_{pn}^{k\tau s}\mathbf{D}_{n\Omega}+\tilde{\mathbf{Z}}_{pn^z}^{k\tau s_z}\mathbf{D}_{n\Omega})-\mathbf{D}_{n\Omega}^T(\tilde{\mathbf{Z}}_{np}^{k\tau s}\mathbf{D}_p\\ &\quad+\tilde{\mathbf{Z}}_{nn}^{k\tau s}\mathbf{D}_{n\Omega}+\tilde{\mathbf{Z}}_{nn^z}^{k\tau s_z})+\tilde{\mathbf{Z}}_{np^z}^{k\tau_z s}\mathbf{D}_p+\tilde{\mathbf{Z}}_{nn^z}^{k\tau_z s}\mathbf{D}_{n\Omega}+\tilde{\mathbf{Z}}_{nn^z}^{k\tau_z s_z}\\ \mathbf{\Pi}_d^{k\tau s}&=\mathbf{I}_p^T(\tilde{\mathbf{Z}}_{pp}^{k\tau s}\mathbf{D}_p+\tilde{\mathbf{Z}}_{pn}^{k\tau s}\mathbf{D}_{n\Omega}+\tilde{\mathbf{Z}}_{pn^z}^{k\tau s_z}\mathbf{D}_{n\Omega})+\mathbf{I}_{n\Omega}^T(\tilde{\mathbf{Z}}_{np}^{k\tau s}\mathbf{D}_p\\ &\quad+\tilde{\mathbf{Z}}_{nn}^{k\tau s}\mathbf{D}_{n\Omega}+\tilde{\mathbf{Z}}_{nn^z}^{k\tau s_z})\\ \mathbf{M}^{k\tau s}&=E_{\tau s}\mathbf{I}.\end{aligned} \quad (18)$$

Mixed Formulation. In the mixed case, equilibrium and compatibility are both formulated in terms of the $\mathbf{u}^k$ and $\boldsymbol{\sigma}_n^k$ unknowns via Reissner's variational equation (Reissner, 1984, 1986; Murakami, 1986),

$$\sum_{k=1}^{N_l}\int_{\Omega^k}\int_{A_k}(\delta\boldsymbol{\epsilon}_{pG}^{kT}\boldsymbol{\sigma}_{pH}^k+\delta\boldsymbol{\epsilon}_{nG}^{kT}\boldsymbol{\sigma}_{nM}^k+\delta\boldsymbol{\sigma}_{nM}^{kT}(\boldsymbol{\epsilon}_{nG}^k-\boldsymbol{\epsilon}_{nH}^k))d\Omega^k dz$$

$$=\sum_{k=1}^{N_l}\int_{\Omega^k}\int_{A_k}\rho^k\delta\mathbf{u}^k\ddot{\mathbf{u}}^k dV+\delta L^e. \quad (19)$$

The L.H.S. includes the variations of the internal work in the plate: the first two terms come from the displacement formulation and lead to variationally consistent equilibrium conditions; the third "mixed" term variationally enforces the compatibility of the transverse strains components.

By using what has been done in the previous subsection and by introducing the following further stress and strain layer-resultants

$$(\mathbf{R}_{pH}^{k\tau},\mathbf{R}_{nM}^{k\tau},\mathbf{R}_{nM}^{kt_z},\mathbf{S}_{nG}^{k\tau},\mathbf{S}_{nH}^{k\tau})=\int_{A_k}(F_\tau\boldsymbol{\sigma}_{pH}^k,F_\tau\boldsymbol{\sigma}_M^k,F_{\tau_z}\boldsymbol{\sigma}_M^k,F_\tau\boldsymbol{\epsilon}_{nG}^k,F_\tau\boldsymbol{\epsilon}_{nH}^k)dz, \quad (20)$$

the Eq. (19) leads to

$$\sum_{k=1}^{N_l}\left(\int_{\Omega^k}(\delta\mathbf{u}_\tau^{kT}(-\mathbf{D}_p^T\mathbf{R}_{pH}^{k\tau}+\mathbf{R}_{nM}^{k\tau_z}-\mathbf{D}_{n\Omega}^T\mathbf{R}_{nM}^{k\tau})+\delta\boldsymbol{\sigma}_{n\tau}^{kT}(\mathbf{S}_{nG}^{k\tau}\right.$$

$$\left.-\mathbf{S}_{nH}^{k\tau}))d\Omega+\int_{\Gamma^k}\delta\mathbf{u}_\tau^{kT}(\mathbf{I}_p^T\mathbf{R}_{pH}^{k\tau}+\mathbf{I}_{n\Omega}^T\mathbf{R}_{nM}^{k\tau})d\Gamma^k\right)$$

$$=\sum_{k=1}^{N_l}\int_{\Omega^k}\delta\mathbf{u}_\tau^{kT}(E_{\tau s}\mathbf{I}\ddot{\mathbf{u}}^k+\mathbf{p}_\tau^k)d\Omega. \quad (21)$$

By imposing the definition of virtual variations for the unknown stress and displacement variables, the differential system of governing equations and related boundary conditions for the k-layer are obtained in terms of the introduced stress and strain resultants.

The equilibrium equations on Ω^k are

$$\delta\mathbf{u}_\tau^k:\ -\mathbf{D}_p^T\mathbf{R}_{pH}^{k\tau}+\mathbf{R}_{nM}^{k\tau_z}-\mathbf{D}_{n\Omega}^T\mathbf{R}_{nM}^{k\tau}=E_{\tau s}\mathbf{I}\ddot{\mathbf{u}}^k+\mathbf{p}_\tau^k. \quad (22)$$

The constitutive equations on Ω^k are

$$\delta\boldsymbol{\sigma}_{n\tau}^k:\ \mathbf{S}_{nG}^{k\tau}-\mathbf{S}_{nH}^{k\tau}=0. \quad (23)$$

The boundary conditions on Γ^k are

geometrical on Γ_g^k $\quad\mathbf{u}_\tau^k=\bar{\mathbf{u}}_\tau^k$

or

mechanical on Γ_m^k $\quad\mathbf{I}_p^T\mathbf{R}_{pH}^{k\tau}+\mathbf{I}_{n\Omega}^T\mathbf{R}_{nM}^{k\tau}=\mathbf{I}_p^T\bar{\mathbf{R}}_{pH}^{k\tau}+\mathbf{I}_{n\Omega}^T\bar{\mathbf{R}}_{nM}^{k\tau}$. (24)

In order to express the governing equation in terms of stress and displacement variables, the extra resultants in Eq. (9) are written in terms of the unknown variables by using Eqs. (2), (4), and (8),

$$\mathbf{R}_{pH}^{k\tau}=\mathbf{Z}_{pp}^{k\tau s}\mathbf{D}_p\mathbf{u}_s^k+\mathbf{Z}_{pn}^{k\tau s}\boldsymbol{\sigma}_{ns}^k,\quad\mathbf{R}_{nM}^{k\tau}=E_{\tau s}^k\boldsymbol{\sigma}_{ns}^k,\quad\mathbf{R}_{nM}^{k\tau_z}=E_{\tau_z s}^k\boldsymbol{\sigma}_{ns}^k,$$

$$\mathbf{S}_{nG}^{k\tau}=E_{\tau s}^k\mathbf{D}_{n\Omega}\mathbf{u}_s^k+E_{\tau s_z}^k\mathbf{u}_s^k,\quad\mathbf{S}_{nH}^{k\tau}=\mathbf{Z}_{np}^{k\tau s}\mathbf{D}_p\mathbf{u}_s^k+\mathbf{Z}_{nn}^{k\tau s}\boldsymbol{\sigma}_{ns}^k. \quad (25)$$

Further layer stiffnesses and integrals have been introduced:

$$(\mathbf{Z}_{pp}^{k\tau s},\mathbf{Z}_{pn}^{k\tau s},\mathbf{Z}_{np}^{k\tau s},\mathbf{Z}_{nn}^{k\tau s})=(\mathbf{C}_{pp}^k,\mathbf{C}_{pn}^k,\mathbf{C}_{np}^k,\mathbf{C}_{nn}^k)E_{\tau s}. \quad (26)$$

The governing equations expressed in terms of displacement and stress variables are

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\delta \mathbf{u}_\tau^k : \mathbf{K}_{uu}^{k\tau s}\mathbf{u}_s^k + \mathbf{K}_{u\sigma}^{k\tau s}\boldsymbol{\sigma}_{ns}^k = \mathbf{M}^{k\tau s}\ddot{\mathbf{u}}_s^k + \mathbf{p}_\tau^k$$

$$\delta \boldsymbol{\sigma}_{n\tau}^k : \mathbf{K}_{\sigma u}^{k\tau s}\mathbf{u}_s^k + \mathbf{K}_{\sigma\sigma}^{k\tau s}\boldsymbol{\sigma}_{ns}^k = 0 \tag{27}$$

with boundary conditions

$$\mathbf{u}_\tau^k = \bar{\mathbf{u}}_\tau^k \quad \text{or} \quad \mathbf{\Pi}_u^{k\tau s}\mathbf{u}_s^k + \mathbf{\Pi}_\sigma^{k\tau s}\boldsymbol{\sigma}_{ns}^k = \mathbf{\Pi}_u^{k\tau s}\bar{\mathbf{u}}_s^k + \mathbf{\Pi}_\sigma^{k\tau s}\bar{\boldsymbol{\sigma}}_{ns}^k \tag{28}$$

where

$$\mathbf{K}_{uu}^{k\tau s} = -\mathbf{D}_p^T\mathbf{Z}_{pp}^{k\tau s}\mathbf{D}_p \quad \mathbf{K}_{u\sigma}^{k\tau s} = -\mathbf{D}_p^T\mathbf{Z}_{pn}^{k\tau s} + E_{\tau_s}\mathbf{I} - E_{\tau s}\mathbf{D}_{n\Omega}^T$$

$$\mathbf{K}_{\sigma u}^{k\tau s} = E_{\tau s}\mathbf{D}_{n\Omega} + E_{\tau s_z}\mathbf{I} - \mathbf{Z}_{np}^{k\tau s}\mathbf{D}_p\mathbf{K}_{\sigma\sigma}^{k\tau s} = -\mathbf{Z}_{nn}^{k\tau s}$$

$$\mathbf{\Pi}_u^{k\tau s} = \mathbf{I}_p^T\mathbf{Z}_{pp}^{k\tau s}\mathbf{D}_p \quad \mathbf{\Pi}_\sigma^{k\tau s} = \mathbf{Z}_{pn}^{k\tau s} + E_{\tau s}\mathbf{I}_{n\Omega}^T. \tag{29}$$

Fulfillment of C_z^0-Requirements: Multilayered Equations

The C_z^0-requirements Eqs. (5) and (6) can be accounted for by simply writing the governing equations at the multilayered level. The necessary steps can be summarized as follows (see Carrera, 1995, for more details on similar techniques):

1 The governing equations for the N_l layers are written.
2 The continuity conditions at Eqs. (5) are linked through index contractions.

By choosing the top and 2-resultant values as unknowns, the arrays of the unknown displacements for the whole multilayer holds:

$$\mathbf{u}^T = \{\mathbf{u}_b^{1^T}, \mathbf{u}_t^{1^T}, \mathbf{u}_2^{1^T}; \mathbf{u}_t^{2^T}, \mathbf{u}_2^{2^T}; \ldots \mathbf{u}_t^{k^T}, \mathbf{u}_2^{k^T}; \mathbf{u}_t^{(k+1)^T}, \mathbf{u}_2^{(k+1)^T}; \ldots \mathbf{u}_t^{(N_l-1)^T}, \mathbf{u}_2^{(N_l-1)^T}; \mathbf{u}_{t_l}^{N^T}, \mathbf{u}_2^{N_l^T}\}. \tag{30}$$

The governing equations for the displacement formulation are formally written

$$\mathbf{K}_d\mathbf{u} = \mathbf{M}\ddot{\mathbf{u}} + \mathbf{p}$$

$$\mathbf{u} = \bar{\mathbf{u}} \quad \text{or} \quad \mathbf{\Pi}_d\mathbf{u} = \mathbf{\Pi}_d\bar{\mathbf{u}}. \tag{31}$$

What has been done for the displacement formulation can be extended to the transverse stress components and this leads to the complete fulfillment of the C_z^0-requirements. The following array for the whole multilayer is introduced

$$\boldsymbol{\sigma}_n^T = \{\boldsymbol{\sigma}_{nt}^{1^T}, \boldsymbol{\sigma}_{n2}^{1^T}; \boldsymbol{\sigma}_{nt}^{2^T}, \boldsymbol{\sigma}_{n2}^{2^T}; \ldots \boldsymbol{\sigma}_{nt}^{k^T}, \boldsymbol{\sigma}_{n2}^{k^T}; \boldsymbol{\sigma}_{nt}^{(k+1)^T}, \boldsymbol{\sigma}_{n2}^{(k+1)^T}; \ldots \boldsymbol{\sigma}_{nt}^{(N_l-1)^T}, \boldsymbol{\sigma}_{n2}^{(N_l-1)^T}; \boldsymbol{\sigma}_{n2}^{N_l^T}\}. \tag{32}$$

The governing system of differential equations at the multilayered level is formally written in the following final form:

$$\mathbf{K}_{uu}\mathbf{u} + \mathbf{K}_{u\sigma}\boldsymbol{\sigma}_n = \mathbf{M}\ddot{\mathbf{u}} + \mathbf{p} + \mathbf{p}_u^{1N_l}$$

$$\mathbf{K}_{\sigma u}\mathbf{u} + \mathbf{K}_{\sigma\sigma}\boldsymbol{\sigma}_n = \mathbf{p}_\sigma^{1N_l}. \tag{33}$$

$\mathbf{p}_u^{1N_l}$, $\mathbf{p}_\sigma^{1N_l}$ are loading due to the nonhomogeneous conditions in Eq. (6). The boundary conditions are

$$\mathbf{u} = \bar{\mathbf{u}} \quad \text{or} \quad \mathbf{\Pi}_u\mathbf{u} + \mathbf{\Pi}_\sigma\boldsymbol{\sigma}_n = \mathbf{\Pi}_u\bar{\mathbf{u}} + \mathbf{\Pi}_\sigma\bar{\boldsymbol{\sigma}}_n + \mathbf{q}_\sigma^{1N_l}. \tag{34}$$

The expanded expressions of the multilayered governing equations have been omitted for the sake of brevity. Exact closed-form solutions of the derived system of differential Eqs. (31), (33), and (34), that govern the dynamic response of a generally laminated multilayered plate, are not available. Approximated solution procedures could be conveniently implemented for this purpose (Carrera 1995). The particular case in which the material has the following properties $\tilde{C}_{16} = \tilde{C}_{26} = \tilde{C}_{36} = \tilde{C}_{45} = 0$ has her been considered. In such a case, Navier-type closed-form solutions can be found by assuming the following harmonic forms for the applied loadings and unknown variables,

Table 1 List of the acronyms used to denote three-dimensional and two-dimensional plate theories, in alphabetic order.

3D	Three Dimensional Solution
3D–1	3D Solution by Noor & Burton (1989a)
3D–2	3D Solution by Noor (1973)
3D–3	3D Solution in Kheider and Librescu (1988)
3D–4	3D Solution by Srinivas et al. (1970)
3D–5	3D Solution by Nosier et al. (1993)
CLT	Classical Lamination Theory
D-p	Present LWM at the Displacements formulated with parabolic fields in the k–layer
D-l	Present LWM at the Displacements formulated with linear fields in the k–layer
ESLM	Equivalent Single Layer Model
FSDT	First order Shear Deformation Theory
HSDT	Higher order Shear Deformation Theory of ESLM–type
HSDT–1	HSDT by Kant & Kommineni (1985)
HSDT–2	HSDT by Ren & Owen (1989)
HSDT–3	HSDT by Kheider & Librescu (1988)
HSDT–4	further HSDT by Kheider & Librescu (1988)
HSDT–5	HSDT by Reddy & Phan (1985)
LWM	Layer Wise Model
LWM-1	LWM by Cho et al. (1981)
LWM-2	LWM by Nosier et al. (1993)
M-p	Present Mixed LWM model with parabolic fields in the k–layer
M-l	Present Mixed LWM model with linear fields in the k–layer

$$(u_{x_\tau}^k, \sigma_{xz_\tau}^k, p_{x_\tau}^k) = \sum_{m,n} (U_{x_\tau}^k e^{i\omega_{mn}\hat{t}}, S_{xz_\tau}^k e^{i\omega_{mn}\hat{t}}, P_{x_\tau}^k(\hat{t})) \cos\frac{m\pi x}{a} \sin\frac{n\pi y}{b}$$

$$(u_{y_\tau}^k, \sigma_{yz_\tau}^k, p_{y_\tau}^k) = \sum_{m,n} (U_{y_\tau}^k e^{i\omega_{mn}t}, S_{yz_\tau}^k e^{i\omega_{mn}\hat{t}}, P_{y_\tau}^k(\hat{t})) \sin\frac{m\pi x}{a} \cos\frac{n\pi y}{b}$$

$$(u_{z_\tau}^k, \sigma_{zz_\tau}^k, p_{z_\tau}^k) = \sum_{m,n} (U_{z_\tau}^k e^{i\omega_{mn}\hat{t}}, S_{zz_\tau}^k e^{i\omega_{mn}\hat{t}}, P_{z_\tau}^k(\hat{t})) \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{b}, \tag{35}$$

which correspond to simply supported boundary conditions. a and b are the plate lengths in the x and y directions, respectively, while m and n are the correspondent wave numbers; $i = \sqrt{-1}$, $\hat{t}$ is the time and ω_{mn} is the circular frequency. Capital letters at R.H.S. denote correspondent maximum amplitudes. Upon substitution of Eqs. (35), the governing equations assume the form of a linear system of ordinary differential equations in the time domain. The free-vibration response leads to an eigenvalue problem: For the displacement formulation one obtains

$$\|\hat{\mathbf{K}}_d - \omega_{mn}^2\hat{\mathbf{M}}\| = 0, \tag{36}$$

while, upon eliminations of the the stress unknowns, the mixed case leads to

$$\|\hat{\mathbf{K}}_{uu} - (\hat{\mathbf{K}}_{u\sigma}(\hat{\mathbf{K}}_{\sigma\sigma})^{-1}\hat{\mathbf{K}}_{\sigma u}) - \omega_{mn}^2\hat{\mathbf{M}}\| = 0. \tag{37}$$

The double bar denotes the determinant, while the caret indicates arrays constituted by real numbers.

Numerical Illustrations

In order to evaluate the accuracy of the proposed models to analyze the free vibrational response of laminated plates, a numerical investigation related to simply supported, cross-ply laminated thick and thin plates has been conducted. Table 1 gives a list of the acronyms used to denote the theories compared in the subsequent analysis.

As a preliminary assessment, related to static bending of a cross-ply plate, Table 2 compares the results of the mixed models with the three-dimensional analysis of Noor and Burton (1989a). In-plane and transverse stresses and displacements, related to displacements and mixed solutions, are given for both

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 2 Comparison of present Analysis to three-dimensional exact solution on stress and displacement fields. Simply supported square bended plates. Cross-ply skew-symmetric laminate with $N_l = 10$, $G_{LT}/E_T = G_{Lz}/E_T = .50$, $G_{TT}/E_T = .35$, $\nu_{LT} = \nu_{Lz} = .30$, $\nu_{TT} = 0.49$.

$\frac{a}{h}$	$\frac{E_L}{E_T}$	z	3D-1	*Present Layer-Wise Analysis*			
				M-p	M-l	D-p	D-l
			Transverse displacement $u_z \times \frac{E_T}{P_{z_t}^{N_l} h}$.				
	3	t	195.2	195.2	195.2	195.2	194.8
10		b	194.8	194.8	194.8	194.8	194.4
	30	t	52.31	52.31	52.27	52.31	52.13
		b	51.92	51.92	51.88	51.99	51.71
	3	t	14.59	14.59	14.59	14.59	14.55
5		b	14.17	14.17	14.17	14.17	14.13
	30	t	6.227	6.226	6.215	6.226	6.183
		b	5.838	5.838	5.827	5.838	5.794
			In-plane displacement $u_y \times \frac{E_T}{P_{z_t}^{N_l} h}$.				
10	3	t	-30.45	-30.45	-30.45	-30.45	-30.38
		b	27.40	27.40	27.40	27.40	27.33
	30	t	-7.017	-7.017	-7.011	-7.171	-6.985
		b	5.561	5.561	5.553	5.561	5.534
5	3	t	-3.784	-3.784	-3.784	-3.784	-3.769
		b	3.499	3.499	3.498	3.499	3.485
	30	t	-1.092	-1.092	-1.089	-1.092	-1.077
		b	0.767	0.767	0.763	0.767	0.754
			In-plane stress $\sigma_{yy} \times \frac{1}{P_{z_t}^{N_l}}$.				
10	3	t	13.02	13.02	13.02	13.02	13.71
		b	-29.62	-29.62	-29.62	-29.62	-30.17
	30	t	3.225	3.225	3.227	3.230	3.357
		b	-53.24	-53.24	-53.17	-53.24	-53.11
5	3	t	3.631	3.631	3.631	3.637	3.821
		b	-7.573	-7.576	-7.571	-7.578	-7.713
	30	t	1.326	1.326	1.323	1.331	1.372
		b	-14.71	-14.71	-14.63	-14.71	-14.51
			Transverse shear stress $\sigma_{yz} \times \frac{1}{P_{z_t}^{N_l}}$.				
10	3	c	2.380	2.380	2.370	2.380	2.375
	30	c	2.344	2.344	2.341	2.344	2.343
5	3	c	1.185	1.185	1.179	1.185	1.182
	30	c	1.141	1.141	1.383	1.141	1.142

cases of parabolic and linear fields in Eqs. (4). $P_{zt}^{N_l}$ denotes the amplitude of the applied z-transverse pressure applied in correspondence to the t-top surface of the N_l-layer. Two values of thickness ratio a/h (h is the plate thickness) and orthotropic ratio E_L/E_T are investigated (standard notations are used to denote mechanical properties of the lamina, see Jones, 1975); t, b, c (plate-top, bottom, and center, respectively) denote the location in the z-direction of the quoted values. The obtained results show the excellent agreement between the present mixed parabolic (M-p) solutions and the three-dimensional analysis. Some minor comments can be made: (1) in some cases the results related to the mixed linear (M-1) cases can be more accurate than the parabolic case related to the displacement formulation (D-p); (2) the D-1 analysis leads to the least accurate description; (3) as expected as a/h and E_L/E_T decrease, the discrepancies between the different modelings increase.

Dynamic responses are treated in the subsequent tables. Table 3 compares the values of the fundamental ($m = n = 1$) frequency parameter of different theories. The plate analyzed in

Table 3 Comparison of present analysis to three-dimensional exact and ESLM available solution on fundamental circular frequency parameter $\bar{\omega}_1 = \omega h\sqrt{\rho/E_T}$ simply supported square plates $a/h = 5$. Cross-ply skew-symmetric and symmetric laminates. The same material of Table 2 is used.

N_l	$\frac{E_L}{E_T}$	3D-1	*Present Layer-Wise Analysis*				ESLM (Kant et alii 1989)	
		[23]	M-p	M-l	D-p	D-l	FSDT	HSDT-1
2	3	0.2392	0.2394	0.2417	0.2394	0.2478	0.2379	0.2388
	30	0.3117	0.3142	0.3134	0.3168	0.3209	0.3165	0.3117
10	3	0.2530	0.2530	0.2530	0.2530	0.2534	0.2527	0.2531
	30	0.4027	0.4027	0.4031	0.4027	0.4042	0.4086	0.4028
3	3	0.2516	0.2516	0.2511	0.2517	0.2556	-	-
	30	0.3739	0.3749	0.3721	0.3763	0.3807	-	-
9	3	0.2535	0.2535	0.2535	0.2535	0.2540	0.2531	0.2536
	30	0.4040	0.4040	0.4044	0.4040	0.4058	0.4067	0.4067

Table 4 Effect fictitious layers on fundamental circular frequency parameter $\bar{\omega}_1 = \omega h\sqrt{\rho/E_T}$. Simply supported square plates $a/h = 5$. Cross-ply skew-symmetric with $N_l = 2$ laminates. The same material of Table 2 is used.

$\frac{E_L}{E_T}$	N_l^*	3D-1	*Present Analysis*			
			M-p	M-l	D-p	D-l
3	2×1	0.2392	0.2394	0.2417	0.2394	0.2478
	2×1		0.2392	0.2396	0.2392	0.2416
	2×4		0.2392	0.2392	0.2392	0.2398
	2×5		0.2392	0.2392	0.2392	0.2395
30	2×1	0.3117	0.3142	0.3134	0.3168	0.3209
	2×2		0.3119	0.3129	0.3121	0.3179
	2×4		0.3117	0.3118	0.3117	0.3135
	2×5		0.3117	0.3117	0.3117	0.3128

Table 2 is referred to. Two symmetric ($N_l = 3, 9$) and two antisymmetric ($N_l = 2, 10$) cross-ply plates are investigated with two different values of the orthotropic ratio. It has been confirmed that the M-p analysis leads to the best description, with respect to three-dimensional solutions. The results related to the ESLM analysis by Kant and Kommineni (1989) are also compared (the HSDT-1 model includes a cubic displacement field but it does not include the zig-zag effects nor fulfill the interlaminar equilibrium). As expected the accuracy of the ESLM analyses are very much subordinate to the layouts and material properties, while present layer-wise models are not affected by these factors.

Layer-wise results can be improved by introducing fictitous interfaces. Such a possibility is discussed in Table 4, where two cases of the previous analysis are treated. N_l^* is the number of the layer used in the calculations. The same number of fictitious interfaces has been considered in each layer. For instance, $N_l^* = N_l \times 2$ signifies that each phisical layer has been simulated by means of two fictitious layers. N_l^* increasing the three-dimensional analysis is approached. Mixed models show higher convergence rates with respect to those of the displacement models.

Further comparisons with available three-dimensional and to ESLM solutions are considered in Table 5. HSDT-2 describes the zig-zag form of the in-plane displacements and fulfills transverse shear continuity. HSDT-3 and HSDT-4 consist of the two different higher-order models described by Kheider and Librescu, (1989). The FSDT and CLT values are also those quoted in this last paper. The three-dimensional-2 and three-dimensional-3 solutions are based on approximated finite difference technique and exact closed-form solutions, respectively. The conclusions reached for the previous analysis have been confirmed.

Table 6 compares the present analysis with that by Reddy and Phan, (1985). Thickness ratios from 2 to 1000 are considered. HSDT-5 fulfills the top and bottom zero-transverse shear plate conditions. The ESLM analysis leads to very large errors for the very thick plate cases. CLT solution is obtained for thin

Table 5 Comparison of present analysis to three-dimensional solutions and available ESLM analyses on fundamental circular frequency parameter $\bar{\omega}_2 = 10 \times \omega h\sqrt{\rho/E_T}$ simply supported square plates $a/h = 5$. Cross-ply skew-symmetric and symmetric laminates. $G_{LT}/E_T = G_{Lz}/E_T = .60$, $G_{TT}/E_T = .50$, $\nu_{LT} = \nu_{Lz} = \nu_{TT} = 0.25$.

Model	N_l	2		10		3		9	
	$\frac{E_L}{E_T}$	3	30	3	30	3	30	3	30
3D-2		2.5031	3.2705	2.6583	4.4011	2.6474	4.1089	2.6640	4.4210
3D-3		-	-	-	-	2.6293	4.1087	2.6392	4.4193
M-p		2.4950	3.2826	2.6382	4.3670	2.6315	4.0919	2.6438	4.3827
M-l		2.5174	3.2788	2.6383	4.3711	2.6358	4.0738	2.6434	4.3916
D-p		2.4954	3.3042	2.6382	4.3670	2.6312	4.1032	2.6438	4.3867
D-l		2.5350	3.3255	2.6407	4.3825	2.6492	4.1526	2.6468	4.4058
HSDT-2		2.4128	3.2529	2.6308	4.3631	2.5560	3.9311	2.6356	4.3582
HSDT-3		-	-	-	-	2.6292	3.9387	2.6436	4.3847
HSDT-4		-	-	-	-	2.6223	3.9301	2.6375	4.3788
FSDT		-	-	-	-	2.6252	3.9540	2.6376	4.3818
CLT		-	-	-	-	3.0143	6.6419	3.0143	6.6419

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 6 Comparison of present analysis to ESLM solution by Reddy and Phan (1985) on fundamental circular frequency parameter $\bar{\omega}_3 = \omega\sqrt{a^4\rho/E_T h^2}$ simply supported square plates. Cross-ply skew-symmetric and symmetric laminates. $E_L/E_T = 40$, $G_{LT}/E_T = G_{Lz}/E_T = .50$, $G_{TT}/E_T = .60$, $\nu_{LT} = \nu_{Lz} = \nu_{TT} = 0.25$.

a/h	*Present Layer-Wise Analysis*				ESLM (Reddy et alii, 1985)		
	M-p	M-l	D-p	D-l	CLT	FSDT	HSDT-5
	stacking sequences 0/90						
2	4.774	4.772	4.804	4.848	8.499	5.191	5.699
4	7.450	7.424	7.528	7.562	10.292	7.975	8.294
5	8.288	8.261	8.355	8.397	10.584	8.757	9.010
10	10.141	10.128	10.178	10.215	11.011	10.355	10.449
20	10.876	10.872	10.888	10.921	11.125	10.941	10.968
50	11.116	11.115	11.118	11.150	11.158	11.127	11.132
100	11.152	11.152	11.152	11.184	11.163	11.155	11.156
1000	11.164	11.164	11.164	11.164	-	-	-
	stacking sequence 0/90/90/0						
2	5.270	5.320	5.277	5.414	15.830	5.492	5.576
4	9.230	9.343	9.235	9.472	17.907	9.369	9.497
5	10.753	10.876	10.756	11.005	18.215	10.820	10.989
10	15.150	15.248	15.152	15.334	18.652	15.083	15.270
20	17.626	17.668	17.627	17.702	18.767	17.583	17.668
50	18.600	18.607	18.600	18.617	18.799	18.590	17.606
100	18.753	18.755	18.753	18.761	18.804	18.751	18.755
1000	18.805	18.805	18.805	18.805	-	-	-

plates. As already observed by Cho et al., (1992), HSDT results can result to be less accurate than FSDT solutions. However, this last theory requires the use of a shear correction factor whose value is not a priori known (Noor and Burton 1989a). A comparison with present layer-wise analysis confirms that the accuracy of ESLM results are layout dependent. They are in particular more accurate in the case of symmetric laminates.

Higher frequencies related to higher modes are considered in Table 7 for an isotropic plate. A comparison with the aforementioned ESLM results and with LWM-1 by Cho et al. (1992) has been performed.

The effect of fictitous interfaces has also been evaluated. As the plate is isotropic, the mixed and displacement formulations lead to the same results. One should notice that the symmetric modes are better evaluated than antisymmetric ones. A slower convergence rate is found for higher antisymmetric modes. N_l^* increasing the present layer-wise model shows a better convergence rate than LWM-1 analysis.

Table 7 Comparison of present analysis to three-dimensional exact and other LWM and ESLM solutions on the lowest five circular frequency parameter $\bar{\omega}_4 = \omega h\sqrt{\rho/G}$. Simply supported square isotropic plates ($\nu = 0.3$). A and S denote modes which are antisymmetric and symmetric about the midplane. I-S and II-S are both thickness-twist modes.

Model	N_l^*	I-A	I-S	II-S	II-A	III-A
	(mh/a =0.1, nh/a =0.1)					
3D-4	-	0.0931	0.4443	0.7498	3.1729	3.2465
LWM-1	1	0.0930	0.4443	0.7416	3.1933	3.2756
LWM-1	4	0.0931	0.4443	0.7499	3.1736	3.2496
HSDT-5	-	0.0931	-	-	-	-
FSDT	-	0.0930	-	-	3.1729	3.2538
M-p & D-p	1	0.0934	0.4443	0.7498	3.4925	3.5699
M-l & D-l	1	0.1029	0.4443	0.7502	3.4985	3.5855
M-p & D-p	2	0.0932	0.4443	0.7498	3.1845	3.2587
M-l & D-l	2	0.0959	0.4443	0.7499	3.4925	3.5744
M-p & D-p	4	0.0931	0.4443	0.7498	3.1736	3.2437
M-l & D-l	4	0.0939	0.4443	0.7499	3.2533	3.4291
	(mh/a =0.2, nh/a =0.2)					
3D-4	-	0.3421	0.8886	1.4923	3.2648	3.5298
LWM-1	1	0.3404	0.8886	1.5066	3.2847	3.5831
LWM-1	4	0.3416	0.8886	1.4932	3.2656	3.5398
HSDT-5	-	0.3411	-	-	-	-
FSDT	-	0.3402	-	-	3.2648	3.5580
M-p & D-p	1	0.3456	0.8886	1.4925	3.5763	3.8589
M-l & D-l	1	0.3763	0.8886	1.4959	3.7627	3.9257
M-p & D-p	2	0.3423	0.8886	1.4923	3.2762	3.5436
M-l & D-l	2	0.3537	0.8886	1.4934	3.5762	3.8748
M-p & D-p	4	0.3421	0.8886	1.4923	3.2656	3.5307
M-l & D-l	4	0.3452	0.8886	1.4925	3.3431	3.6172

Table 8 Comparison of present mixed analysis to three-dimensional exact and other LWM and ESLM solutions on the lowest ten circular frequency parameters $\bar{\omega}_1 = \omega h\sqrt{\rho/E_T}$ simply supported square plates $a/h = 10$. Cross-ply symmetric laminates 0/90/0. $E_L = 35.1 \times 10^6$ psi, $E_T = 4.8 \times 10^6$ psi, $E_z = 0.75 \times 10^6$ psi, $G_{LT} = 1.36 \times 10^6$ psi, $G_{Lz} = 1.2 \times 10^6$ psi, $G_{Tz} = 0.47 \times 10^6$ psi, $\nu_{LT} = 0.036$, $\nu_{Lz} = 0.25$, $\nu_{TT} = 0.171$.

3D-5	Present M-p		LWM-2	ESLM (Nosier et alii, 1993)		
	$N_l \times 1$	$N_l \times 3$	$N_l \times 6$	HSDT-5	FSDT	CLT
			(m=n=1)			
0.06715	0.06716	0.06715	0.06716	0.06839	0.06931	0.07769
0.50350	0.50349	0.50349	0.50355	0.50897	0.50897	0.50897
0.63775	0.63775	0.63775	0.63782	0.64174	0.64174	0.64174
1.2429	1.2438	1.2429	1.2445	1.4106	1.2719	
1.2790	1.2804	1.2790	1.2812	1.4638	1.5845	
1.3292	1.3305	1.3292	1.3301			
2.1533	2.1459	2.1533	2.1621			
2.4894	2.4943	2.4895	2.5022			
2.7419	2.7556	2.7422	2.7612			
3.5416	3.7728	3.5437	3.5917			
			(m=n=2)			
0.20798	0.20809	0.20798	0.20808	0.21526	0.22055	0.31077
0.97517	0.97516	0.97517	0.97560	1.0179	1.0179	1.0179
1.2034	1.2040	1.2035	1.2047	1.2835	1.2835	1.2835
1.2916	1.2920	1.2916	1.2927	1.5695	1.4343	
1.4326	1.4336	1.4326	1.4348	1.9300	2.0309	
1.8037	1.8065	1.8037	1.8048			
2.3653	2.3560	2.3653	2.3740			
2.4877	2.4938	2.4878	2.5008			
2.9047	2.9176	2.9051	2.9237			
3.6361	3.7688	3.6382	3.6864			

Finally, Table 8 compares the present layer-wise mixed models with those by Nosier et al. (1993). Fundamental and higher frequencies related to two half-waves modes are considered for a symmetrically cross-ply laminated plate. The effects of a fictitious interfaces is also analyzed. The present mixed analysis shows results that are very close to the three-dimensional exact solutions. The quoted frequencies, especially for the higher mode cases, are much better evaluated using present models than those by Nosier et al. (1993). This results once again confirms the effectiveness of the a priori fulfillment of the transverse stress continuity, i.e., the superiority of the present mixed analysis with respect to other layer-wise theories based on displacement formulations.

Concluding Remarks

By employing a Reissner's mixed variational equation, this paper has presented the differential equations that govern the dynamics of multilayered plates modeled using mixed layer-wise theories. Classical displacement formulations are also derived for comparison purposes. A numerical investigation has been conducted for the free-vibration response of cross-ply laminated, simply supported, thick and thin plates for which closed-form solutions are given.

From a two-dimensional modeling point of view, one can notice that the proposed mixed theory consists of the unique layer-wise analysis which permits one to completely and a priori fulfill the C_z^0-continuity conditions at the interfaces for both displacement and transverse stress components (transverse displacement and transverse normal stress included).

Nevertheless, the conducted numerical analyses have shown, by means of several comparisons, that the mixed models lead to a better description than to standard displacement ones. Furthermore, the related results match the three-dimensional elasticity solutions very well and their accuracy has been confirmed for symmetrically, unsymmetrically, as well as arbitrarily laminated plates.

Acknowledgments

The author is deeply indebted to the late Prof. Eric Reissner for his encouragement to work on the subject of this paper.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Further acknowledgments are directed to the late Prof. Placido Cicala for the many suggestions and criticisms given during the review process of the paper (Carrera, 1995).

References

Bhashar, B., and Varadan, T. K., 1989, "Refinement of Higher-Order laminated plate theories," *American Institute of Aeronautics and Astronautics Journal,* Vol. 27, pp. 1830–1831.

Carrera, E., 1995, "A class of Two Dimensional Theories for Multilayered Plates Analysis," *Atti Academia delle Scienze di Torino, Mem Sci. Fis.,* Vol. 19–20, pp. 49–87.

Carrera, E., 1997, "C_z^0-Requirements: Models for the Two-Dimensional Analysis of Multilayered Structures," *Composite Structures,* Vol. 37, pp. 373–384.

Carrera, E., 1998a, "Mixed Layer-Wise Models for Multilayered Plates Analyses," preprint in DIAS Report No. 1/98, Politecnico di Torino, Torino, Italy.

Carrera, E., 1998b, "Evaluation of Mixed Layer-Wise Models for Multilayered Plates Analyses," *American Institute of Aeronautics and Astronautics Journal,* Vol. 36, No. 5, pp. 830–839.

Cho, K. N., Bert, C. W., and Striz, A. G., 1991, "Free Vibrations of Laminated Rectangular Plates Analyzed by Higher order Individual-Layer Theory," *Journal of Sound and Vibration,* Vol. 145, pp. 429–442.

Cho, M., and Parmerter, R. R., 1993, "Efficient higher order composite plate theory for general lamination configurations," *American Institute of Aeronautics and Astronautics Journal,* Vol. 31, pp. 1299–1305.

Di Sciuva, M., 1987, "An Improved Shear Deformation Theory for Moderately Thick Multilayered Anisotropic Shells and Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 54, pp. 589–586.

Hildebrand, F. B., Reissner, E., and Thomas, G. B., "Notes on the foundations of the theory of small displacement of orthotropic shells," NACA TN-1833, Washington, DC.

Jones, R. M., 1975, *Mechanics of Composite Materials,* McGraw-Hill, New York.

Kant, T., and Kommineni, J. R., 1989, "Large Amplitude Free Vibration Analysis of Cross-Ply Composite and Sandwich Laminates with a Refined Theory and C^0 Finite Elements," *Computer & Structures,* Vol. 50, pp. 123–134.

Kapania, R. K., and Raciti, S., 1989, "Recent advances in analysis of laminated beams and plates," *American Institute of Aeronautics and Astronautics Journal,* Vol. 27, pp. 923–946.

Kao, J., and Ross, R. J., 1968, "Bending of Multilayer sandwich beams," *American Institute of Aeronautics and Astronautics Journal,* Vol. 6, pp. 1583–1585.

Kheider, A. A., and Librescu, L., 1988, "Analysis of Symmetric Cross-ply Laminated Elastic Plates Using a Higher order Theory: Part II—Buckling and Free Vibration," *Composite Structures,* Vol. 9, pp. 259–277.

Kirchhoff, G., 1850, "Über das Gleichwich und die Bewegung einer Elastischen Sheibe," *J. Angew. math.,* Vol. 40, pp. 51–88.

Lekhnitskii, S. G., 1935, "Strength Calculation of Composite Beams," *Vestnik inzhen. i tekhnikov,* No. 9.

Librescu, L., and Reddy, J. N., 1987 "A Critical Review and Generalization of Transverse Shear Deformable Anisotropic Plates," *Refined dynamical Theories of Beams, Plates and Shells and their Applications,* Euromech Colloquium 219, Kassel, Sept. 1996, I. Ellshakoff and H. Irretier, eds., Springer-Verlag, Berlin, pp. 32–43.

Lo, K. H., Christensen, R. M., and Wu, E. M., 1977, "A Higher-Order Theory of Plate Deformation, Part II: Laminated Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 44, pp. 669–676.

Mindlin, R. D., 1951, "Influence of Rotatory Inertia and Shear on Flexural Motions of Isotropic Elastic Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 18, pp. 1031–1036.

Mindlin, R. D., Schacknow, A., and Deresiewicz, H., 1956, "An exact analysis for vibration of simply supported homogeneous and laminated thick rectangular plates," *Journal of Sound and Vibration,* Vol. 12, pp. 187–195.

Murakami, H., 1986, "Laminated Composite Plate Theory With Improved In-Plane Response," ASME JOURNAL OF APPLIED MECHANICS, Vol. 53, pp. 661–666.

Noor, A. K., 1973, "Free Vibration of Multilayered Plates," *American Institute of Aeronautics and Astronautics Journal,* Vol. 11, pp. 1038–1039.

Noor, A. K., and Burton, W. S., 1989a, "Stress and Free Vibration Analyses of Multilayered Composite Plates," *Composite Structures,* Vol. 11, pp. 183–204.

Noor, A. K., and Burton, W. S., 1989b, "Assessment of shear deformation theories for multilayered composite plates," ASME *Applied Mechanics Reviews,* Vol. 41, pp. 1–18.

Nosier, A., Kapania, R. K., and Reddy, J. N., 1993, "Free Vibration Analysis of Laminated Plates Using a Layer-Wise Theory," *American Institute of Aeronautics and Astronautics Journal,* Vol. 31, pp. 2335–2346.

Pagano, N. J., 1969, "Exact solutions for Composite Laminates in Cylindrical Bending," *Journal of Composite Materials,* Vol. 3, pp. 398–411.

Poisson, S. D., 1829, "Memoire sur l'equilibre et le mouvement des corps elastique," *Mem. Acad. Sci.,* Vol. 8, p. 357.

Reddy, J. N., and Phan, N. D., 1985, "Stability and Vibration of Isotropic, Orthotropic, and Laminated Plates According to a Higher order Shear Deformation Theory," *Journal of Sound and Vibration,* Vol. 98, pp. 157–170.

Reddy, J. N., 1987, "A generalization of Two-Dimensional Theories of Laminated Composite Plates," *Communication in Applied Numerical Methods,* Vol. 3, pp. 173–180.

Reddy, J. N., 1997, *Mechanics of Laminated Composite Plates: Theory and Analysis,* CRC Press, Boca Raton, FL.

Reissner, E., 1945, "The Effects of Transverse Shear Deformation on the Bending of Elastic Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 12, pp. 69–76.

Reissner, E., 1984, "On a certain mixed variational theory and a proposed applications," *International Journal of Numerical Methods in Engineering,* Vol. 20, pp. 1366–1368.

Reissner, E., 1985, "Reflections on the theory of elastic plates," ASME *Applied Mechanics Reviews,* Vol. 38, pp. 1433–1464.

Reissner, E., 1986, "On a mixed variational theorem and on a shear deformable plate theory," *International Journal of Numerical Methods in Engineering,* Vol. 23, pp. 193–198.

Ren, J. G., and Owen, D. R. J., 1989, "Vibration and Buckling of Laminated Plates," *International Journal of Solids and Structures,* Vol. 23, pp. 111–131.

Seide, P., 1980, "An approximate theory for the bending of laminated plates," *Mechanics Today,* Vol. 5, pp. 451–466.

Srinivas, S., Joga Rao, C. V., and Rao, A. K., 1970, "Flexural Vibration of Rectangular Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 23, pp. 430–436.

Srinivas, S., 1973, "A refined analysis of composite laminates," *Journal of Sound and Vibration,* Vol. 30, pp. 495–50.

Sun, C. T., and Whitney, J. M., 1973, "On the theories for the dynamic response of laminated plates," *American Institute of Aeronautics and Astronautics Journal,* Vol. 11, pp. 372–398.

Swift, G. W., and Heller, R. A., 1974, "Layered beams analysis," *ASCE Journal of Engineering Mechanics Division,* Vol. 100, pp. 267–282.

Toledano, A., and Murakami, H., 1987a, "A Higher-Order Laminated Plate Theory with Improved In-Plane Responses," *International Journal of Solids and Structures,* Vol. 23, pp. 111–131.

Toledano, A., and Murakami, H., "A composite plate theory for arbitrary laminated configurations," ASME JOURNAL OF APPLIED MECHANICS, Vol. 54, pp. 181–189.

Valisetty, R. R., and Rehfield, L. W., 1988, "Application of ply level analysis to flexure wave vibration," *Journal of Sound and Vibration,* Vol. 126, pp. 183–194.

Vlasov, B. F., 1957, "Ob uravnieniakh isgiba palstinok (On the equations of bending of plates)," Dokla Ak. Nauk Azerbeijanskoi-SSR, Vol. 3, pp. 955–979.

Yang, P. C., Norris, C. H., and Stavsky, Y., 1966, "Elastic Wave Propagation in Heterogenous Plates," *International Journal of Solids and Structures,* Vol. 2, pp. 85–97.

APPENDIX

Explicit Form of Basic Arrays

The arrays in the mixed Hooke's law referred to a generic fiber orientation with respect to the x-axis are

$$\tilde{\mathbf{C}}_{pp}^k = \begin{bmatrix} \tilde{C}_{11}^k & \tilde{C}_{12}^k & \tilde{C}_{16}^k \\ \tilde{C}_{12}^k & \tilde{C}_{22}^k & \tilde{C}_{26}^k \\ \tilde{C}_{16}^k & \tilde{C}_{26}^k & \tilde{C}_{66}^k \end{bmatrix}; \quad \tilde{\mathbf{C}}_{pn}^k = \tilde{\mathbf{C}}_{np}^{kT} \begin{bmatrix} 0 & 0 & \tilde{C}_{13}^k \\ 0 & 0 & \tilde{C}_{23}^k \\ 0 & 0 & \tilde{C}_{36}^k \end{bmatrix};$$

$$\tilde{\mathbf{C}}_{nn}^k = \begin{bmatrix} \tilde{C}_{44}^k & \tilde{C}_{45}^k & 0 \\ \tilde{C}_{45}^k & \tilde{C}_{55}^k & 0 \\ 0 & 0 & \tilde{C}_{33}^k \end{bmatrix}.$$

The introduced arrays of differential operators on Ω^k are

$$\mathbf{D}_p = \begin{bmatrix} \partial_x & 0 & 0 \\ 0 & \partial_y & 0 \\ \partial_y & \partial_x & 0 \end{bmatrix}; \quad \mathbf{D}_n = \begin{bmatrix} \partial_z & 0 & \partial_x \\ 0 & \partial_z & \partial_y \\ 0 & 0 & \partial_z \end{bmatrix};$$

$$\mathbf{D}_{n\Omega} = \begin{bmatrix} 0 & 0 & \partial_x \\ 0 & 0 & \partial_y \\ 0 & 0 & 0 \end{bmatrix}; \quad \mathbf{D}_{nz} = \begin{bmatrix} \partial_z & 0 & 0 \\ 0 & \partial_z & 0 \\ 0 & 0 & \partial_z \end{bmatrix}$$

related not differential operators are

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{I}_p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix};$$

$$\mathbf{I}_{n\Omega} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

The integrals at the Eqs. (15) related to a linear field in the k-layer hold,

$$E^k_{tt} = h_k\tfrac{1}{3},\quad E^k_{tb} = h_k\tfrac{1}{6},\quad E^k_{bt} = h_k\tfrac{1}{6},\quad E^k_{bb} = h_k\tfrac{1}{3}$$

$$E_{t_zt} = \tfrac{1}{2},\quad E_{t_zb} = \tfrac{1}{2},\quad E_{b_zt} = -\tfrac{1}{2},\quad E_{b_zb} = -\tfrac{1}{2}$$

$$E_{tt_z} = \tfrac{1}{2},\quad E_{tb_z} = -\tfrac{1}{2},\quad E_{bt_z} = \tfrac{1}{2},\quad E_{bb_z} = -\tfrac{1}{2}$$

$$E^k_{t_zt_z} = \frac{1}{h_k},\ E^k_{t_zb_z} = -\frac{1}{h_k},\quad E^k_{b_zt_z} = -\frac{1}{h_k},\quad E^k_{b_zb_z} = \frac{1}{h_k}.$$

For the parabolic case the F_2 functions are also involved,

$$E^k_{tt} = \frac{h_k}{3},\quad E^k_{t2} = -\frac{h_k}{2},\quad E^k_{tb} = \frac{h_k}{6},$$

$$E^k_{2t} = -\frac{h_k}{2},\quad E^k_{22} = \frac{6h_k}{5},\quad E^k_{2b} = -\frac{h_k}{2},$$

$$E^k_{bt} = \frac{h_k}{6},\quad E^k_{b2} = -\frac{h_k}{2},\quad E^k_{bb} = \frac{h_k}{3}$$

$$E_{t_zt} = \tfrac{1}{2},\quad E_{t_z2} = -1,\quad E_{t_zb} = \tfrac{1}{2},$$

$$E_{2_zt} = 1,\quad E_{2_z2} = 0,\quad E_{2_zb} = -1,$$

$$E_{b_zt} = -\tfrac{1}{2},\quad E_{b_z2} = 1,\quad E_{b_zb} = -\tfrac{1}{2},$$

$$E_{tt_z} = \tfrac{1}{2},\quad E_{t2_z} = 1,\quad E_{tb_z} = -\tfrac{1}{2},$$

$$E_{2t_z} = -1,\quad E_{22_z} = 0,\quad E_{2b_z} = 1,$$

$$E_{bt_z} = \tfrac{1}{2},\quad E_{b2_z} = -1,\quad E_{bb_z} = -\tfrac{1}{2},$$

$$E^k_{t_zt_z} = \frac{1}{h_k},\quad E^k_{t_z2_z} = 0,\quad E^k_{t_zb_z} = \frac{-1}{h_k},$$

$$E^k_{2_zt_z} = 0,\quad E^k_{2_z2_z} = \frac{12}{h_k},\quad E^k_{2_zb_z} = 0.$$

$$E^k_{b_zt_z} = \frac{-1}{h_k},\quad E^k_{b_z2_z} = 0,\quad E^k_{b_zb_z} = \frac{1}{h_k}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Stress Singularity Analysis of a Crack Terminating at the Interface of an Anisotropic Layered Composite

P. Poonsawat
Graduate Student

A. C. Wijeyewickrema
Associate Professor

P. Karasudhi
Chair Professor

School of Civil Engineering,
Asian Institute of Technology,
Bangkok, Thailand

The order of stress singularities at the tip of an inclined crack terminating at the interface of an anisotropic layered composite is investigated. Both fully bonded and frictional interfaces are considered. The expressions for stresses and displacements are obtained by using the Stroh formalism. The stresses at the crack tip are expressed in the form $\sigma_{ij} = r^{-k}F_{ij}(\theta)$, where k is the crack-tip singularity. The singularity k is obtained by solving a characteristic equation which incorporates the effects of the interface and the crack faces. The problem can be visualized as two wedges created by a crack, pressing on a half-plane. For the frictional interface, depending on the relative slip directions of the two wedges, both the case of the two wedges slipping in opposite directions and the case of the two wedges slipping in the same direction are treated. In the numerical calculation of the singularities, a high modulus graphite/epoxy layered composite is used and the effect of the crack inclination on the stress singularity k is graphically presented. In general, there are three roots of k for the fully bonded interface, while there are only two roots of k for the slipping interface.

1 Introduction

Layered composites are widely used in many applications due to many advantages such as high strength, good fracture toughness, lightweightness, etc. An angle-ply layered composite consists of many layers of fiber-reinforced material as shown in Fig. 1. The fibers in all the layers lie parallel to the interface plane (the x_1-x_3 plane) but their directions can vary from layer to layer. In practice, the fibers in alternate layers have the same direction and the ply angles of adjacent layers generally differ by 90 deg. The failure mechanisms and failure criteria of layered composites are important factors considered in the design of the composites. When a layered composite is subjected to external loads, failure may be caused by delamination of the interface or development of cracks in the layers. Cracks originating from the interface of two dissimilar layers can propagate either along the interface or transverse into a layer. A crack that has developed in a layer (i.e., an intralayer crack) is shown in Fig. 1. The stresses at the tip of the crack are singular and proportional to r^{-k}, where k is the order of singularity and r is the distance from the crack tip. The three principle methods used in stress singularity analysis of composites are (i) separation of variables method, (ii) complex potential method, and (iii) Mellin integral transform method. The separation of variables method was initially used by Williams (1959) to analyze the interface crack and by Zak and Williams (1963) to study the crack perpendicular to the bimaterial interface. Gdoutos and Theocaris (1975) used the complex potential method to investigate the singular stresses of a wedge in contact with a half-plane while Bogy (1968) used the Mellin integral transform technique to analyze the problem of bonded dissimilar elastic quarter planes.

The initial studies of crack problems in composites were limited to elastically isotropic materials. The singularity at the tip of a crack along the bonded interface between two isotropic materials was investigated by Williams (1959), Erdogan (1965), and Hein and Erdogan (1971). The stress singularity at the tip of a crack perpendicular to the interface was studied by Zak and Williams (1963) and Cook and Erdogan (1972). The stress singularity at the tip of a crack terminating at a slipping interface was studied by Gharpuray et al. (1991) while Wijeyewickrema et al. (1995) considered the singular stress field of a crack terminating at a frictional interface. The composite wedge problem is related to the crack-tip singularity problem, since the tip of an interface crack can be modeled as a bimaterial wedge where each material domain occupies a wedge angle of π. The singularity at the vertex of a wedge pressing on a half-plane with different interfacial conditions was studied by Gdoutos and Theocaris (1975). The singularity analysis of stress at the vertex of a composite wedge was studied by Dempsey and Sinclair (1979) by using the separation of variables method.

The stress singularity problem of anisotropic materials was first investigated by Bogy (1972) and Kuo and Bogy (1974) who studied the singularity of a symmetrically orthotropic elastic wedge by using the complex representation scheme of Green and Zerna (1954). Eshelby et al. (1953) proposed a solution scheme to solve problems in anisotropic elasticity by using displacements as the basic unknowns, where the displacements and stresses are independent of one Cartesian coordinate. This scheme was more general than the previous scheme and was further developed by Stroh (1958, 1962). Another scheme which is equivalent to the Stroh formalism is the solution scheme of Leknitskii (1963). The Stroh formalism has been used in the study of many wedge problems in anisotropic elasticity. Line force loadings on anisotropic half-spaces and wedges have been considered by Barnett and Lothe (1975), while anisotropic elastic wedges at critical wedge angles have been studied by Hwu and Ting (1990). The order of singularity at the tip of a crack normal to the interface of two anisotropic layers was studied by Ting and Hoang (1984) and at the tip of a crack

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Sept. 25, 1997; final revision, Mar. 2, 1998. Associate Technical Editor: W. J. Drugan.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

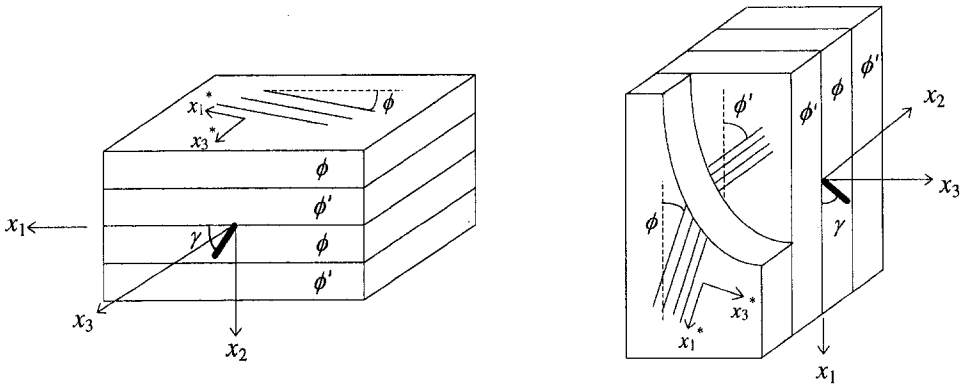


Fig. 1 An angle-ply layered composite with an intralayer crack

along the interface of two different anisotropic materials was studied by Ting (1986). Singularities at the tip of an inclined crack terminating at the interface of two anisotropic materials have been investigated by Lin and Sung (1997) and detailed results were given for the in-plane case where the materials are both orthotropic. The monograph by Ting (1996) contains the Stroh formalism in detail and also gives more examples of wedge problems in anisotropic elasticity.

In this study, the order of singularity at the tip of an inclined crack terminating at the interface of two different anisotropic layers is investigated. Hence, it is sufficient to consider two wedges created by a crack, pressing on a half-plane (see Fig. 2). The crack makes an angle γ with the x_1-axis. The interfaces considered in the study are both a fully bonded interface and a frictional interface. The frictional interface is governed by Coulomb's friction law. The direction of frictional stresses depends on the relative slip direction between the two wedges and the half-plane. Following the work of Wijeyewickrema et al. (1995), two configurations are considered; Case I: the two wedges slip in opposite directions (see Fig. 3), and Case II: the two wedges slip in the same direction (see Fig. 4).

In this paper, the Stroh formalism (Stroh, 1958) is adopted and the necessary expressions for displacements and stresses are given in Section 2. Following the work of Ting and Chou (1981), the stresses are expressed as $\sigma_{ij} = r^{-k}F_{ij}(\theta)$ to study the power singularities. The characteristic equation to determine the singularity at the tip of the crack terminating at a fully bonded interface is obtained in Section 3, and at a frictional interface is obtained in Section 4. The characteristic equations which consist of transcendental functions are obtained by equating the determinant of an 18×18 coefficient matrix to zero. A numerical method, i.e., the secant method, is used to solve

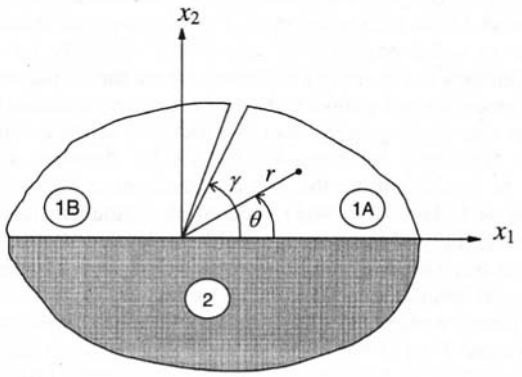


Fig. 2 Crack terminating at the interface of two anisotropic layers

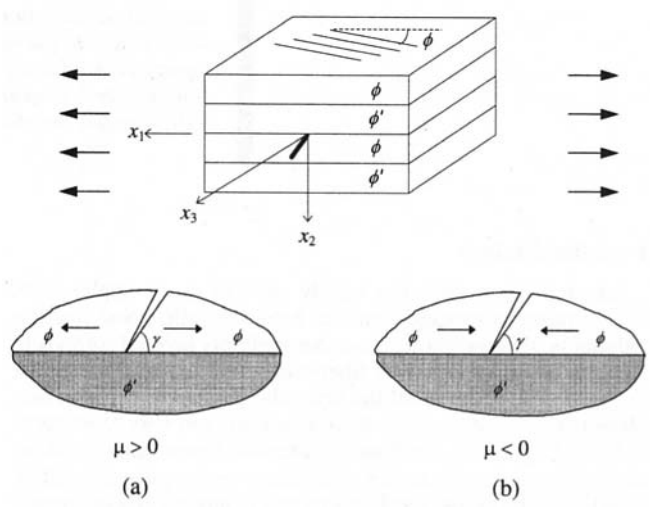


Fig. 3 Case I: the two wedges slip in opposite directions (*a*) away from each other, (*b*) toward each other

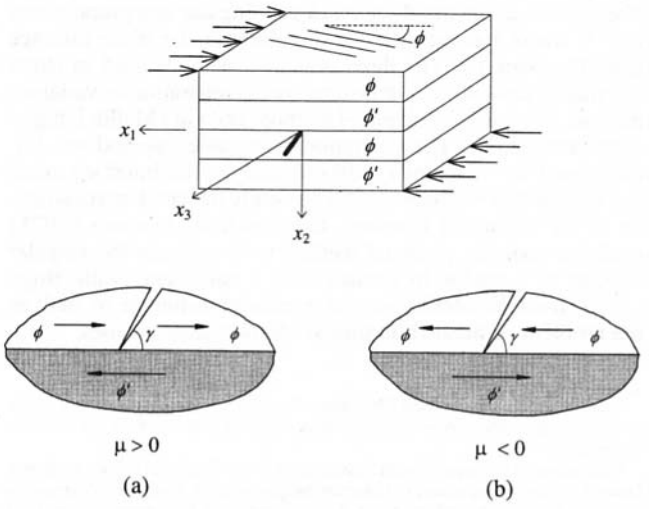


Fig. 4 Case II: the two wedges slip in the same direction (*a*) to the right relative to the bottom half-plane, (*b*) to the left relative to the bottom half-plane

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

the characteristic equation and numerical results for the singularity of a high modulus graphite/epoxy layered composite are illustrated in Section 5. The numerical results of the singularities obtained in this study agree with those obtained by Ting and Hoang (1984) when the crack is normal to a fully bonded interface.

2 Basic Equations

For two-dimensional problems in anisotropic elasticity where all physical quantities depend only on the x_1 and x_2 coordinates of a rectangular Cartesian coordinate system, the displacement field can be written as

$$u_i(x_1, x_2) = a_i f(x_1 + px_2), \tag{1}$$

where $f(z)$ is an analytic function of the complex variable $z = x_1 + px_2$, (Stroh, 1958). The constant p is determined from the sextic equation

$$| c_{i1k1} + p(c_{i1k2} + c_{i2k1}) + p^2 c_{i2k2}| = 0, \tag{2a}$$

and the constant vector a_i is determined from

$$[c_{i1k1} + p(c_{i1k2} + c_{i2k1}) + p^2 c_{i2k2}]a_k = 0, \tag{2b}$$

where repeated Latin indices imply summation. Eshelby et al. (1953) have shown that the roots of Eq. (2a) cannot be real. Hence, when all six roots are assumed distinct, the three different pairs of complex conjugates are denoted by p_α and $\bar{p}_\alpha$, ($\alpha = 1, 2, 3$), where the imaginary part of p_α is taken to be positive and an overbar represents the complex conjugate. The value of a_i corresponding to p_α and $\bar{p}_\alpha$ are denoted by $a_{i\alpha}$ and $\bar{a}_{i\alpha}$, respectively. A general expression for the displacement can then be written as

$$u_i(x_1, x_2) = \sum_{\alpha=1}^{3} [a_{i\alpha} f_\alpha(x_1 + p_\alpha x_2) + \bar{a}_{i\alpha} g_\alpha(x_1 + \bar{p}_\alpha x_2)]. \tag{3}$$

The stress fields are related to the displacement fields in Eq. (3) by

$$\sigma_{ij}(x_1, x_2) = c_{ijkl}\epsilon_{kl} = c_{ijkl}u_{k,l} = \sum_{\alpha=1}^{3} [\tau_{ij\alpha} f'_\alpha(x_1 + p_\alpha x_2) + \bar{\tau}_{ij\alpha} g'_\alpha(x_1 + \bar{p}_\alpha x_2)], \tag{4}$$

where

$$\tau_{ij\alpha} = (c_{ijk1} + p_\alpha c_{ijk2})a_{k\alpha}. \tag{5}$$

In the study of singular stress fields, where the focus is on power singularities, it is convenient to express the stresses as $\sigma_{ij} = r^{-k}\bar{F}_{ij}(\theta)$, where k is the order of the singularity. To achieve this, following the work of Ting and Chou (1981), the unknown functions f_α and g_α in Eqs. (3) and (4) are chosen as

$$\left.\begin{aligned} f_\alpha(x_1 + p_\alpha x_2) &= (x_1 + p_\alpha x_2)^{1-k} M_\alpha/(1-k), \\ g_\alpha(x_1 + \bar{p}_\alpha x_2) &= (x_1 + \bar{p}_\alpha x_2)^{1-k} N_\alpha/(1-k), \end{aligned}\right\} \tag{6}$$

where M_α and N_α are arbitrary constants. Substituting Eq. (6) into Eqs. (3) and (4) yields

$$u_i(r, \theta) = r^{1-k} \sum_{\alpha=1}^{3} [a_{i\alpha} M_\alpha \xi_\alpha^{1-k}(\theta) + \bar{a}_{i\alpha} N_\alpha \bar{\xi}_\alpha^{1-k}(\theta)]/(1-k), \tag{7}$$

$$\sigma_{ij}(r, \theta) = r^{-k} \sum_{\alpha=1}^{3} [\tau_{ij\alpha} M_\alpha \xi_\alpha^{-k}(\theta) + \bar{\tau}_{ij\alpha} N_\alpha \bar{\xi}_\alpha^{-k}(\theta)], \tag{8}$$

where

$$\xi_\alpha(\theta) = \cos\theta + p_\alpha \sin\theta. \tag{9}$$

By using Eqs. (2b), (5), (8), and (9), the traction on a radial plane with a unit outward normal vector $n_j \equiv (-\sin\theta, \cos\theta, 0)$ can be obtained as

$$t_i(r, \theta) = \sigma_{ij} n_j = r^{-k} \sum_{\alpha=1}^{3} [b_{i\alpha} M_\alpha \xi_\alpha^{1-k}(\theta) + \bar{b}_{i\alpha} N_\alpha \bar{\xi}_\alpha^{1-k}(\theta)], \tag{10}$$

where

$$b_{i\alpha} = -\frac{1}{p_\alpha}(c_{i1k1} + p_\alpha c_{i1k2})a_{k\alpha} = (c_{i2k1} + p_\alpha c_{i2k2})a_{k\alpha}. \tag{11}$$

For convenience in the formulation, making use of Eqs. (5) and (11), Eq. (8) is rewritten as

$$\sigma_{i1}(r, \theta) = -r^{-k} \sum_{\alpha=1}^{3} [p_\alpha b_{i\alpha} M_\alpha \xi_\alpha^{-k}(\theta) + \bar{p}_\alpha \bar{b}_{i\alpha} N_\alpha \bar{\xi}_\alpha^{-k}(\theta)], \tag{12a}$$

$$\sigma_{i2}(r, \theta) = r^{-k} \sum_{\alpha=1}^{3} [b_{i\alpha} M_\alpha \xi_\alpha^{-k}(\theta) + \bar{b}_{i\alpha} N_\alpha \bar{\xi}_\alpha^{-k}(\theta)]. \tag{12b}$$

3 Singularity at the Tip of a Crack Terminating at a Fully Bonded Interface

The crack located in the upper half-plane makes an angle γ with the x_1-axis and splits the upper half-plane into two planar wedges as shown in Fig. 2. The domain under consideration is given by $\gamma - 2\pi \le \theta \le \gamma$. All quantities associated with the upper and the lower half-planes are denoted by superscripts (1) and (2). The superscript (1A) refers to the quantities corresponding to the wedge on the right and the superscript (1B) refers to the quantities corresponding to the wedge on the left. The characteristic equation which yields the crack-tip singularity can be derived as follows:

The traction-free boundaries at the crack faces are

$$t_i = 0 \quad \text{at} \quad \theta = \gamma \quad \text{and} \quad \theta = \gamma - 2\pi. \tag{13}$$

From Eq. (9),

$$\xi_\alpha(\gamma) = \chi_\alpha = \cos\gamma + p_\alpha^{(1)} \sin\gamma. \tag{14}$$

Using Eqs. (10) and (14) in Eq. (13) yields

$$\sum_{\alpha=1}^{3} b_{i\alpha}^{(1)} \chi_\alpha^{1-k} M_\alpha^{(1A)} + \bar{b}_{i\alpha}^{(1)} \bar{\chi}_\alpha^{1-k} N_\alpha^{(1A)} = 0, \tag{15a}$$

$$\sum_{\alpha=1}^{3} e^{2\pi k i} b_{i\alpha}^{(1)} \chi_\alpha^{1-k} M_\alpha^{(1B)} + e^{-2\pi k i} \bar{b}_{i\alpha}^{(1)} \bar{\chi}_\alpha^{1-k} N_\alpha^{(1B)} = 0. \tag{15b}$$

The continuity of the displacements and the stresses at the interface are

$$u_i^{(1A)} = u_i^{(2)} \quad \text{and} \quad \sigma_{2i}^{(1A)} = \sigma_{2i}^{(2)} \quad \text{at } \theta = 0, \tag{16a, b}$$

$$u_i^{(1B)} = u_i^{(2)} \quad \text{and} \quad \sigma_{2i}^{(1B)} = \sigma_{2i}^{(2)} \quad \text{at } \theta = -\pi. \tag{16c, d}$$

By using Eqs. (7) and (12b) in Eq. (16),

$$\sum_{\alpha=1}^{3} a_{i\alpha}^{(1)} M_\alpha^{(1A)} + \bar{a}_{i\alpha}^{(1)} N_\alpha^{(1A)} = \sum_{\alpha=1}^{3} a_{i\alpha}^{(2)} M_\alpha^{(2)} + \bar{a}_{i\alpha}^{(2)} N_\alpha^{(2)}, \tag{17a}$$

$$\sum_{\alpha=1}^{3} b_{i\alpha}^{(1)} M_\alpha^{(1A)} + \bar{b}_{i\alpha}^{(1)} N_\alpha^{(1A)} = \sum_{\alpha=1}^{3} b_{i\alpha}^{(2)} M_\alpha^{(2)} + \bar{b}_{i\alpha}^{(2)} N_\alpha^{(2)}, \tag{17b}$$

$$\sum_{\alpha=1}^{3} e^{\pi k i} a_{i\alpha}^{(1)} M_\alpha^{(1B)} + e^{-\pi k i} \bar{a}_{i\alpha}^{(1)} N_\alpha^{(1B)} = \sum_{\alpha=1}^{3} e^{\pi k i} a_{i\alpha}^{(2)} M_\alpha^{(2)} + e^{-\pi k i} \bar{a}_{i\alpha}^{(2)} N_\alpha^{(2)}, \tag{17c}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\sum_{\alpha=1}^{3} e^{\pi k i} b_{i\alpha}^{(1)} M_\alpha^{(1B)} + e^{-\pi k i} \bar{b}_{i\alpha}^{(1)} N_\alpha^{(1B)} = \sum_{\alpha=1}^{3} e^{\pi k i} b_{i\alpha}^{(2)} M_\alpha^{(2)} + e^{-\pi k i} \bar{b}_{i\alpha}^{(2)} N_\alpha^{(2)}. \quad (17d)$$

Equations (15) and (17) give a system of 18 linear equations which can be written in the matrix form

$$\mathbf{K}_1(k)\mathbf{q} = \mathbf{0}, \quad (18)$$

where $\mathbf{K}_1$ is an 18 × 18 square matrix (see Appendix A) and $\mathbf{q}$ is given by

$$\mathbf{q} = \{\mathbf{M}^{(1A)}, \mathbf{N}^{(1A)}, \mathbf{M}^{(2)}, \mathbf{N}^{(2)}, \mathbf{M}^{(1B)}, \mathbf{N}^{(1B)}\}^T, \quad (19)$$

where $\mathbf{M}^{(n)} = \{M_1^{(n)}, M_2^{(n)}, M_3^{(n)}\}^T$ and $\mathbf{N}^{(n)} = \{N_1^{(n)}, N_2^{(n)}, N_3^{(n)}\}^T$. For nontrivial solution of $\mathbf{q}$,

$$|\mathbf{K}_1(k)| = 0. \quad (20)$$

The above equation is the characteristic equation to determine the order of singularity k. The root k can be either real or complex depending on the inclined angle and the material properties of each half-plane. Since the strain energy must be bounded in a finite region where r is close to zero, the range of interest is $0 < \text{Re}\ (k) < 1$. If the root is complex, an oscillatory singular stress field will result. For the special case when both domains consist of the same material, there is no interface and the singularities are 0.5 of multiplicity 3 (Ting, 1986). When the two half-planes consist of different materials and the crack is along the interface, i.e., an interface crack, the singularities are 0.5 and $0.5 \pm \epsilon i$, where ϵ is a positive real constant which depends on the material constants of the two half-planes (Ting, 1986; Suo, 1990). For arbitrary γ the singularities are determined by solving Eq. (20) numerically.

4 Singularity at the Tip of a Crack Terminating at a Frictional Interface

It is assumed that the normal stress along the frictional interface is always compressive, i.e., both domains adjacent to the interface are in contact and press against each other at the interface which is governed by Coulomb's law of friction. Following the work of Wijeyewickrema et al. (1995), it is convenient to consider two configurations classified as, Case I: the two wedges slip in opposite directions and Case II: the two wedges slip in the same direction (see Figs. 3 and 4).

4.1 Case I: The Two Wedges Slip in Opposite Directions. This mode of slip might occur when the composite is subjected to tension or compression. Both the cases of the two wedges slipping away from each other and slipping toward each other can be treated at the same time by allowing the coefficient of friction μ to be either positive or negative. Positive values of μ are taken to correspond to the two wedges slipping away from each other while negative values of μ correspond to the two wedges slipping toward each other, as shown in Fig. 3.

The traction-free boundaries along the crack faces and the continuity of stresses along the interface are given by Eqs. (13), (16*b*), and (16*d*), i.e., Eqs. (15), (17*b*), and (17*d*) derived in the previous section are still valid here. At the interface the stresses are related by

$$\sigma_{21}^{(2)} = -\mu\sigma_{22}^{(2)}, \quad (\sigma_{22}^{(2)} < 0) \quad \text{at} \quad \theta = 0, \quad (21a)$$

$$\sigma_{21}^{(2)} = +\mu\sigma_{22}^{(2)}, \quad (\sigma_{22}^{(2)} < 0) \quad \text{at} \quad \theta = -\pi. \quad (21b)$$

The continuity of the displacements in x_2 and x_3-directions along the interface are

$$u_2^{(1A)} = u_2^{(2)} \quad \text{and} \quad u_3^{(1A)} = u_3^{(2)} \quad \text{at} \quad \theta = 0, \quad (21c, d)$$

$$u_2^{(1B)} = u_2^{(2)} \quad \text{and} \quad u_3^{(1B)} = u_3^{(2)} \quad \text{at} \quad \theta = -\pi. \quad (21e, f)$$

By using Eqs. (7) and (12*b*), Eqs. (21*a–f*) yield

$$\sum_{\alpha=1}^{3} (b_{1\alpha}^{(2)} + \mu b_{2\alpha}^{(2)}) M_\alpha^{(2)} + (\bar{b}_{1\alpha}^{(2)} + \mu \bar{b}_{2\alpha}^{(2)}) N_\alpha^{(2)} = 0, \quad (22a)$$

$$\sum_{\alpha=1}^{3} e^{\pi k i} (b_{1\alpha}^{(2)} - \mu b_{2\alpha}^{(2)}) M_\alpha^{(2)} + e^{-\pi k i} (\bar{b}_{1\alpha}^{(2)} - \mu \bar{b}_{2\alpha}^{(2)}) N_\alpha^{(2)} = 0, \quad (22b)$$

$$\sum_{\alpha=1}^{3} a_{2\alpha}^{(1)} M_\alpha^{(1A)} + \bar{a}_{2\alpha}^{(1)} N_\alpha^{(1A)} = \sum_{\alpha=1}^{3} a_{2\alpha}^{(2)} M_\alpha^{(2)} + \bar{a}_{2\alpha}^{(2)} N_\alpha^{(2)}, \quad (22c)$$

$$\sum_{\alpha=1}^{3} a_{3\alpha}^{(1)} M_\alpha^{(1A)} + \bar{a}_{3\alpha}^{(1)} N_\alpha^{(1A)} = \sum_{\alpha=1}^{3} a_{3\alpha}^{(1)} M_\alpha^{(2)} + \bar{a}_{3\alpha}^{(1)} N_\alpha^{(2)}, \quad (22d)$$

$$\sum_{\alpha=1}^{3} e^{\pi k i} a_{2\alpha}^{(1)} M_\alpha^{(1B)} + e^{-\pi k i} \bar{a}_{2\alpha}^{(1)} N_\alpha^{(1B)} = \sum_{\alpha=1}^{3} e^{\pi k i} a_{2\alpha}^{(2)} M_\alpha^{(2)} + e^{-\pi k i} \bar{a}_{2\alpha}^{(2)} N_\alpha^{(2)}, \quad (22e)$$

$$\sum_{\alpha=1}^{3} e^{\pi k i} a_{3\alpha}^{(1)} M_\alpha^{(1B)} + e^{-\pi k i} \bar{a}_{3\alpha}^{(1)} N_\alpha^{(1B)} = \sum_{\alpha=1}^{3} e^{\pi k i} a_{3\alpha}^{(2)} M_\alpha^{(2)} + e^{-\pi k i} \bar{a}_{3\alpha}^{(2)} N_\alpha^{(2)}. \quad (22f)$$

Equations (15), (17*b*), (17*d*), and (22) which form a system of 18 linear equations can be written in the matrix form

$$\mathbf{K}_2(k)\mathbf{q} = \mathbf{0}, \quad (23)$$

where $\mathbf{q}$ is given by Eq. (19). Consequently, the order of singularity is determined numerically from the condition that the determinant of the matrix $\mathbf{K}_2$ is identically zero.

4.2 Case II: The Two Wedges Slip in the Same Direction. When the composite is subjected to shear loading the two wedges may slip in the same direction. Here, positive values of μ correspond to the two wedges slipping to the right relative to the lower half-plane, and negative values of μ correspond to the two wedges slipping to the left relative to lower half-plane as shown in Fig. 4. All boundary and interfacial conditions in Case I are still valid except Eq. (21*b*), i.e., the relation between the shear and normal stresses at $\theta = -\pi$, which is replaced by

$$\sigma_{21}^{(2)} = -\mu\sigma_{22}^{(2)}, \quad (\sigma_{22}^{(2)} < 0) \quad \text{at} \quad \theta = -\pi. \quad (24)$$

Equations (12*b*) and (24) yield

$$\sum_{\alpha=1}^{3} e^{\pi k i} (b_{1\alpha}^{(2)} + \mu b_{2\alpha}^{(2)}) M_\alpha^{(2)} + e^{-\pi k i} (\bar{b}_{1\alpha}^{(2)} + \mu \bar{b}_{2\alpha}^{(2)}) N_\alpha^{(2)} = 0. \quad (25)$$

The order of singularity of this case can be obtained by writing Eqs. (15), (17*b, d*), (22*a*), (22*c–f*) and (25) in the matrix form

$$\mathbf{K}_3(k)\mathbf{q} = \mathbf{0}, \quad (26)$$

where $\mathbf{q}$ is given by Eq. (19). Once again the singularity k is determined by equating the determinant $|\mathbf{K}_3|$ to zero.

5 Numerical Examples and Discussion

The stress-strain relationship can be written in the abbreviated form

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

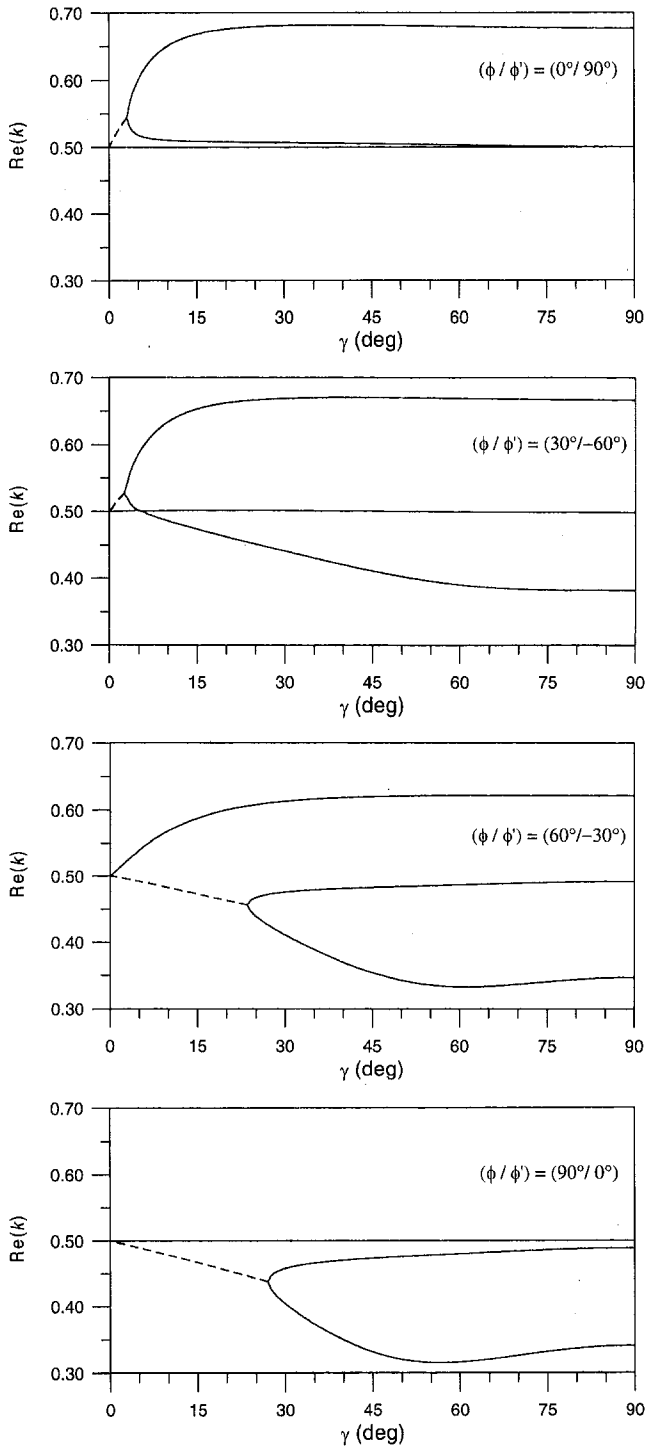


Fig. 5 The order of singularity of a crack terminating at a fully bonded interface. —— k is real; ------ k is complex.

$$\sigma_i = c_{ij}\epsilon_j \quad \text{or} \quad \epsilon_i = s_{ij}\sigma_j \tag{27}$$

where

$$\sigma_1 = \sigma_{11}, \quad \sigma_2 = \sigma_{22}, \quad \sigma_3 = \sigma_{33},$$
$$\sigma_4 = \sigma_{23}, \quad \sigma_5 = \sigma_{31}, \quad \sigma_6 = \sigma_{12},$$
$$\epsilon_1 = \epsilon_{11}, \quad \epsilon_2 = \epsilon_{22}, \quad \epsilon_3 = \epsilon_{33},$$
$$\epsilon_4 = 2\epsilon_{23}, \quad \epsilon_5 = 2\epsilon_{31}, \quad \epsilon_6 = 2\epsilon_{12}, \tag{28}$$

and c_{ij}, s_{ij} are the stiffnesses and the compliances, respectively. A layered orthotropic composite of high modulus graphite/epoxy with the elastic constants

$$E_1 = 20 \times 10^6 \text{ psi}, E_2 = E_3 = 2.1 \times 10^6 \text{ psi},$$
$$G_{12} = G_{23} = G_{13} = 0.85 \times 10^6 \text{ psi},$$
$$\nu_{12} = \nu_{23} = \nu_{13} = 0.21, \tag{29}$$

taken from Ting and Hoang (1984), is used in the numerical calculation. Here E_i, G_{ij}, and ν_{ij} are the Young modulus, the shear modulus, and the Poisson's ratio, respectively. Each layer of the composite has a local coordinate system (x_1^*, x_2^*, x_3^*) which is selected in such a way that the x_1^*-axis is in the longitudinal direction of the fiber and the x_2^*-axis is normal to the layer and coincides with the x_2-axis of the global coordinate system (x_1, x_2, x_3), as shown in Fig. 1. Alternate layers of the composite have the same fiber orientation. Since each layer is considered as an orthotropic material, the elastic compliance matrix s_{ij}^* referred to the local coordinate system is obtained from the given set of engineering constants (see Christensen, 1979, p. 154). The stiffness matrix c_{ij}^* which is obtained by

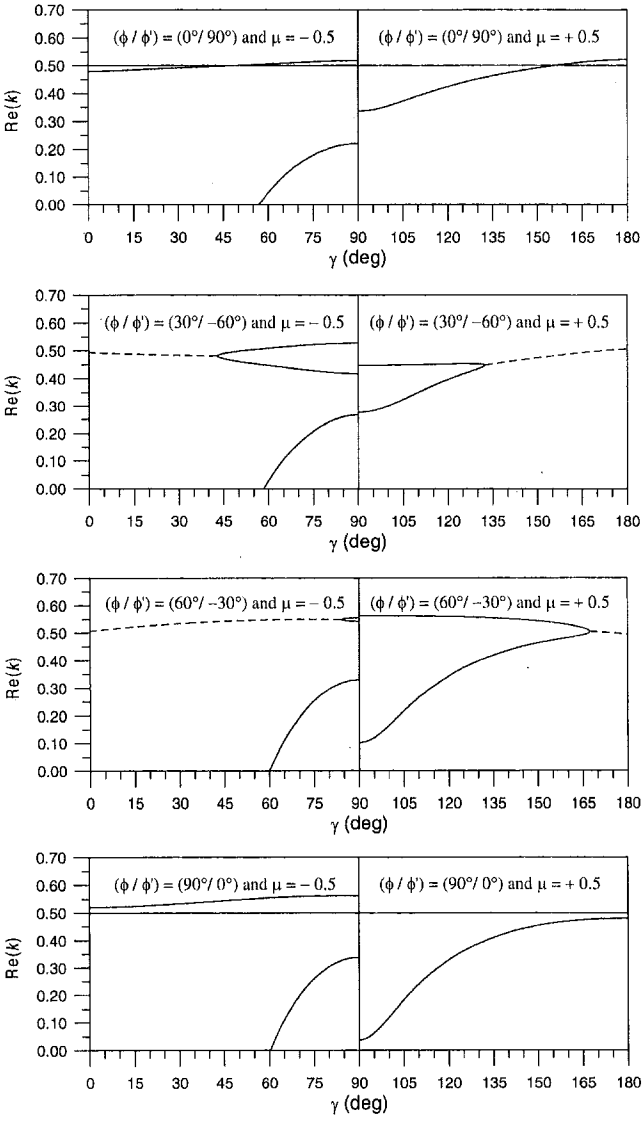


Fig. 6 The order of singularity of a crack terminating at a frictional interface when the two wedges slip in opposite directions (Case I). —— k is real; ------ k is complex.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

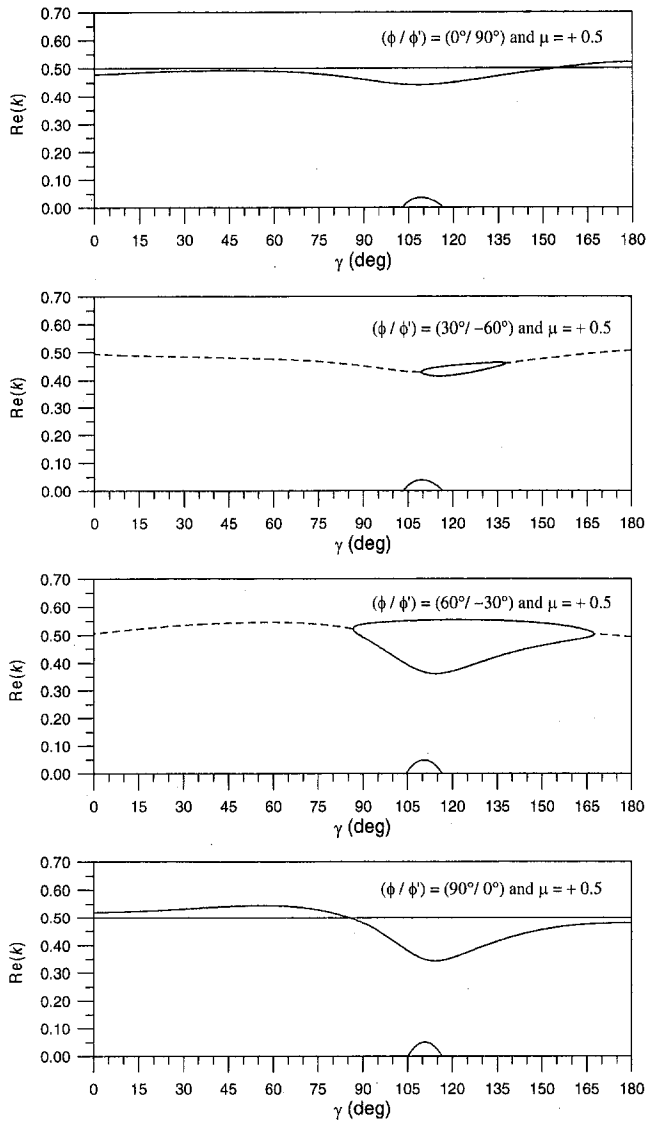


Fig. 7 The order of singularity of a crack terminating at a frictional interface when the two wedges slip in the same direction (Case II). —— k is real; ------ k is complex.

inverting the s_{ij}^* matrix is the same for all layers since it is referred to the local coordinate system. The stiffness matrix c_{ij} referred to the global coordinate system is obtained by transforming the c_{ij}^* matrix according to the ply-angle. The ply-angles of adjacent layers are ϕ and ϕ' as shown in Fig. 1. For the ϕ-layer, the global coordinates are related to the local coordinate system by

$$\begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \begin{Bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{Bmatrix}. \quad (30)$$

Following the work of Auld (1973), the stiffness matrices c_{ij} and c'_{ij} are related by

$$[c] = [m][c^*][m]^T, \quad (31)$$

where the superscript T denotes the transpose of a matrix and

$$[m] = \begin{bmatrix} \cos^2\phi & 0 & \sin^2\phi & 0 & \sin 2\phi & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \sin^2\phi & 0 & \cos^2\phi & 0 & -\sin 2\phi & 0 \\ 0 & 0 & 0 & \cos\phi & 0 & -\sin\phi \\ -(\sin 2\phi)/2 & 0 & (\sin 2\phi)/2 & 0 & \cos 2\phi & 0 \\ 0 & 0 & 0 & \sin\phi & 0 & \cos\phi \end{bmatrix}. \quad (32)$$

Similarly, the c_{ij}^* matrix of the ϕ'-layer is transformed by the angle ϕ' and matrix c'_{ij} is obtained. In many applications, ϕ

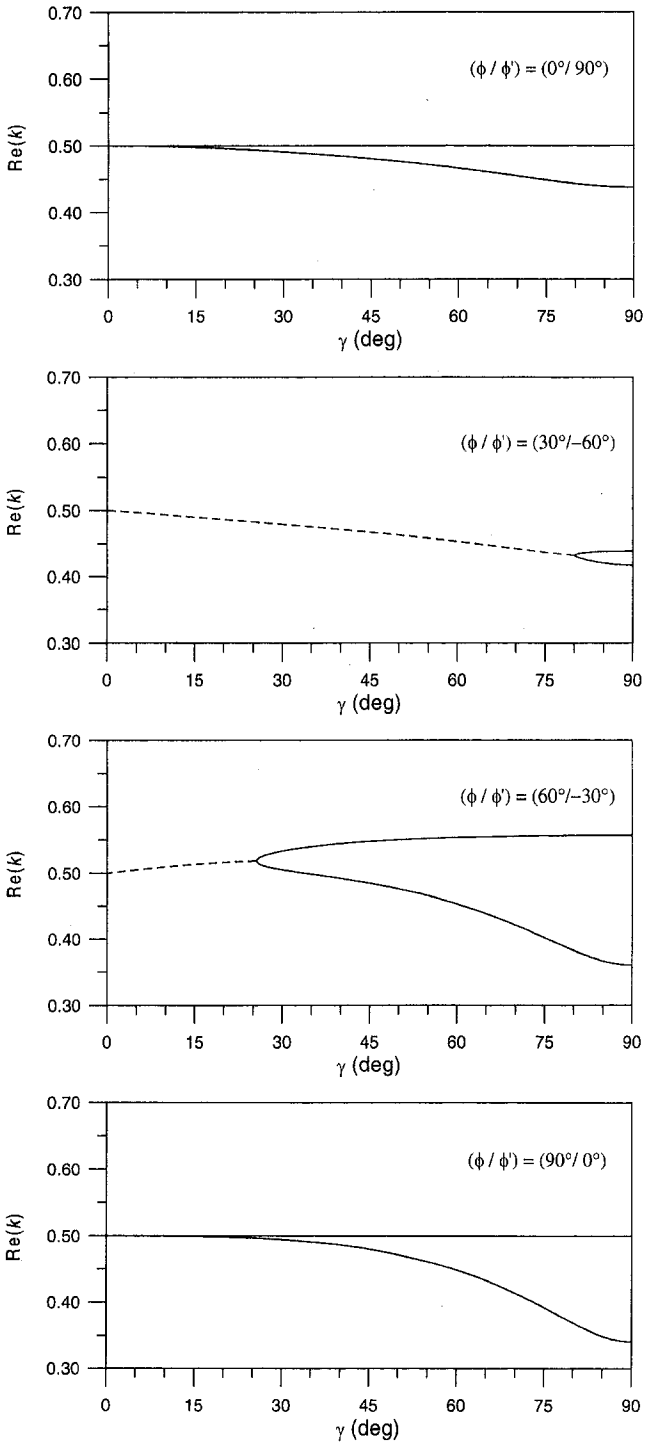


Fig. 8 The order of singularity of a crack terminating at a slipping interface. —— k is real; ------ k is complex.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

and ϕ' differ by $\pi/2$ and hence in the present study the (ϕ/ϕ') combinations considered are (0 deg/90 deg), (30 deg/−60 deg), (60 deg/−30 deg) and (90 deg/0 deg). For a crack which is normal to the interface, i.e., $\gamma = \pi/2$, due to geometric considerations the singularity k is identical for the combinations ($\phi/-\phi'$) and ($-\phi/\phi'$). In the present analysis it was found that even when the crack is inclined to the interface, i.e., $\gamma \neq \pi/2$, the singularity is identical for the combinations ($\phi/-\phi'$) and ($-\phi/\phi'$). Numerical results are shown in Figs. 5–8 where the real part of complex conjugate roots are shown by dashed lines. In the numerical calculation of the singularity, the complex arithmetic secant method with deflation is used. For a specified crack angle γ, once one root of the characteristic equation is found, this root is factored out of the equation by using synthetic division. The deflated equation is then solved for the other roots. The procedure is repeated until all roots in the range $0 < \text{Re}\ (k) < 1$ are obtained.

For the fully bonded interface, the singularities are identical when $\gamma = \gamma^0$ and $\pi - \gamma^0$ due to geometric considerations, hence, the singularities are shown in the range $0 \leq \gamma \leq \pi/2$, in Fig. 5. The numerical results reveal that there are three roots in the range $0 < \text{Re}\ (k) < 1$. These roots can be either, one conjugate pair of complex roots and one real root; or three real roots. Three real roots are always obtained when the crack inclination $\gamma > \pi/6$. For a crack normal to the interface, i.e., $\gamma = \pi/2$, the three real roots are the same as those obtained by Ting and Hoang (1984). In addition, the results show that for (0 deg/90 deg) and (90 deg/0 deg) combinations, one of the roots is 0.5 and does not depend on the crack angle. This is because for these combinations, the x_1-x_2 plane coincides with the plane of material symmetry, and consequently the problems of in-plane deformation and antiplane deformation with respect to this plane are uncoupled. It is shown in Appendix B that $k = 0.5$ is the root that corresponds to the problem of antiplane deformation.

For the frictional interface, the coefficient of friction μ is chosen as 0.5. When the two wedges slip in opposite directions (Case I), the singularities are identical for $\gamma = \gamma^0$ and $\pi - \gamma^0$. Hence, it is sufficient to plot the singularity for the two wedges slipping toward each other ($\mu < 0$) in the range $0 \leq \gamma \leq \pi/2$ and for the two wedges slipping away from each other ($\mu > 0$) in the range $\pi/2 \leq \gamma \leq \pi$ as in Fig. 6. Only two roots of k are found for wedges slipping away from each other, but there exist either two roots or three roots of k for wedges slipping toward each other. For an interface crack, the roots of k are two real roots or two complex conjugate roots. For a crack normal to the interface, three real roots of k exist when the wedges slip toward each other while there are only two real roots when the wedges slip away from each other.

For Case II, the crack-tip singularity when the wedges slip to the right for $\mu = \mu^0$ and $\gamma = \gamma^0$ is identical to the crack tip singularity when the wedges slip to the left for $\mu = -\mu^0$ and $\gamma = \pi - \gamma^0$. Hence, results are shown in Fig. 7 only when the two wedges slip to the right ($\mu > 0$) for $0 \leq \gamma \leq \pi$. At any crack angles γ either two or three roots of k exist. Two roots exist for a wide range of γ while three roots exist in a narrow range 100 deg $< \gamma <$ 120 deg. For an interface crack or a crack normal to the interface, two roots of k exist.

Finally, the numerical results for the special case of the slipping interface ($\mu = 0$) are given in Fig. 8. The singularities are identical for $\gamma = \gamma^0$ and $\pi - \gamma^0$ hence the singularities are shown in the range $0 \leq \gamma \leq \pi/2$. There exist two roots of k for any crack angle γ.

The singularities obtained here are important in the determination of stress distributions in critical regions of composites. In addition, the singularities play an important role as exact solutions to check the numerical solutions at a crack tip terminating at an interface obtained from other approximate methods, such as boundary element methods and finite element methods.

References

Auld, B. A., 1973, *Acoustic Fields and Waves in Solids,* Vol. 1, John Wiley and Sons, New York.

Barnett, D. M., and Lothe, J., 1975, "Line Force Loadings on Anisotropic Half-Spaces and Wedges," *Physica Norvegica,* Vol. 8, pp. 13–22.

Bogy, D. B., 1968, "Edge-Bonded Dissimilar Orthogonal Elastic Wedges Under Normal and Shear Loading," ASME JOURNAL OF APPLIED MECHANICS, Vol. 35, pp. 460–466.

Bogy, D. B., 1972, "The Plane Solution for Anisotropic Elastic Wedges Under Normal and Shear Loading," ASME JOURNAL OF APPLIED MECHANICS, Vol. 39, pp. 1103–1109.

Christensen, R. M., 1979, *Mechanics of Composite Materials,* John Wiley and Sons, New York.

Cook, T. S., and Erdogan, F., 1972, "Stresses in Bonded Materials with a Crack Perpendicular to the Interface," *International Journal of Engineering Science,* Vol. 10, pp. 677–697.

Dempsey, J. P., and Sinclair, G. B., 1979, "On the Stress Singularities in the Plane Elasticity of the Composite Wedge," *Journal of Elasticity,* Vol. 9, pp. 373–391.

Erdogan, F., 1965, "Stress Distribution in Bonded Dissimilar Materials With Cracks," ASME JOURNAL OF APPLIED MECHANICS, Vol. 32, pp. 403–410.

Eshelby, J. D., Read, W. T., and Shockley, W., 1953, "Anisotropic Elasticity with Applications to Dislocation Theory," *Acta Metallurgica,* Vol. 1, pp. 251–259.

Gdoutos, E. E., and Theocaris, P. S., 1975, "Stress Concentrations at the Apex of a Plane Indenter Acting on an Elastic Half-plane," ASME JOURNAL OF APPLIED MECHANICS, Vol. 42, pp. 688–692.

Gharpuray, V. M., Dundurs, J., and Keer, L. M., 1991, "A Crack Terminating at a Slipping Interface Between Two Materials," ASME JOURNAL OF APPLIED MECHANICS, Vol. 58, pp. 960–963.

Green, A. E., and Zerna, W., 1954, *Theoretical Elasticity,* Clarendon Press, Oxford, U.K.

Hein, V. L., and Erdogan, F., 1971, "Stress Singularities in a Two-Material Wedge," *International Journal of Fracture Mechanics,* Vol. 7, No. 3, pp. 317–330.

Hwu, C., and Ting, T. C. T., 1990, "Solutions for the Anisotropic Elastic Wedge at Critical Wedge Angles," *Journal of Elasticity,* Vol. 24, pp. 1–20.

Kuo, M. C., and Bogy, D. B., 1974, "Plane Solutions for the Displacement and Traction-Displacement Problems for Anisotropic Wedges," ASME JOURNAL OF APPLIED MECHANICS, Vol. 41, pp. 197–203.

Lekhnitskii, S. G., 1963, *Theory of Elasticity of an Anisotropic Elastic Body,* Holden-Day, San Francisco, CA.

Lin, Y. Y., and Sung, J. C., 1997, "Singularities of an Inclined Crack Terminating at an Anisotropic Bimaterial Interface," *International Journal of Solids and Structures,* Vol. 34, pp. 3727–3754.

Stroh, A. N., 1958, "Dislocations and Cracks in Anisotropic Elasticity," *Philosophical Magazine,* Vol. 7, pp. 625–646.

Stroh, A. N., 1962, "Steady State Problems in Anisotropic Elasticity," *Journal of Mathematics and Physics,* Vol. 44, pp. 77–103.

Suo, Z., 1990, "Singularities, Interfaces and Cracks in Dissimilar Anisotropic Media," *Proceedings of the Royal Society of London,* Vol. A427, pp. 331–358.

Ting, T. C. T., 1986, "Explicit Solution and Invariance of the Singularities at an Interface Crack in Anisotropic Composites," *International Journal of Solids and Structures,* Vol. 22, pp. 965–983.

Ting, T. C. T., 1996, *Anisotropic Elasticity,* Oxford University Press, New York.

Ting, T. C. T., and Chou, S. C., 1981, "Edge Singularities in Anisotropic Composites," *International Journal of Solids and Structures,* Vol. 17, pp. 1057–1068.

Ting, T. C. T., and Hoang, P. H., 1984, "Singularities at the Tip of a Crack Normal to the Interface of an Anisotropic Layered Composite," *International Journal of Solids and Structures,* Vol. 20, pp. 439–454.

Wijeyewickrema, A. C., Dundurs, J., and Keer, L. M., 1995, "The Singular Stress Field of a Crack Terminating at a Frictional Interface Between Two Materials," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 289–293.

Williams, M. L., 1959, "The Stresses Around a Fault or Crack in Dissimilar Media," *Bulletin of the Seismological Society of America,* Vol. 49, pp. 199–204.

Zak, A. R., and Williams, M. L., 1963, "Crack Point Stress Singularities at a Bimaterial Interface," ASME JOURNAL OF APPLIED MECHANICS, Vol. 30, pp. 142–143.

APPENDIX A

Coefficient Matrices $\mathbf{K}_1$, $\mathbf{K}_2$, $\mathbf{K}_3$

For a *fully bonded interface,* the coefficient matrix $\mathbf{K}_1$ in Eq. (18) can be written as

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\mathbf{K}_1 = \begin{bmatrix} \mathbf{B}^{(1)}\mathbf{X} & \bar{\mathbf{B}}^{(1)}\mathbf{Y} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{A}^{(1)} & \bar{\mathbf{A}}^{(1)} & -\mathbf{A}^{(2)} & -\bar{\mathbf{A}}^{(2)} & \mathbf{0} & \mathbf{0} \\ \mathbf{B}^{(1)} & \bar{\mathbf{B}}^{(1)} & -\mathbf{B}^{(2)} & -\bar{\mathbf{B}}^{(2)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -e^{\pi ki}\mathbf{A}^{(2)} & -e^{-\pi ki}\bar{\mathbf{A}}^{(2)} & e^{\pi ki}\mathbf{A}^{(1)} & e^{-\pi ki}\bar{\mathbf{A}}^{(1)} \\ \mathbf{0} & \mathbf{0} & -e^{\pi ki}\mathbf{B}^{(2)} & -e^{-\pi ki}\bar{\mathbf{B}}^{(2)} & e^{\pi ki}\mathbf{B}^{(1)} & e^{-\pi ki}\bar{\mathbf{B}}^{(1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & e^{2\pi ki}\mathbf{B}^{(1)}\mathbf{X} & e^{-2\pi ki}\bar{\mathbf{B}}^{(1)}\mathbf{Y} \end{bmatrix}_{18\times 18}, \quad \text{(A1)}$$

where $\mathbf{A}^{(n)} = [a_{i\alpha}^{(n)}]$, $\mathbf{B}^{(n)} = [b_{i\alpha}^{(n)}]$, $\mathbf{X} = \text{diag}\,[\chi_1^{1-k}, \chi_2^{1-k}, \chi_3^{1-k}]$ and $\mathbf{Y} = \text{diag}\,[\bar{\chi}_1^{1-k}, \bar{\chi}_2^{1-k}, \bar{\chi}_3^{1-k}]$.

For a *frictional interface,* the matrices $\mathbf{K}_2$ in Eq. (23) and $\mathbf{K}_3$ in Eq. (26) are given by

$$\mathbf{K}_{2,3} = \begin{bmatrix} \mathbf{B}^{(1)}\mathbf{X} & \bar{\mathbf{B}}^{(1)}\mathbf{Y} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{DA}^{(1)} & \mathbf{D}\bar{\mathbf{A}}^{(1)} & -(\mathbf{DA}^{(2)} + \mathbf{C}^{+}\mathbf{B}^{(2)}) & -(\mathbf{D}\bar{\mathbf{A}}^{(2)} + \mathbf{C}^{+}\bar{\mathbf{B}}^{(2)}) & \mathbf{0} & \mathbf{0} \\ \mathbf{B}^{(1)} & \bar{\mathbf{B}}^{(1)} & -\mathbf{B}^{(2)} & -\bar{\mathbf{B}}^{(2)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -e^{\pi ki}(\mathbf{DA}^{(2)} + \mathbf{C}^{\pm}\mathbf{B}^{(2)}) & -e^{-\pi ki}(\mathbf{D}\bar{\mathbf{A}}^{(2)} + \mathbf{C}^{\pm}\bar{\mathbf{B}}^{(2)}) & e^{\pi ki}\mathbf{DA}^{(1)} & e^{-\pi ki}\mathbf{D}\bar{\mathbf{A}}^{(1)} \\ \mathbf{0} & \mathbf{0} & -e^{\pi ki}\mathbf{B}^{(2)} & -e^{-\pi ki}\bar{\mathbf{B}}^{(2)} & e^{\pi ki}\mathbf{B}^{(1)} & e^{-\pi ki}\bar{\mathbf{B}}^{(1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & e^{2\pi ki}\mathbf{B}^{(1)}\mathbf{X} & e^{-2\pi ki}\bar{\mathbf{B}}^{(1)}\mathbf{Y} \end{bmatrix}_{18\times 18} \quad \text{(A2)}$$

where

$$\mathbf{C}^{\pm} = \begin{bmatrix} 1 & \pm\mu & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad \text{(A3)}$$

In the fourth row of matrix $\mathbf{K}_{2,3}$, the (+) sign corresponds to the two wedges slipping in the same direction and the (−) sign corresponds to the two wedges slipping in opposite directions.

APPENDIX B

Roots of (0 Deg/90 Deg) and (90 Deg/0 Deg) Combinations

Consider an orthotropic layered composite where the layers have the same elastic properties. For the combinations (0 deg/90 deg) and (90 deg/0 deg) the x_1-x_2 plane is a plane of symmetry, hence the antiplane and in-plane problems with respect to this plane are uncoupled. The matrices $\mathbf{A}^{(n)}$ and $\mathbf{B}^{(n)}$ are given by

$$\mathbf{A}^{(n)} = \begin{bmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & a_{33}^{(n)} \end{bmatrix}, \quad \mathbf{B}^{(n)} = \begin{bmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & b_{33}^{(n)} \end{bmatrix}, \quad \text{(A4)}$$

where * denotes a possible nonzero element.

For a layer with $\phi = 0$ deg,

$$a_{33} = -i(G_{23}G_{13})^{-1/2} \quad \text{and} \quad b_{33} = 1, \quad \text{(A5)}$$

and for a layer with $\phi = 90$ deg,

$$a_{33} = -i(G_{12}G_{13})^{-1/2} \quad \text{and} \quad b_{33} = 1. \quad \text{(A6)}$$

The (0 deg/90 deg) or (90 deg/0 deg) composites given by Eq. (29), where $G_{12} = G_{23}$, then has

$$a_{33}^{(1)} = a_{33}^{(2)} = a_{33} = -i(G_{23}G_{31})^{-1/2}. \quad \text{(A7)}$$

Consequently, from matrices $\mathbf{K}_1$ and $\mathbf{K}_{2,3}$ in Eqs. (A1) and (A2) it is possible to obtain the matrix $\mathbf{K}_a$ corresponding to antiplane deformation as

$$\mathbf{K}_a = \begin{bmatrix} \chi_3^{1-k} & \bar{\chi}_3^{1-k} & 0 & 0 & 0 & 0 \\ a_{33} & -a_{33} & -a_{33} & a_{33} & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & -e^{\pi ki}a_{33} & e^{-\pi ki}a_{33} & e^{\pi ki}a_{33} & -e^{-\pi ki}a_{33} \\ 0 & 0 & -e^{\pi ki} & -e^{-\pi ki} & e^{\pi ki} & e^{-\pi ki} \\ 0 & 0 & 0 & 0 & e^{2\pi ki}\chi_3^{1-k} & e^{-2\pi ki}\bar{\chi}_3^{1-k} \end{bmatrix}_{6\times 6}. \quad \text{(A8)}$$

In Eq. (A8) the matrix $\mathbf{K}_a$ is associated with $\mathbf{q}_a = \{M_3^{(1A)}, N_3^{(1A)}, M_3^2, N_3^2, M_3^{(1B)}, N_3^{(1B)}\}^T$ while the determinant of $\mathbf{K}_a$ is given by

$$|\mathbf{K}_a| = -4a_{33}^2\chi_3^{1-k}\bar{\chi}_3^{1-k}(e^{2\pi ki} - e^{-2\pi ki}). \quad \text{(A9)}$$

Hence, the root of $|\mathbf{K}_a| = 0$, in the range $0 < \text{Re}\,(k) < 1$, is $k = 0.5$ as shown in Figs. 5–8.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Zhu Wu
Ph.D. Student,
wu@univ-troyes.fr

Jian Lu
Professor,
lu@univ-troyes.fr,
Mem. ASME

LASMIS Laboratory,
Department of Mechanical
Systems Engineering,
University of Technology of Troyes,
12, rue Marie Curie,
BP2060,
10010 Troyes Cedex, France

Bongtae Han
Associate Professor,
Mechanical Engineering department,
248 Flour Daniel Building EIB,
Clemson University,
Clemson, SC 29634
bhan@clemson.edu
Mem. ASME

Study of Residual Stress Distribution by a Combined Method of Moiré Interferometry and Incremental Hole Drilling, Part I: Theory

A new method combining Moiré interferometry and the incremental hole-drilling method is developed to determine both uniform and nonuniform residual stress distribution in depth. The study is reported in two parts. In this first part, the theoretical development of the moiré interferometry hole-drilling method is presented. The relationship between the in-plane surface displacements produced by introducing a blind hole and the corresponding residual stresses is established by employing the existing theoretical solution containing a set of undetermined coefficients. The coefficients are calibrated by the three-dimensional finite element method and they are processed nondimensionally for general use. The whole field in-plane surface displacements data U_x and U_y produced by the relaxation of residual stresses are obtained from moiré interferometry after each increment of hole drilling. The high signal-to-noise ratio provided by moiré interferometry allows accurate determination of fringe orders near the hole boundary which is essential for enhancing fidelity of residual stress determination. The experimental procedure to determine the signs of residual stresses is described and the accuracy of the method is also discussed.

Introduction

The hole-drilling method of residual stress measurement was first introduced in the 1930s. In the original attempts, the uniaxial residual stress was determined approximately by measuring the changes in the diameter of a drilled hole (Mathar, 1934). Soon after, the strain gage method was adopted to measure surface deformations around a drilled hole due to the relaxation of residual stress (Soete, 1949; Soete and Vancrombrugge, 1950). In this method, the theoretical solution for an infinite plate with a small through hole was used to establish the relationship between experimental strain data and residual stresses. The method has been improved continuously since the size of strain gages became smaller and standard rosettes became available. However, Soete's through hole strain gage method had inherent disadvantages. First, residual stresses are self-balanced stresses; it is often inappropriate to assume a uniform distribution of residual stresses throughout the thickness of a plate. Secondly, for many industrial applications, a through hole often renders the tested structure or component useless.

In some cases, residual stresses exist only in the vicinity of surface layers, where a through hole approach would mislead the results significantly. In order to make it suitable to a blind hole problem, the experimental calibration method was developed to determine the nonuniform distribution of residual stresses in depth (Kelsey, 1956). In this method, an extra calibration experiment was introduced, assuming that the average relaxation of residual stresses for any incremental depth of a tested specimen would produce the same strain increment as in the calibration specimen with the external loads acting on the ends. Indeed, this method provided an improved means for determination of residual stress. Several followers (Rendler and Vigness, 1966; Gupta, 1968, Bert and Thomson, 1968; Newwar and Shewchuk, 1978) also made contributions to improve this method for new applications.

The experimental calibration methods were time-consuming and they were often impractical to prepare twin specimens with and without residual stresses. Furthermore, the surface deformation, around a blind hole, caused by external loads acting on the ends, would not represent the deformation caused by the relaxation of nonuniform residual stresses in depth. Numerical coefficient calibration methods were developed to cope with the problems (Schajer, 1981, 1988a, b; Niku-Lari, Lu, and Flavenot, 1985; Lu and Flavenot, 1989). In the methods, external loads were applied to the side surface of the hole to simulate the relaxation of residual stresses, which produced a more accurate residual stress distribution. The integral finite element calibration method combined with strain gage incremental hole drillings was also developed to determine nonuniform residual stress distributions in depth (Niku-Lari, Lu, and Flavenot, 1985; Flaman and Manning, 1985; Lu and Flavenot, 1986, 1989; Schajer, 1988a, b; Lu, ed., 1996). This method supported by a powerful software has been widely used in industry. It has been reported recently that the accuracy of this method can be further improved by employing the three-pair rosette method (Schajer and Tootoonian, 1997).

In the past decades, numerous laser techniques for displacement measurement have been developed for small deformation analyses. They include holographic interferometry, laser speckle interferometry and moiré interferometry, etc. Recently, the applicability of these methods has been extended to residual stress measurement to cope with the problems associated with the strain gage method. Holographic interferometry was first adopted for residual stress measurement (Antonov, 1983; Nel-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, June 17, 1997; final revision, Mar. 31, 1998. Associate Technical Editor: I. M. Daniel.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

son et al., 1986). An improved method combining the single axis holographic interferometry and blind hole-drilling method was proposed to determine residual stresses in depth (Makino and Nelson, 1994; Nelson et al., 1994). Later the method was developed in conjunction with incremental hole drillings for an elastic-plastic bending residual stress problem (Makino and Nelson, 1997). As indicated in their articles, however, the displacement measurement was carried out at a distance from the hole boundary due to inherent optical noise produced by the technique. Moiré interferometry was also utilized to determine residual stresses (Furgiuele, Pagnotta, and Poggialini, 1991; Nicoletto, 1988, 1991). In their studies, the through holes were introduced and the closed-form solution was obtained by assuming a uniform residual stress distribution throughout the thickness of a plate.

The objective of this paper is to introduce a practical and accurate method for determining the residual stress distribution in depth with high fidelity. The method combines moiré interferometry with the incremental hole-drilling method. The paper is divided into two parts. In this first part a theoretical development of the moiré interferometry incremental hole-drilling method is presented. The relationship between the in-plane surface displacement and residual stresses is derived. A detailed procedure of determining calibration coefficients and residual stresses is described.

Background: Moiré Interferometry

Moiré interferometry is an optical method, providing real-time and whole-field contour maps of in-plane displacements. The high displacement measurement sensitivity and high spatial resolution made it suitable for a broad range of problems in solid mechanics (Post, Han, and Ifju, 1994). The principle of moiré interferometry is depicted schematically in Fig. 1. The specimen grating of frequency f_s is replicated or made directly on the surface of the specimen. The specimen grating intercepts two beams of mutually coherent light which are incident at symmetrically. When the relation of Eq. (1) is satisfied, the specimen grating diffracts the incident light such that +1 and −1 diffraction orders emerge normal to the specimen (w_1' and w_2' in Fig. 1),

$$\sin \alpha = f_s \lambda, \tag{1}$$

where α is the incident angle, λ is the wavelength of the laser, and f_s is the frequency of the specimen grating.

The two diffracted beams coexist in the space but their angle of intersection is zero. Their mutual interference produces a uniform intensity throughout the field. When the specimen deforms, the specimen grating deforms together with it, which results in warped wavefronts of w_1'' and w_2''. The resulting interference patterns represent contours of constant U_x and U_y displacements. The displacements can then be determined from fringe orders by the following relationships:

$$U_x = \frac{1}{2f_s} N_x, \quad U_y = \frac{1}{2f_s} N_y \tag{2}$$

where N_x and N_y are fringe orders in the U_x and U_y fields, respectively. In routine practice, a frequency of 1200 lines/mm is used, providing a contour interval of 0.417 μm/fringe order.

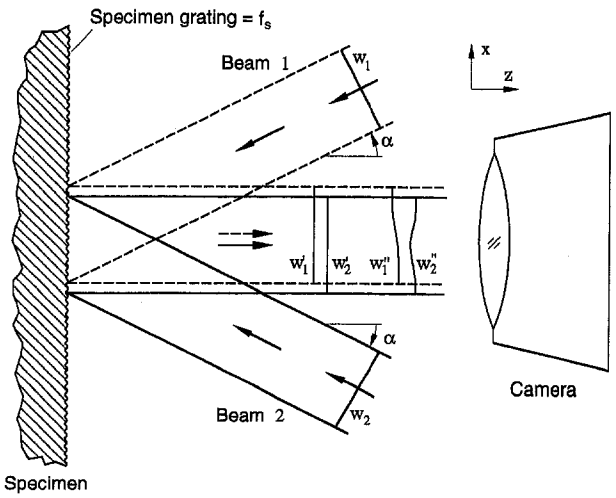


Fig. 1 Principle of moiré interferometry

Displacement Field for the Hole Drilling Method

For blind hole problems, a Fourier expansion solution was proposed by Schajer (1981) to account for the effect of out-of-plane displacement and was expressed as

$$u_r(r, \theta, z) = \sum_{n=0}^{\infty} u_{rn}(r, z) \cos n\theta$$

$$u_\theta(r, \theta, z) = \sum_{n=0}^{\infty} u_{\theta n}(r, z) \sin n\theta$$

$$u_z(r, \theta, z) = \sum_{n=0}^{\infty} u_{zn}(r, z) \cos n\theta \tag{3}$$

where u_{rn}, $u_{\theta n}$, and u_{zn} are the contributions of the nth displacement components to the total displacement components u_r, u_θ, and u_z, respectively, in the cylindrical coordinate system. The corresponding load distribution for each value of n has the same harmonic form, and thus $n = 0, \pm 2, \pm 4, \ldots$

As the first-order approximation, $n = 0, 2$, comparing the through hole solution when a biaxial residual stress exists, the above Fourier expansion can be expressed as

$$u_r(r, \theta) = A(\sigma_{xx} + \sigma_{yy}) + B[(\sigma_{xx} - \sigma_{yy}) \cos 2\theta + 2\tau_{xy} \sin 2\theta]$$

$$u_\theta(r, \theta) = C[(\sigma_{xx} - \sigma_{yy}) \sin 2\theta - 2\tau_{xy} \cos 2\theta]$$

$$u_z(r, \theta) = F(\sigma_{xx} + \sigma_{yy}) + G[(\sigma_{xx} - \sigma_{yy}) \cos 2\theta + 2\tau_{xy} \sin 2\theta] \tag{4}$$

where σ_{xx}, σ_{yy}, and τ_{xy} are the residual stress components in the Cartesian coordinate system; u_r, u_θ, and u_z are the surface displacements around the hole in the cylindrical coordinate system; and A, B, C, F, and G are undetermined coefficients which depend on the material constants and the geometrical parameters of the blind hole (Makino and Nelson, 1994).

Determination of Calibration Coefficients

The coefficients A, B, C, F, and G in Eq. (4) can be determined by three-dimensional finite element analyses by using two special loading cases:

1 $\sigma_{xx} = \sigma_{yy} = \sigma$, $\tau_{xy} = 0$, an equibiaxial residual stress field, which is equivalent to a uniform pressure $p = \sigma$ acting on the hole boundary as shown in Fig. 2. By using Eq. (4), the coefficients A and F are determined as

$$A\left(E, \nu, r_0, r, \frac{h}{d}\right) = \frac{u_r(r, \theta)}{2\sigma}$$

$$F\left(E, \nu, r_0, r, \frac{h}{d}\right) = \frac{u_z(r, \theta)}{2\sigma}. \tag{5}$$

(2) $\sigma_{xx} = -\sigma_{yy} = \sigma$, $\tau_{xy} = 0$, a pure shear residual stress field, which is equivalent to the harmonic distributions of the normal stress $\sigma_{rr} = -\sigma \cos 2\theta$ and the shear stress $\tau_{r\theta} =$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

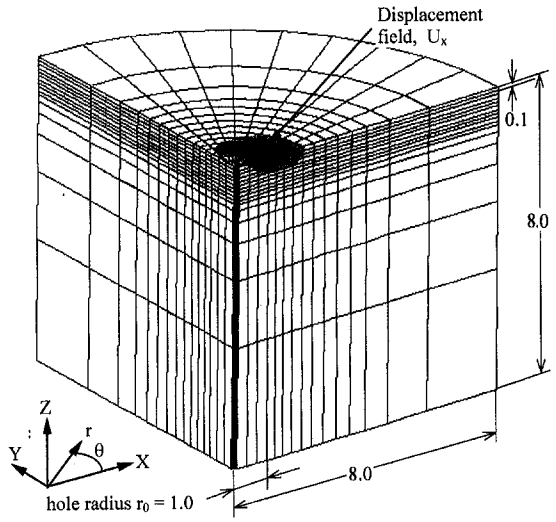


Fig. 2 Finite element model of a specific loading condition, a uniform pressure acts on the first incremental layer, used to determine calibration coefficients; the removal of material by hole drilling was substituted by elements with tiny Young's modulus

$\sigma \sin 2\theta$ acting on the hole boundary. By using Eq. (4), the coefficients B, C, and G are determined as

$$B\left(E, \nu, r_0, r, \frac{h}{d}\right) = \frac{u_r(r, \theta)}{2\sigma \cos 2\theta},$$

$$C\left(E, \nu, r_0, r, \frac{h}{d}\right) = \frac{u_\theta(r, \theta)}{2\sigma \sin 2\theta},$$

$$G\left(E, \nu, r_0, r, \frac{h}{d}\right) = \frac{u_z(r, \theta)}{2\sigma \cos 2\theta}. \quad (6)$$

In Eqs. (5) and (6), u_r, u_θ, and u_z are the surface displacements calculated from the finite element analysis. It is to be noted that the calibration coefficients are independent of angle θ. The coefficients depend on the material constants of Young's modulus E, Poisson ratio ν, the geometrical parameters of the hole radius r_0, hole depth-to-diameter ratio h/d, and the radio coordinate r. They can be converted into nondimensional parameters so that they can be used for different materials and hole diameters with the same h/d parameter. According to the theoretical solution, the Poisson's ratio should be coupled in Eq. (4). For most engineering materials, Poisson's ratio does not vary significantly, the nondimensionally processed coefficients can be expressed as

$$A' = \frac{E}{r_0}A, \quad B' = \frac{E}{r_0}B, \quad C' = \frac{E}{r_0}C, \quad F' = \frac{E}{r_0}F, \quad G' = \frac{E}{r_0}G.$$

Moiré interferometry provides in-plane displacement data on the entire region of interest. A finite element analysis was conducted to define the ideal locations of the calibration point for optimum accuracy. A three-dimensional finite element model of eight-node isoparametric brick elements was constructed by using a finite element code ABAQUS-standard 5.6 and ABAQUS-pre. In the analysis, a blind hole-drilling procedure was simulated by replacing elements representing removed material with zero Young's modules. The parameters used in the analysis were σ = 100 MPa; $E = 7.0 \times 10^4$ MPa; ν = 0.3; r_0 = 1.0 mm; incremental hole depth = 0.1 mm. Figure 3 shows results for the two special loading conditions, obtained after the first incremental step. As can be seen, the relaxed surface displacements are highly localized in a small region around the blind hole. Furthermore, the displacement gradients vary at different radial calibration points and diminish rapidly as r increases. This implies that it is essential to use displacement data at the points very close to hole boundary to achieve high accuracy in residual stress measurement. The results also indicate that u_z is always smaller than u_x or u_y and its gradient diminishes more rapidly. The out-of-plane displacement can be measured by other interferometric techniques such as the Twyman/Green interferometry. However, the sensitivity would not be sufficient to produce a sufficient number of fringes for accurate determination of residual stress (Wu and Lu, 1997).

The above results imply that the optimum combination of displacement data should be in-plane displacement data determined at the same radial coordinate r_c with different polar angle θ and can be written as

$$u_r(r_c, \theta_k) = (A + B \cos 2\theta_k)\sigma_{xx} + (A - B \cos 2\theta_k)\sigma_{yy} + (2B \sin 2\theta_k)\tau_{xy}, \quad k = 1, 2, 3; \quad (7)$$

and

$$u_\theta(r_c, \theta_k) = C[(\sigma_{xx} - \sigma_{yy}) \sin 2\theta_k - 2\tau_{xy} \cos 2\theta_k], \quad k = 1, 2, 3. \quad (8)$$

Generally Eqs. (7) and (8) can be used independently or they can be used together to determine the three residual stress components. Equations (7) and (8) indicate that more than three linear homogeneous equations are available since more than three displacement data can be obtained from the displacement contour maps of U_x and U_y. The residual stress can then be determined from different simultaneous equations independently, the average value of the results would eliminate the random errors in displacement measurement. However, for simplicity, only Eq. (7) will be used in the following sections.

Determination of Uniform Residual Stress

Moiré interferometry provides displacement data with the Cartesian coordinates. The radial displacement U_r can be expressed in terms of U_x and U_y as

$$U_r(r_c, \theta_k) = U_x(x_k, y_k) \cos \theta_k + U_y(x_k, y_k) \sin \theta_k \quad (9)$$

where

$$(x_k)^2 + (y_k)^2 = (r_c)^2 \text{ and } \theta_k = \tan^{-1}\left(\frac{y_k}{x_k}\right), \quad k = 1, 2, 3.$$

By combining Eqs. (2), (7), and (9), the relationship between the residual stress and the fringe orders can be rewritten in a matrix form as

$$[N_x(x_k, y_k) \quad N_y(x_k, y_k)] \begin{bmatrix} \cos \theta_k \\ \sin \theta_k \end{bmatrix} = 2f_s[A + B \cos 2\theta_k \quad A - B \cos 2\theta_k \quad 2B \sin 2\theta_k] \times \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \tau_{xy} \end{bmatrix}, \quad k = 1, 2, 3. \quad (10)$$

The three unknown residual stress components σ_{xx}, σ_{yy}, τ_{xy} can be determined by solving the above three linear homogeneous equations. The principal stresses σ_1, σ_2 and their direction φ then can be determined as

$$\sigma_1, \sigma_2 = \frac{\sigma_{xx} + \sigma_{yy}}{2} \pm \sqrt{\left(\frac{\sigma_{xx} - \sigma_{yy}}{2}\right)^2 + \tau_{xy}^2}$$

$$\tan \varphi = \frac{\sigma_1 - \sigma_{xx}}{\tau_{xy}}, \quad (11)$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

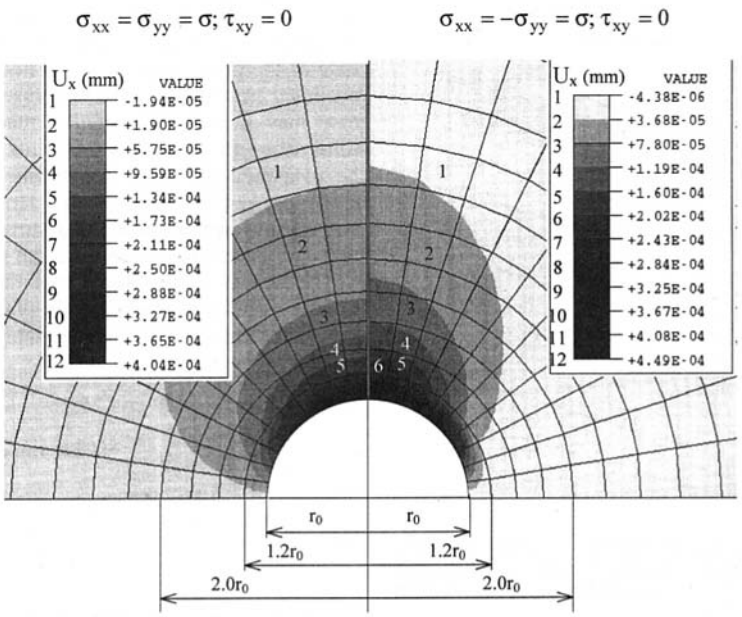


Fig. 3 Contour maps of U_x for the two specific loading conditions when hole depth $h = 0.1$ mm

where φ is the angle between the maximum principal stress σ_1 and coordinate X (the directions of the specimen grating are coincident with the directions of coordinates X and Y in the Cartesian coordinate system).

Determination of Nonuniform Residual Stress

In many cases, residual stress gradients in the depth direction can be very large. For example, in surface coatings or shot peening processes, the magnitude of residual stress changes from its maximum value to zero within a very thin surface layer. The method described in the previous section can only determine an average value of residual stress over the depth of the hole. One can consider using the method to determine residual stress distribution by employing a series of separate blind holes with different hole depths. However, this approach can provide only approximate solutions since residual stresses released from each incremental hole-drilling step produce nonlinear contributions to existing displacement fields. This can be further explained by using the same finite element model used in the previous calibration process. An identical uniform pressure $p = \sigma$ was applied on different layers to simulate the relaxation of the equibiaxial residual stresses ($\sigma_{xx} = \sigma_{yy} = \sigma$, $\tau_{xy} = 0$) for two incremental steps. The results are plotted in Figs. 4–6. It is important to note that $A^{21} > A^{11}$, $B^{21} > B^{11}$ and $A^{21} > A^{22}$, $B^{21} > B^{22}$, where the first and second superscripts denote the incremental step and the number of layers on which the pressure was applied, respectively. This analysis clearly indicates that the nonuniform residual stress cannot be determined accurately by the method. A similar conclusion was reported for the strain gage hole-drilling method (Schajer, 1998a, b; Lu and Flavenot, 1989). However, it should be emphasized that, in moiré interferometry, this nonlinear effect would be more significant since the displacement data are to be obtained at the points very close to the hole boundary.

The surface displacements of any point around the hole can be measured continuously by moiré interferometry while the depth of a blind hole is increased by a known increment. In reference to the solution of Eq. (10), the relationship between the total surface displacements and the residual stresses of each incremental layer can be expressed as

$$[N_x^i(x_k, y_k) \quad N_y^i(x_k, y_k)] \begin{bmatrix} \cos\theta_k \\ \sin\theta_k \end{bmatrix} = \sum_{j=1}^{i} 2f_s \times [A^{ij} + B^{ij}\cos 2\theta_k \quad A^{ij} - B^{ij}\cos 2\theta_k \quad 2B^{ij}\sin 2\theta_k] \times \begin{bmatrix} \sigma_{xx}^j \\ \sigma_{yy}^j \\ \tau_{xy}^j \end{bmatrix}, \quad k = 1, 2, 3; \quad i = 1, 2, \ldots, n \quad (12)$$

where n is the total number of incremental steps; A^{ij}, B^{ij} are the calibration coefficients of the jth layer after the ith incremental step is drilled; $N_x^i(x_k, y_k)$ and $N_y^i(x_k, y_k)$ are the fringe orders of the U_x and U_y fields, respectively, after the ith incremental step is drilled; f_s is the frequency of the specimen grating; and σ_{xx}^j, σ_{yy}^j, and τ_{xy}^j are the residual stress components

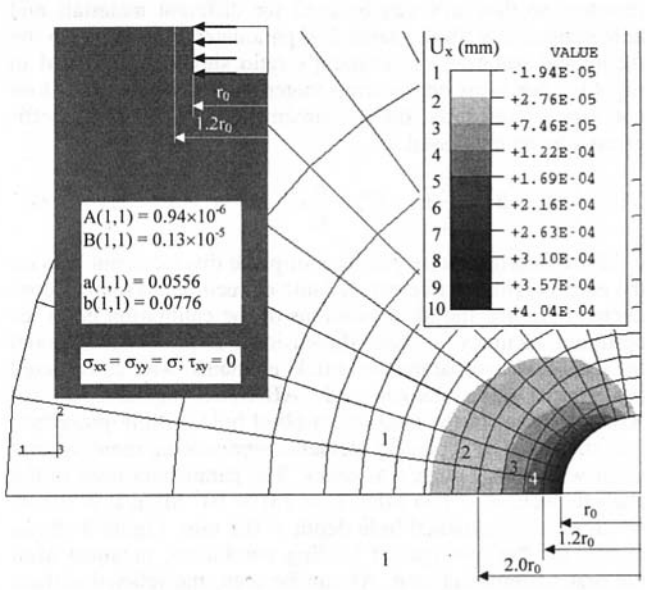


Fig. 4 Contour map of U_x displacement when a uniform pressure acts on the first incremental layer for a single-step hole drilling

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

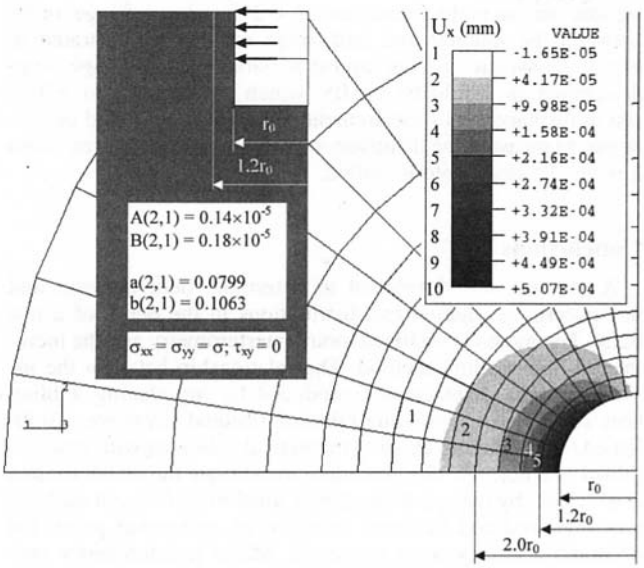


Fig. 5 Contour map of U_x displacement when a uniform pressure acts on the first incremental layer after the second incremental layer is drilled

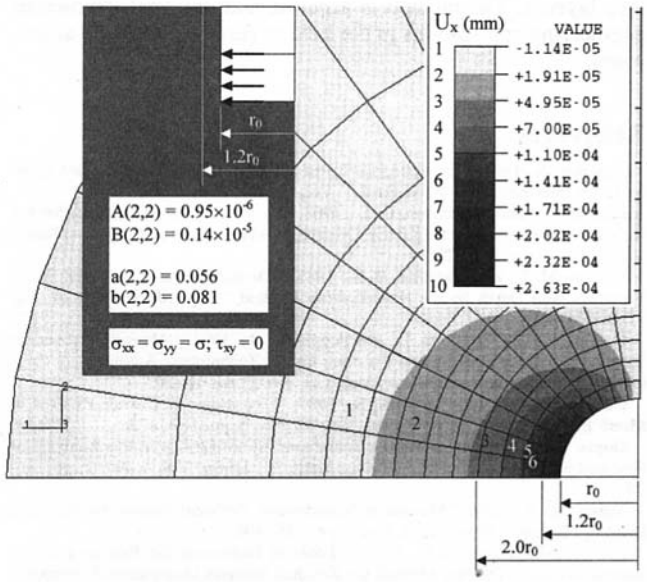


Fig. 6 Contour map of U_x displacement when a uniform pressure acts on the second incremental layer after the second incremental layer is drilled

of jth layer. The corresponding principal stress and its direction can be determined by using Eq. (11).

As can be seen from Eq. (12), when the total number of n increments are used, the same number of finite element models with different hole depths are required and a total of $2\sum_{n=1}^{n} n$ times of finite element analyses are required to calibrate the coefficients. In addition, $\sum_{n=1}^{n} n$ groups of the three linear homogenous equations are to be solved for determination of residual stress at each layer. These calculations can be done automatically by a pre-programmed computer routine.

Determination of the Sign of Residual Stress

Moiré fringes produced by one set of residual stresses can be identical to those produced by another set of residual stresses of the same magnitude with the opposite sign. The direction of the fringe gradient should be determined during the experiments since the sign of residual stresses was not known previously. In moiré interferometry, carrier fringes can be used effectively for the analysis of complex fringe patterns (Guo et al., 1989). They can be obtained by changing an incident angle (carrier fringes of extension and compression) and by rotating the specimen grating or by rotating one incident beam (carrier fringes of rotation). The results are a series of uniformly spaced fringes parallel (carrier fringes of extension and compression) or perpendicular (carrier fringes of rotation) to the grating lines. The resultant displacement field represents a constant displacement gradient. When carrier fringes are added to a load-induced fringe pattern, the resultant fringe gradient vector in every point of the fringe pattern will be the superposition of the original and the carrier fringe gradient vectors.

An example is illustrated in Fig. 7, where carrier fringes of extension and compression were added to the right and left side of a typical U_x fringe pattern, respectively. The fringe gradient in the X direction was increased in the right side, as revealed by closer fringe spacing near the hole, while the fringe gradient in the left side was decreased near the hole boundary, as revealed by a saddle-type fringe pattern; the positive gradient near the hole boundary was canceled by the negative gradient of

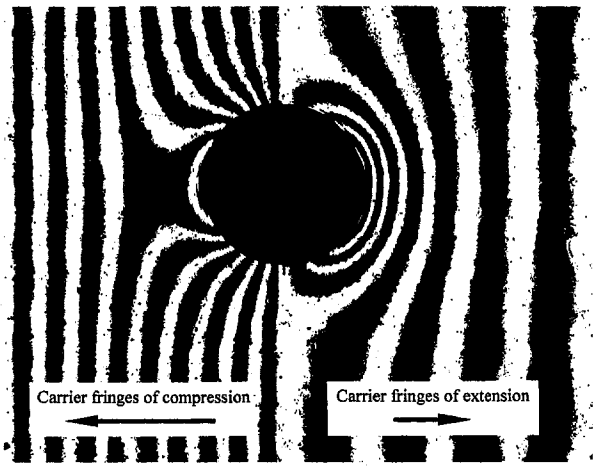


Fig. 7 U_x field moiré pattern modified by carrier fringes of compression (left side) and extension (right side)

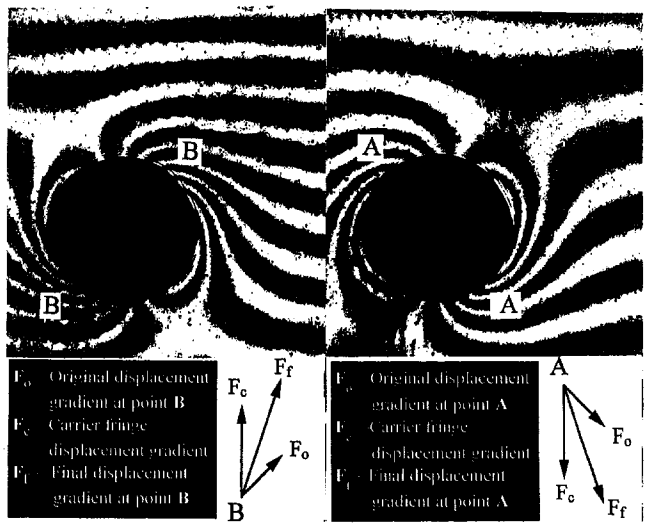


Fig. 8 U_x field moiré pattern modified by carrier fringes of rotation with opposite signs. The inserts illustrate the resultant fringe gradient vector after carrier fringes are added.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

carrier fringes. Another example is shown in Fig. 8, where the moiré fringes are modified by two known carrier fringes of rotation with opposite sign. By examining the resultant fringe gradient vector, the signs of the fringe gradient at points A and B can be readily determined, as illustrated in the inserts. Determination of the sign of the fringe gradient is necessary only for the first incremental step. The sign of fringe gradient for the following increments can be determined by comparing the variation of fringe patterns. The sign of the residual stress components σ_{xx} and σ_{yy} and τ_{xy} are consequently determined by solving Eq. (12).

Fringe Counting and Accuracy

In the proposed method, $r_c = 1.2r_0$ is suggested for displacement measurements. The radius was based on numerous experimental observations. A typical magnified U_x field moiré fringe pattern around the drilled blind hole is shown in Fig. 9. The plots in the inserts represent the intensity distributions along the lines shown in the figure. It is evident from the figure that the distance of $0.2r_0$ is sufficient to avoid uncertainties associated with small plastic zones produced during the drilling process.

As mentioned earlier, the surface displacement fields obtained after a first tiny increment are localized in a very small region around the hole. The accuracy in residual stress determination for the following incremental steps will be highly effected by the accuracy in the first step. The choice of angle θ for three points required in Eq. (12) can be arbitrary. However, if the angles between these points are relatively small, the condition numbers of Eq. (12) are relatively large, and thus produces large relative errors in determination of residual stress. In order to increase the accuracy of displacement measurement and to simplify the fringe counting and calculation process, the three characteristic points are suggested as $r_c = 1.2r_0$ with angles θ = 0 deg, 45 deg, 90 deg or $\theta = -45$ deg, 0 deg, 45 deg.

In moiré interferometry, integral and half-fringe orders can be readily determined. However, they are not sufficient for accurate determination of residual stresses with the incremental hole-drilling method since the change in fringe orders for each hole drilling increment is often smaller than a whole fringe order and furthermore the points for displacements measurement would not coincide with the locations of integer or half-fringe orders. Moiré interferometry features high signal-to-noise ratio. With the aid of intensity distribution analysis of the images captured by a high-resolution CCD camera, it is reasonable to assume an accurate extraction of 0.2 fractional fringe order between the integral and half-fringe orders, as illustrated in Fig. 9. Consequently, the largest possible error in fringe order determination would be ± 0.05, which corresponds to ± 0.02 μm in displacement measurement. This accuracy would be sufficient to measure the disturbance of each increment, even when a relatively deep hole is drilled.

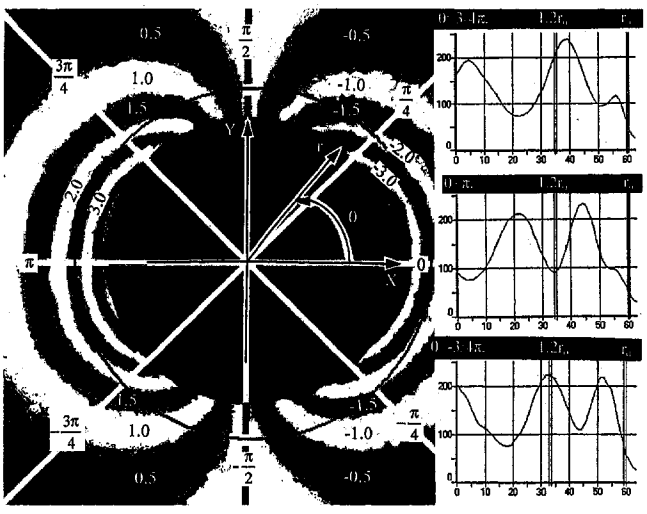


Fig. 9 Magnified view of U_x field moiré pattern around the hole. The plots represent the intensity distributions along the lines shown in the figure.

Conclusions

A method was developed to determine both uniform and nonuniform residual stress distributions in the depth of a test body. The method combined moiré interferometry and the incremental hole-drilling method. The relationship between the in-plane surface displacement produced by introducing a blind hole and the corresponding released residual stress was established by employing existing theoretical solutions with undetermined coefficients. The procedure to calibrate the undetermined coefficients by using a three-dimensional finite element analysis was reviewed and the ideal locations of calibration points for optimum accuracy were suggested. Moiré interferometry provides fringe patterns with high contrast and excellent visibility, even in the region very close to the boundary of the hole (which ensured the accuracy in determination of residual stresses in each incremental layers, especially for the surface and subsurface layers). The method is accurate and practical. Numerous engineering applications in the field of residual stress are anticipated.

References

Antonov, A., 1983, "Inspecting the Level of Residual Stresses in Welded Joints by Laser Interferometry, Weld Prod.," Vol. 30, No. 9, pp. 29–31.

Bert, C. W., and Thompson, G. L., 1968, "A Method for Measuring Planner Residual Stresses in Rectangularly Othotropic Materials," *J. Compos. Mater.*, Vol. 2, No. 4, pp. 244–253.

Flaman, M. T., and Manning, B. H., 1985, "Determination of Residual Stress Variation with Depth by the Hole-Drilling Method," *Exp. Mech.*, Vol. 25, No. 3, pp. 205–207.

Furgiuele, F. M., Pagnotta, L., and Poggialini, A., 1991, "Measuring Residual Stresses by HOLE-Drilling and Coherent Optics Techniques: A Numerical Calibration," *J. Eng. Mater. Techno.*, Vol. 113, No. 1, pp. 41–50.

Guo, Y., Post, D., and Czarnek, R., 1989, "The Magic of Carrier Patterns in Moiré Interferometry," *Exp. Mech.*, Vol. 29, No. 2, pp. 169–173.

Gupta, B., 1968, "Hole Drilling Technique: Modifications in the Analysis of Residual Stresses by the Hole-Drilling Method," *Strain*, Vol. 4, No. 2, pp. 20–29.

Kelsey, R. A., 1956, "Measuring Non-uniform Residual Stresses by the Hole Drilling Method," *Proc. SESA*, No. 1, pp. 181–194.

Lu, J., and Flavenot, J. F., 1986, "Trials in Increasing the Precision of the Incremental Hole-Drilling Method for Residual Stresses Distribution Determination," Residual Stresses in Science and Technology, Int. Conf. on Residual Stress 1986, Germany.

Lu, J., and Flavenot, J. F., 1989, "Applications of the Incremental Hole-Drilling Method for Measurements of Residual Stress Distribution," *Exp. Tech.*, Vol. 13, No. 11, pp. 18–24.

Lu, J., ed., 1996, *Handbook of Measurement of Residual Stresses*, SEM, Fairmont Press.

Makino, A., and Nelson, D., 1994, "Residual Stress Determination by Single Axis Holographic interferometry and Hole Drilling, Part I: Theory," *Exp. Mech.*, Vol. 34, No. 1, pp. 66–78.

Makino, A., and Nelson, D., 1997, "Determination of Sub-Surface Distributions of Residual Stresses by a Holographic-Hole Drilling Technique," *J. Eng. Mater. Techno.*, Vol. 119, No. 1, pp. 95–103.

Mathar, J., 1934, "Determination of Initial Stresses by Measuring the Deformations around Drilled Holes," *Trans. ASME*, Vol. 56, pp. 249–254.

Nelson, D. V., and McCrickerd, J. T., 1986, "Residual-Stress Determination Through Combined Use of Holographic Interferometry and Blind Hole Drilling," *Exp. Mech.*, Vol. 26, No. 4, pp. 371–378.

Nelson, D., Fuchs, E., Makino, A., and Williams, D., 1994, "Residual Stress Determination by Single Axis Holographic Interferometry and Hole Drilling, Part II: Experiments," *Exp. Mech.*, Vol. 34, No. 1, pp. 79–88.

Newwar, A. M., and Shewchuk, J., 1978, "On the Measurement of Residual Stress Gradients. in Aluminum-Alloy Specimen," *Exp. Mech.*, Vol. 18, No. 7, pp. 269–276.

Nicoletto, G., 1988, "Theoretical Fringe Analysis for a Coherent Optics Method of Residual Stress Measurement," *J. Strain Anal.*, Vol. 23, No. 4, pp. 169–178.

Nicoletto, G., 1991, "Moiré Interferometry Determination of Residual Stresses in the Presence of Gradients," *Exp. Mech.*, Vol. 31, No. 3, pp. 252–256.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Niku-Lari, A., Lu, J., and Flavenot, J. F., 1985, "Measurement of Residual Stress Distribution by the Incremental Hole Drilling Method," *Exp. Mech.*, Vol. 25, No. 2, pp. 175–185.

Post, D., Han, B., and Ifju, P., 1994, "High Sensitivity Moiré (Experimental Analysis for Mechanics and Materials)," Springer-Verlag, New York.

Rendler, N. J., and Vigness, I., 1966, "Hole Drilling Strain Gage Method of Measuring Residual Stresses," *Exp. Mech.*, Vol. 6, No. 12, pp. 577–586.

Schajer, G. S., 1981, "Application of Finite Element Calculations to Residual Stress Measurements," *J. Eng. Mater. Tech.*, Vol. 103, No. 4, pp. 157–163.

Schajer, G. S., 1988a, "Measurement of Non-Uniform Residual Stresses Using the Hole Drilling Method, Part I---Stresses Calculation Procedures," *J. Eng. Mater. Technol.*, Vol. 110, No. 4, pp. 338–343.

Schajer, G. S., 1988b, "Measurement of Non-Uniform Residual Stresses Using the Hole Drilling Method, Part II—Practical Applications of the Integral Method," *J. Eng. Mater. Technol.*, Vol. 110, No. 4, pp. 344–349.

Schajer, G. S., and Tootoonian, M., 1997, "A New Rosette Design for More Reliable Hole-Drilling Residual Stress Measurements," *Experimental Mechanics*, Vol. 37, No. 3, pp. 299–306.

Soete, W., 1949, "Measurement and Relaxation of Residual Stress," *Sheet Met. Ind.*, Vol. 26, No. 266, pp. 1269–1281.

Soete, W., and Vancrombrugge, R., 1950, "An Industrial Method for the Determination of Residual Stress," *Proc. SESA*, No. 1, pp. 17–18.

Wu, Z., and Lu, J., 1997, "A New Optical Method for Shot Peening Residual Stress Measurement," *Analysis of Residual Stresses from Materials to Bio-Materials, MAT-TEC 97*, Reins, Nov. 4–5, pp. 287–295.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Study of Residual Stress Distribution by a Combined Method of Moiré Interferometry and Incremental Hole Drilling, Part II: Implementation

Zhu Wu
Ph.D. Student,
wu@univ-troyes.fr

Jian Lu
Professor,
lu@univ-troyes.fr,
Mem. ASME

LASMIS Laboratory,
Department of Mechanical Systems Engineering,
University of Technology of Troyes,
12, rue Marie Curie,
BP2060,
10010 Troyes Cedex, France

Bongtae Han
Associate Professor,
Mechanical Engineering department,
248 Flour Daniel Building EIB,
Clemson University,
Clemson, SC 29634
bhan@clemson.edu
Mem. ASME

An experiment is devised to implement the combined method of moiré interferometry and incremental hole drilling which was proposed in the companion paper. A unique experimental apparatus is designed to combine an optical set up for moiré interferometry with an incremental hole-drilling device. The apparatus maintains a constant relative position of a drilling device to the specimen while alternating moiré interferometry measurements with incremental hole drilling. The drill-bit can reenter the hole for each incremental step. The incremental hole-drilling device produces a precise control of incremental hole depth with an accuracy of 0.1 μm. The method is utilized to measure the residual stresses of two shot-peened materials: AS10U3NG aluminum alloy and the ten percent SiC_p reinforced aluminum composite-F3K10S Duralcan with a T6 heat treatment. Moiré fringe patterns with excellent contrast and high signal-to-noise ratio are obtained, which allows the extraction displacement data at the points very close to a hole boundary. In the experiment, a total number of ten steps of incremental hole drilling with an identical increment of 0.1 mm are performed. The fringe patterns of U_x and U_y displacement fields are recorded after each increment and the displacement data at the points of $r = 1.2r_0$ are extracted consequently. They are used to determine the nonuniform residual stress distributions of surface and subsurface layers of shot-peened materials. The results indicate that the medium level of approximately equibiaxial compressive residual stresses exist within a 0.3 mm layer.

Introduction

The first part of the paper described in detail the theoretical basis of the moiré interferometry incremental hole-drilling method for determination of nonuniform residual stress distributions. Although the strain gage hole-drilling method has some advantages and has been applied for engineering residual stress problems for many years (Lu et al., 1990), the method proposed in this study has numerous advantages over the strain gage hole-drilling method, especially when the method is to be implemented for applications including shot-peening, coating, composite materials, and electronic packaging. When a blind hole is drilled in those applications, the surface deformation released by drilling a blind hole is highly localized near the hole boundary and it diminishes very rapidly. The nonlinear distributions of the deformations depend on both the hole depth-to-diameter ratio and the distributions of residual stresses in depths which was unknown previously. As the size of a standard strain gage rosette is typically two to four times that of the hole diameter, r_0 and 30 deg of θ (ASTM E837-92, 1992), the average strains measured over the area covered by the rosette often deviate from the exact strains to be used at the calibration points. The hole eccentricity is another source of error with the method.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, June 17, 1997; final revision, Mar. 31, 1998. Associate Technical Editor: I. M. Daniel.

The single-axis holographic interferometry hole-drilling method was introduced to cope with these problems (Makino and Nelson, 1994). The method was further developed in conjunction with incremental hole drilling (Makino and Nelson, 1997). Due to the poor contrast and low signal-to-noise ratio of fringe pattern in holographic interferometry, however, the choice of measurement points was based on a tradeoff between the need to obtain counts as high as possible to minimize experimental errors and the reduced ability to distinguish fringes at smaller r/r_0 ratios due to speckle effects. For the problems mentioned above, residual stresses with a high gradient can be often found in a very thin surface layer. Consequently, the high-measurement accuracy at points with a small r/r_0 ratio is essential to detect small displacement changes during hole drilling with a tiny increment (Wu, Lu, and Han, 1998).

Unlike the strain gage incremental hole-drilling method which can be done by using any standard hole-drilling machine, moiré interferometry requires that the incremental hole-drilling procedures should be combined with the optical system and performed alternately with moiré interferometry measurements so that the replacing accuracy for each increment could be ensured. This is an important reason why moiré interferometry was not used to treat incremental hole-drilling problems in the past (Nicoletto, 1991). In this study, a unique experimental apparatus is devised to cope with the above-mentioned problems.

Combined System of Moiré Interferometry and Hole-Drilling Device

For a uniform residual stress, a single step of blind hole drilling is sufficient. Generally, the drilling machine can be

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

separated from the optical system since the released permanent surface deformations can be recorded by moiré interferometry after drilling a hole. However, for the nonuniform residual stress distribution in the depth of a body, a combined system of moiré interferometry and the hole-drilling device is essential for the accurate determination of residual stress distributions.

The calibration coefficients determined by using the finite element method were based on a given equal increment for each step. They are highly sensitive to the hole depth-to-diameter ratio and they are valid only when the incremental depth-to-diameter ratio in the experiment is identical to the ratio used in the finite element model. Consequently, it is extremely important to maintain a constant position of a drilling device relative to the specimen and to provide a precise control of incremental hole depth during the entire incremental hole-drilling process.

A system combining an optical setup for moiré interferometry and a drilling device was designed to meet the above requirements. A schematic diagram of the experimental setup used in the hole drilling process is illustrated schematically in Fig. 1. The specimen (with its specimen grating) was mounted on a rotary stage for adjustments of angular orientation. This rotary stage was mounted on a rigid thick T-shape notched plate, which was fixed to the optical table. A small electric drilling machine with a variable speed of 5000 rpm to 50,000 rpm was connected to a heavy-duty magnetic base through a holding fixture, which in turn was positioned into a U-shaped base. The magnetic base provides a convenient and accurate way to remove and reposition the drilling machine subassembly, consisting of the drilling machine, the holding fixture, and the magnetic base. The ends of the magnetic base were aligned with the U-shaped base to achieve a high precision in repositioning of the subassembly. The U-shaped base was mounted on a stack of translation stages. The translation stages consisted of a manual driven XY micrometer stage, with a maximum traveling distance of 20 mm, and an automatic Z translation stage, with a maximum traveling distance of 200 mm. The XY stage was used for initial positioning of the drilling device at the desired location of the specimen. The Z translation stage was a stepper motor and controlled by a computer program, providing a precise movement of the drilling bit in the direction perpendicular to the specimen with an accuracy of 0.1 μm. This high precision was essential to ensure the constant hole depth increment at each step. The translation speed varies from 0.1 μm to 10.0 mm per second, thus any desired drilling speed can be achieved for different applications and materials. All of the components depicted in Fig. 1 were mounted on a vibration-isolated optical table. A flat mirror was also fixed to the optical table and it was used for in-plane moiré interferometry.

After each incremental hole-drilling process, the drilling machine subassembly was removed temporarily for moiré interferometry measurement. An optical arrangement for moiré interferometry is also depicted schematically in Fig. 1. In this arrangement, half the incident beam impinges directly on the specimen surface, while the other half impinges indirectly in a symmetrical direction after reflection from the plane mirror. Light diffracted in the ± 1 diffraction orders of the specimen grating, passing above the translation stages, was captured by the camera to form the U_x field moiré fringe pattern. The camera used in the system was a high resolution (1317 $\times$ 1035 pixels) digital Kodak CCD camera (MEGAPLUS Camera, Model 1.4i) and the image processing was performed on a PC platform, using commercial software IMAGE PRO PLUS (Version 1.3, Media Cybernetics). The resolution of the computer screen had a matching setup (1280 $\times$ 1024 pixels) with the CCD camera.

Test Specimen

Shot-peening processing is widely used in engineering to increase the working life under fatigue load or to enhance the resistance against surface corrosion of various metallic components in the aerospace industry. This is achieved by introducing moderate compressive residual stresses within the tiny near surface layer of the material (Bates, 1973; Fabr, 1972; Webster and Cunningham, 1993). The magnitude and distribution of residual stresses through a surface layer should be determined accurately so that the shot-peening process can be optimized to satisfy the requirements of different applications in engineering.

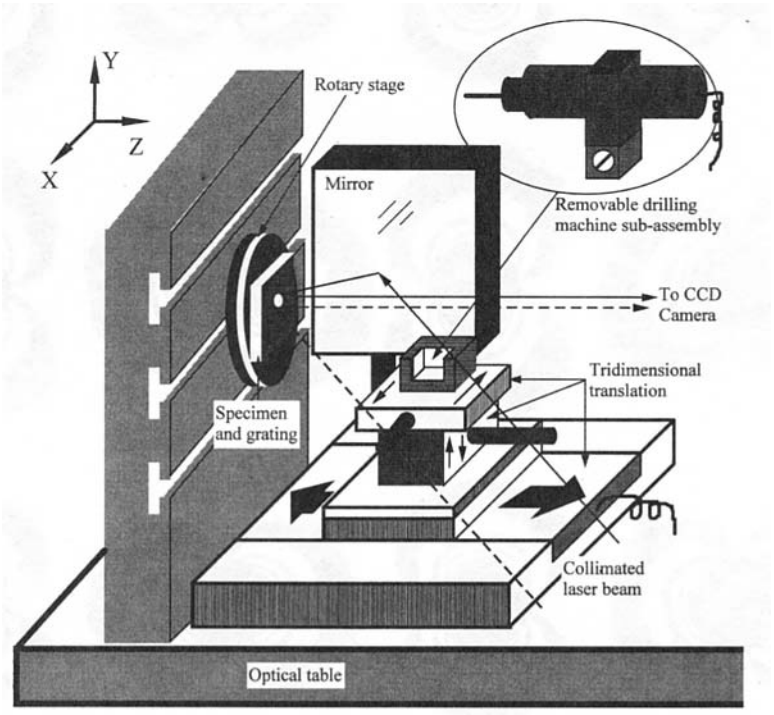


Fig. 1 Schematic diagram of optical and mechanical assembly of moiré interferometry and incremental hole-drilling device

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

The strain gage hole-drilling method, the X-ray diffraction method, and the neutron diffraction method have ever been used (Lu et al., 1990; Fathallah et al., 1994; Li et al., 1995). However, the stain gage hole-drilling method has drawbacks as introduced in the previous section. The X-ray diffraction method is not reliable for the materials with large grain size. The neutron diffraction method is an expensive technique and not very suitable for near-surface analysis.

The specimens were cut from large plates of AS10U3NG aluminum alloy and ten percent SiC_p reinforced aluminum alloy composite-F3K10S, Duralcan, one side of which was shot peened. The specimens were 60-mm square plates with a thickness of 13 mm. The material properties of each material was $E = 7.95 \times 10^4$ MPa, $\nu = 0.3$ and $E = 9.3 \times 10^4$ MPa, $\nu = 0.3$ for the AS10U3NG aluminum alloy the SiC_p reinforced aluminum alloy composite, respectively.

The holographic phase-type cross-line specimen grating with a frequency of 1200 lines/mm was replicated on the shot-peened surface of the specimen. In order to minimize the effect of shear lag through the epoxy grating thickness, the surface of the specimen was slightly polished on a lapping wheel before replication; then, a large force was applied to press the grating mold against the specimen while the epoxy was polymerized. This procedure ensured a specimen grating with a uniform thickness of less than 5 μm, which was much smaller than the incremental depth used in the experiment. Note that only a few microns of surface layer was removed during specimen perpetration and its effect on the residual stress distribution was negligible. If the surface of the specimen is not polished, the rough surface will be filled with epoxy during the replication process and then the maximum thickness of the grating layer could be 20.0 μm for a shot-peened surface. The influence of the extra layer can still be ignored.

Experimental Procedure

After mounting the specimen on the thick T-shape notched plate, the optical system was first adjusted to produce an initial null-field condition, which was devoid of fringes. The drilling machine subassembly was securely locked into the U-shaped base and the XY translation stage was adjusted to locate the drill bit at the desired position of the specimen.

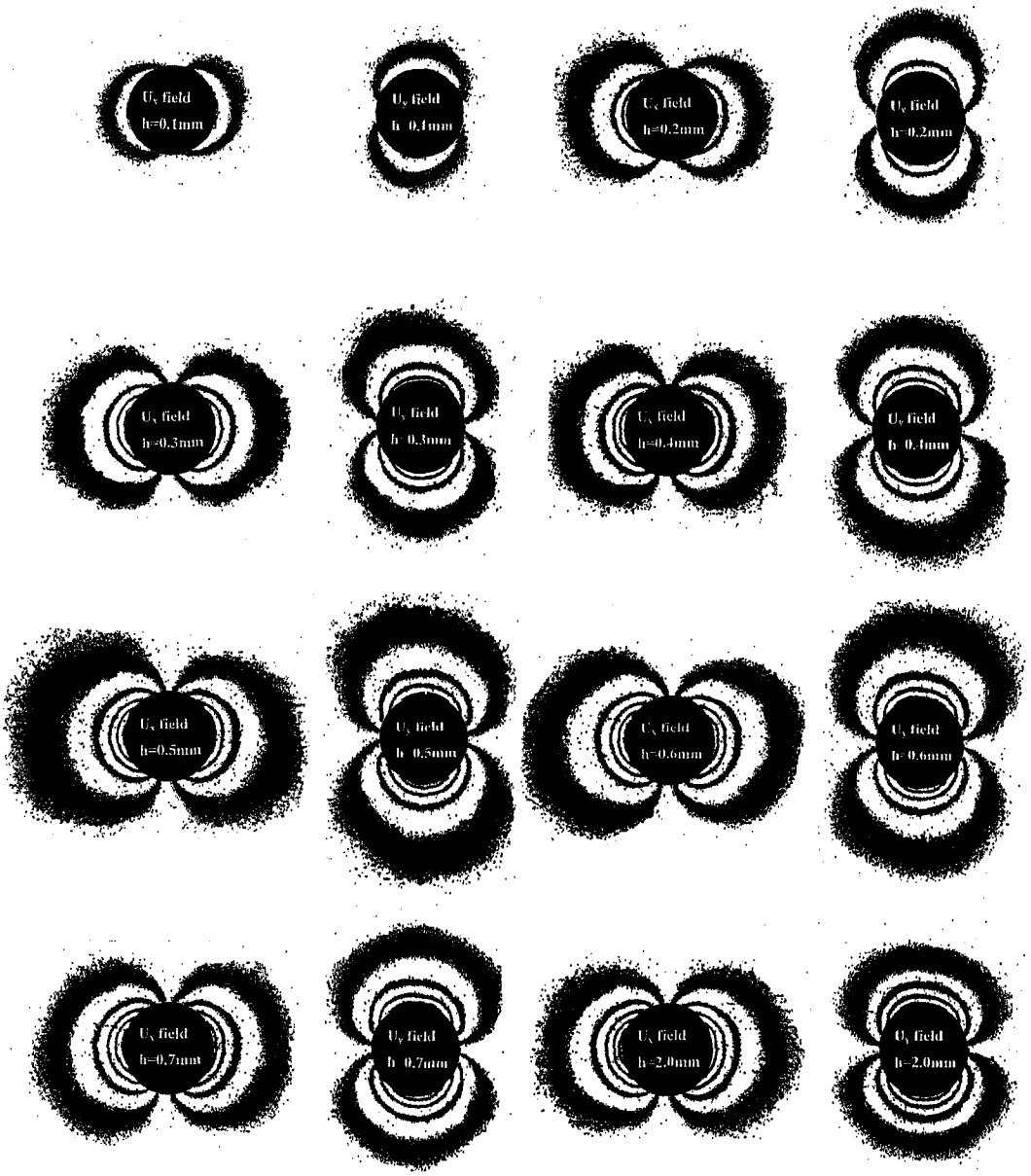


Fig. 2 U_x **and** U_y **displacement fields obtained from AS10U3NG aluminum alloy specimen**

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

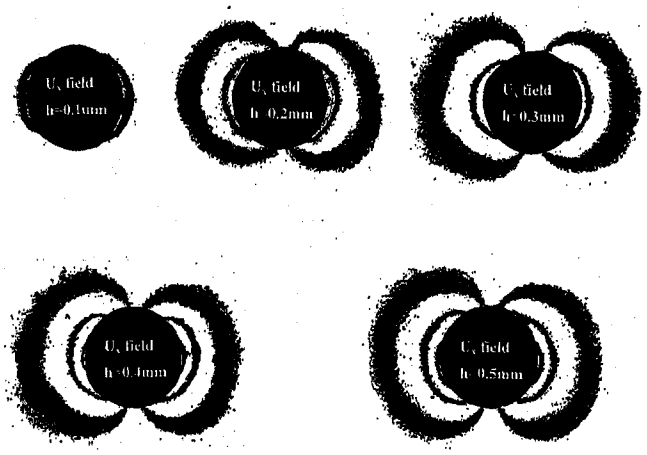


Fig. 3 U_x displacement fields obtained from ten percent SiC_p reinforced aluminum alloy composite-F3K10S, Duralcan specimen

The diameter of the drill bit was 2 mm and the drill was operated at 6000 rpm. The incremental depth was 0.1 mm and it was achieved by advancing the computer programmed Z translation stage at a speed of 1.0 μm/sec. The drilling machine subassembly was then removed and the U_x displacement field was recorded by moiré interferometry. The experiment continued with rotating the specimen by 90 deg and the corresponding U_y displacement field was obtained. It is important to note that a small amount of undesired displacements associated with a rigid-body rotation would be produced in the U_y displacement field even though the rotation of specimen was controlled precisely. This rigid-body displacement was routinely canceled by controlling a fine adjustment screw on the rotary stage until the region away from the hole became a uniform bright field since the original null-field state for the far field was not altered by the presence of the blind hole. After the U_y displacement was recorded, the specimen was repositioned to its precise original U_x field. The drilled hole was located at the center of the rotary stage, which ensured that the reposition error was less than 3.0 μm when the drill bit reentered the hole precisely for the next step-hole drilling. The total number of ten incremental steps was used in each experiment and the above procedure was repeated at each step.

The moiré fringes obtained from the AS10U3NG aluminum alloy specimen are presented in Fig. 2, where a contour interval was 0.417 μm per fringe order. In general, the in-plane residual stress produced by the shot-peening process is regarded as a uniform equibiaxial residual stress state ($\sigma_{xx} = \sigma_{yy} = -\sigma$, $\tau_{xy} = 0$) in macro scale. For this case, U_x and U_y displacement fields should be identical and they should also be symmetrical about the Y and X-axis, respectively. However, the U_x displacement was larger than the U_y displacement and the axial symmetry of the U_x displacement was slightly distorted. This was ascribed to the inherent local variations of the random nature of the shot-peening process. The results from the SiC_p reinforced aluminum alloy composite specimen are shown in Fig. 3. Only the U_x displacement fields are reported here since the U_y fields were identical to the corresponding U_x displacement fields.

Analysis and Results

The relationship between the fringe order and the residual stress, established in the first part of the paper, was used to determine the residual stress distribution and repeated here,

$$[N_x^i(x_k, y_k) \quad N_y^i(x_k, y_k)]\begin{bmatrix}\cos\theta_k \\ \sin\theta_k\end{bmatrix} = \sum_{j=1}^{i} 2f_s \times [A^{ij} + B^{ij}\cos 2\theta_k \quad A^{ij} - B^{ij}\cos 2\theta_k \quad 2B^{ij}\sin 2\theta_k] \times \begin{bmatrix}\sigma_{xx}^j \\ \sigma_{yy}^j \\ \tau_{xy}^j\end{bmatrix}, \; k = 1, 2, 3; \; i = 1, 2, \ldots, n. \quad (1)$$

After confirming the sign of the fringe gradient by using the carrier fringe technique, the fringe orders were assigned to the fringe patterns, as shown in Fig. 4. A circle with a radius of $1.2r_0$ was drawn in both fields and the fringe orders on the characteristic points were determined. These points are separated by 45 deg increments and they are denoted as A through G. As discussed previously, any three points around the hole can be chosen to be used in Eq. (1). However, the characteristic points listed above are often the most desired because the calculation can be much simplified and they provide optimum accuracy (Wu, Lu, and Han, 1998).

For the AS10U3NG aluminum alloy specimen, the U_x displacements were not symmetric. The nonuniform deformations were observed at the opposite points of the left and right sides of Fig. 2. The nonuniform deformations can be regarded as the result of the nonuniform residual stress field in the opposite side of the hole since the measurement points were very close to the boundary of the hole, $r = 1.2r_0$, where the displacements were mainly ascribed to the relaxation of the local residual stresses. The detailed explanation was given for a welding resid-

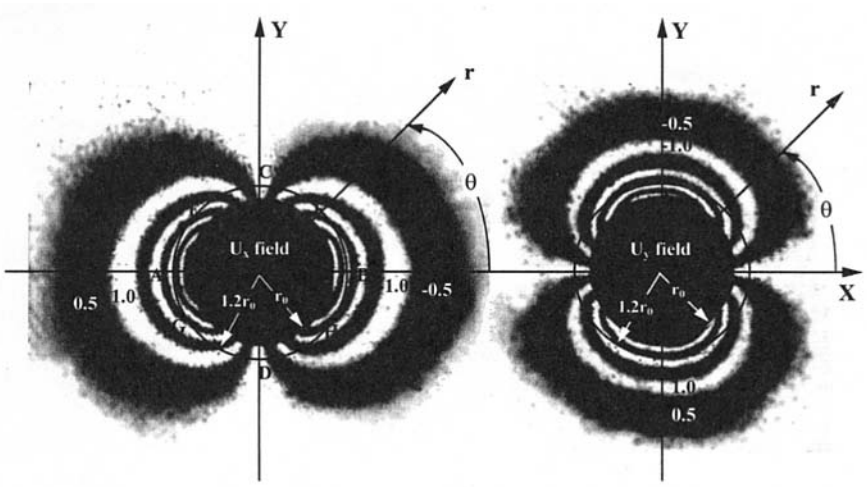


Fig. 4 A pair of typical moiré fringe patterns, U_x and U_y displacement fields used in the analysis

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

ual stress problem (Wu and Lu, 1998). Therefore, three points E ($1.2r_0$, 135 deg), A ($1.2r_0$, 180 deg) and G ($1.2r_0$, 225 deg) were chosen to calculate the residual stresses for the left side of the hole in the U_x field and the other three points F ($1.2r_0$, 45 deg), B ($1.2r_0$, 0 deg) and H ($1.2r_0$, −45 deg) were used for the right side. The residual stress distributions were finally determined and the results are plotted in Figs. 5(a) and 5(b) for the left and the right side, respectively. The maximum compressive residual stresses occurred within the first 0.1-mm incremental layer and the average level of compressive residual stress σ_{xx} was 274 MPa and 232 MPa for the left and right side, respectively. The variations of residual stresses are evident in the fringe patterns. The average level of compressive residual stress σ_{yy} in the first 0.1-mm incremental layer was 241 MPa and 221 MPa for the left and right side, respectively. The magnitude of the compressive residual stress decreased as the hole depth increased, and the residual stress changed its sign at approximately 0.35 mm from the surface. The compressive residual stresses were balanced by the tensile stress in deeper layers. The magnitude of the maximum shear stress was only 10 percent to 20 percent of the maximum compressive stress but it was sufficient to alter the symmetry in the displacement fields.

As mentioned earlier, the identical U_x, and U_y displacement fields were observed for the SiC_p reinforced aluminum alloy composite specimen. For this special case of planar equibiaxial residual stress, Eq. (1) can be further simplified as

$$N_x^i(1.2r_0, \theta) = \sum_{j=1}^{i} 4fA^{ij}\sigma^j. \tag{2}$$

It is clear from Eq. (2) that only one calibration coefficient A and the fringe order of a single point are sufficient to determine the residual stress. Point A ($1.2r_0$, 0 deg) was chosen for the calculation and results are plotted in Fig. 6. The magnitude of the maximum compressive residual stress and the stress distribution are similar to those obtained from the AS10U3NG aluminum alloy specimen. However, the released displacements for the same incremental hole drilling were smaller, as indicated by the fewer fringes appeared in the fringe patterns of Fig. 3. It was ascribed to the larger Young's modulus of the SiC_p reinforced aluminum alloy composite material.

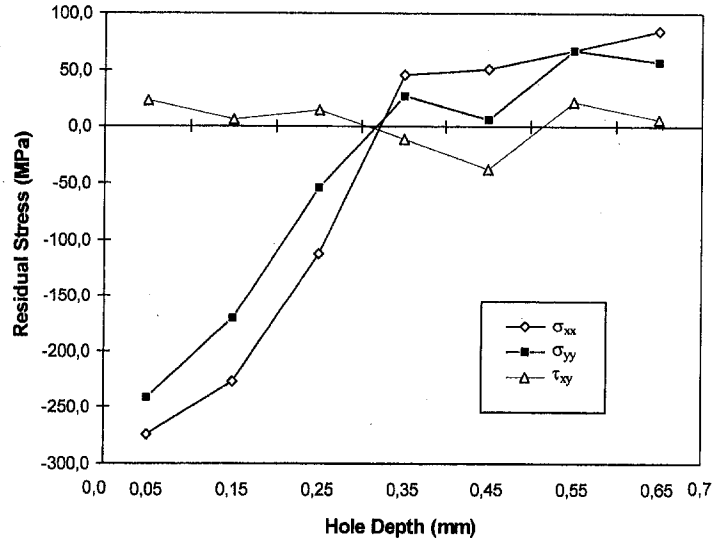


Fig. 5(a)

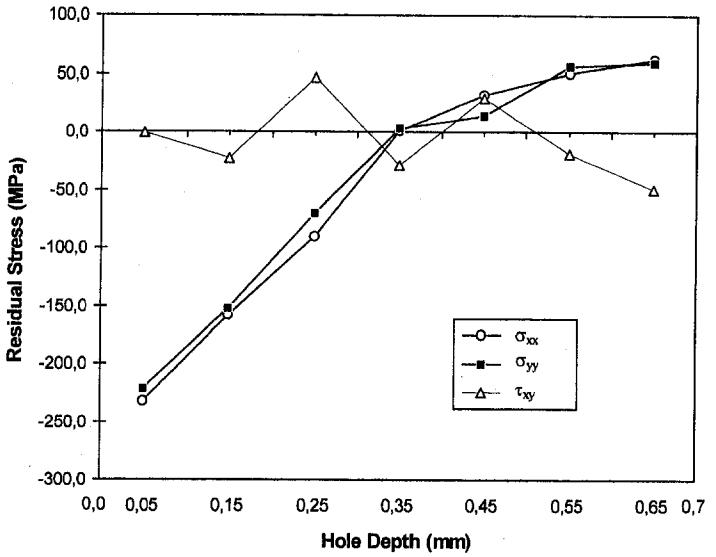


Fig. 5(b)

Fig. 5 Residual stress distribution of AS10U3NG aluminum alloy specimen, obtained from (a) the left side and (b) the right side of the blind hole

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

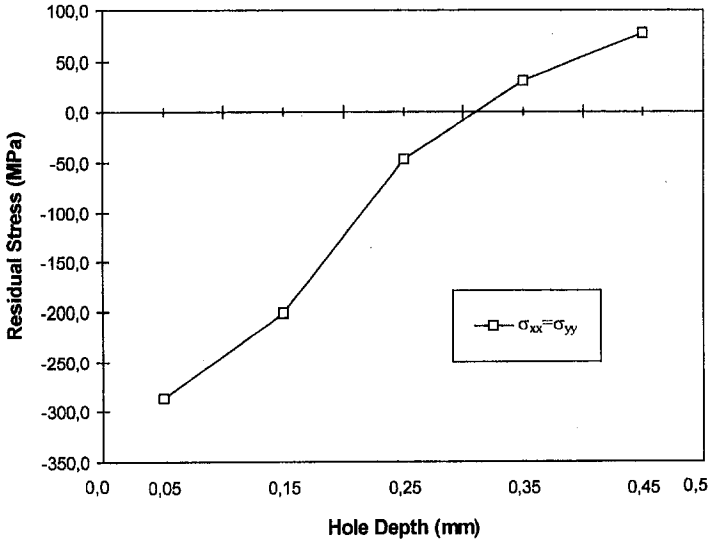


Fig. 6 Residual stress distribution of ten percent SiC_p reinforced aluminum alloy composite-F3K10S, Duralcan specimen

Discussions

The stresses reported in Figs. 5 and 6 are averaged values over each increment of 0.1 mm used in the experiments. Knowing that the stresses vary with a continuous distribution when approaching the surface, the actual peak compressive residual stress could be larger or smaller than the maximum compressive stresses reported in this study. If one desires to determine the maximum compressive residual stress more accurately, it is inevitable to reduce the increment. This reduction of the increment will result in releasing the proportionally lowered surface deformation. For the materials tested in this study, a change in the number of fringes was only 0.5 ~ 1.5 fringe orders for each increment with the given displacement sensitivity. Enhanced displacement sensitivity can be essentially achieved by using a UV laser with low wavelength or implementing a visible laser in a refractive medium (Post, 1994). The measurement sensitivity can also be improved by using the phase shifting method or based on the analysis of the intensity distribution of the fringe pattern.

In the analysis, the displacement data were extracted at the points along $r = 1.2r_0$. If the displacements were taken at the further distance from the hole boundary, the expected measurement errors would be much greater according to the localized nature of the released surface displacements. It is also important to note that the proposed method used the displacement data directly obtained from the fringe patterns, whereas strain values were used in the previous studies (Nicoletto, 1988, 1991). The differentiation of experimental data to calculate strain involves a loss of accuracy typically by an order of magnitude. The enhanced accuracy provided by the proposed method is evident.

When the hole depth became larger than 1.0 mm, no change in fringe patterns was observed. It was ascribed to the small residual stress induced by the shot-peening process and decreased sensitivity of residual stress in deeper layers to surface deformation.

Conclusions

An experimental apparatus combining an optical setup for moiré interferometry and an incremental hole-drilling device was developed to implement the proposed method of moiré interferometry and incremental hole drilling. The method was applied successfully for the determination of residual stress distribution in the depth of two shot-peened specimens-AS10U3NG aluminum alloy and the 10% SiC_p reinforced aluminum composite-F3K10S Duralcan with a T6 heat treatment. The nonuniform compressive residual stress distributions in the depth of the surface layer were obtained with an increment of 0.1 mm. The method used the displacement data obtained at the points $r = 1.2r_0$, which improved the accuracy of the residual stress determination significantly. The method is practical and accurate. A broad range of applications in the field of residual stress is anticipated.

Acknowledgments

Special thanks to the Institutions of CETIM and SNECMA, Metal Improvement Company, France, for their opinion on the design of hole drilling system, for their high quality shot-peened specimen, and for their financial support.

Reference

Ajovalasit, A., 1979, "Measurement of Residual Stresses By The Hole-Drilling Method: Influence of Hole Eccentricity," *J. of Strain Analysis,* Vol. 14, No. 4, pp. 171–178.

ASTM, 1992, *Standard Test Method for Determining Residual Stresses by the Hole Drilling Strain-Gage Method,* Annual Book of ASTM Standards, E837-92, Section 3, Vol. 03.01, American Society for Testing and Materials, Philadelphia, PA, pp. 747–753.

Bates, J. F., 1973, "Peening Stops Stress Corrosion Cracking in Stainless," *Iron Age,* No. 22.

Cheng, W., Finnie, I., Gremaud, M., Rosselet, A., and Streit, R. D., 1994, "The Compliance Method for Measurement of Near Surface Residual Stresses-Application and Validation for Surface Treatment by Laser and Shot-Peening," ASME *Journal of Engineering Materials and Technology,* Vol. 116, pp. 556–560.

Fabr, Ch. W., 1972, "Shot Peening Extends Fatigue Life of Tractor Parts," *Iron Age,* May 18.

Fathallah, R., et al., 1994, "Comparison of Residual Stresses Determined by X-ray diffraction, Neutron Diffraction and the Hole Drilling Method in Aerospace Shot peened Materials," *Proceedings of the Fourth International Conference on Residual Stresses,* June, Baltimore, MD, pp. 834–840.

Jaensson, B., and Ahmad, M., 1994, "Influence of Shot Peening on the fatigue Life and Surface Deformation Phenomena of Torsionally Loaded Aluminum Alloy Specimens," *Proceedings of the Fourth International Conference on Residual Stresses,* June, Baltimore, MD, pp. 1031–1037.

Li, J., et al., 1995, "Study of the Residual Stress in Cold-Rolled 7075 AL6SiC Whisker-Reinforced Composites by X-Ray and Neutron Diffraction," *J of Composites Technology and Research,* pp. 194–198.

Jinkui, L. et al., 1994, "An Understanding of the Effect of Residual Stress Induced by Shot Peening on Fatigue Strength," *Proceedings of the Fourth International Conference on Residual Stresses,* June, Baltimore, MD, pp. 1002–1007.

Lu, J., et al., 1990, "Study on the Effect of the Surface Treatment on the Residual Stress Gradient in Silicon Carbide (SiC) Reinforced Aluminum Metal Matrix Composites," *J. of Composites Technology and Research,* pp. 232–238.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Makino, A., and Nelson, D., 1994, "Residual stress Determination by Single Axis Holographic Interferometry and Hole Drilling, Part I: Theory," *Exp. Mech.,* Vol. 34, No. 1, pp. 66–78.

Makino, A., and Nelson, D., 1997, "Determination of Sub-Surface Distributions of Residual Stresses by a Holographic-Hole Drilling Technique," ASME *Journal of Engineering Materials and Technology,* Vol. 119, pp. 95–103.

Mathar, J., 1934, "Determination of Initial Stresses by Measuring the Deformations around Drilled Holes," *Transactions of the ASME,* Vol. 56, pp. 249–254.

Nelson, D., Fuchs, E., Makino, A., and Williams, D., 1994, "Residual stress Determination by Single Axis Holographic Interferometry and Hole Drilling, Part II: Experiments," *Exp. Mech.,* Vol. 34, No. 1, pp. 79–88.

Nicoletto, G., 1988, "Theoretical Fringe Analysis for a Coherent Optics Method of Residual Stress Measurement," *J. Strain Anal.,* Vol. 23, No. 4, pp. 169–178.

Nicoletto, G., 1991, "Moiré Interferometry Determination of Residual Stresses in the Presence of Gradients," *Exp. Mech.,* Vol. 31, No. 3, pp. 252–256.

Post, D., Han, B., and Ifju, P., 1994, *High Sensitivity Moiré (Experimental Analysis for Mechanics and Materials),* Springer-Verlag, New York.

Webster, P. S., and Cunningham, T. P., 1993, *Shot Peening: Technology and applications,* K. Marsh, ed., Engineering Materials Advisory Services Ltd.

Wu, Z., and Lu, J., 1998, "Study of Welding Residual Stress by High Sensitivity Moiré Interferometry and Incremental Hole-Drilling Method," *1998 SEM Spring Conf. on Exp. & Appl. Mech.,* June 1–3, Houston, pp. 282–284.

Wu, Z., Lu, J., and Guo, Y., 1998, "A New Method of Residual Stress Measurement on Electronic Packaging," *11th Int. Conf. on Exp. Mech.,* Aug. 25–28, Oxford, pp. 987–991.

Wu, Z., Lu, J., and Han, B., 1998, "Study of Residual Stress Distribution by A Combined Method of Moiré Interferometry and Incremental Hole Drilling, Part I: Theory," ASME JOURNAL OF APPLIED MECHANICS, Vol. 65, pp. 837–843.

Wu, Z., Lu, J., and Joulaud, P., 1997, "Study of Residual Stress Distribution by Moiré Interferometry Incremental Hole Drilling Method," *The Fifth Int. Conf. on Residual Stresses,* June 16–18, Linkoping, Sweden, pp. 804–809.

Wu, Z., Niu, X., Lu, J., and Ifju, P., 1998, "Study of Process-Induced Residual Stress in Orthotropic Composite Laminate—Carbon/Epoxy $[0_2/90_2]2s$," *1998 SEM Spring Conf. on Exp. & Appl. Mech.,* June 1–3, Houston, pp. 179–181.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

M. J. Leamy
Graduate Student Research Assistant

J. R. Barber
Professor,
Mem. ASME

N. C. Perkins
Associate Professor,
Mem. ASME

Department of Mechanical Engineering
and Applied Mechanics,
The University of Michigan,
2250 G. G. Brown,
Ann Arbor, MI 48109-2125

Distortion of a Harmonic Elastic Wave Reflected From a Dry Friction Support

This study is motivated by the need to understand the elastodynamic response of belts in frictional contact with pulleys. To this end, a simplified model for belt/pulley contact is used to investigate the dynamic response of a belt subject to a train of harmonic tension waves. Through a nondimensionalization, a single dimensionless parameter Ω is identified which governs the dynamic response. A numerical solution is developed and exercised over a wide range of values of Ω. An approximate closed-form solution is derived assuming the belt stretches quasi-statically, and is shown to yield accurate results for small values of Ω. Reported results include the distortion of an initially harmonic tension wave, the energy reflected from the frictional support, and the distance harmonic waves penetrate into the support. The results suggest that the quasi-static stretching assumption may be further utilized as a modeling simplification for belt drives characterized by values of $\Omega < \frac{1}{3}$.

1 Introduction

This study is principally motivated by belt drive mechanics which have been studied by many researchers, starting with Leonard Euler (1762). Euler proposed the well-known *capstan* formula for the tension distribution in a belt wrapped around a fixed pulley or a capstan (a fixed post used in docking a ship). A comprehensive survey of belt drive mechanics after Euler and up to 1981 is provided by Fawcett (1981). Much of the research cited in Fawcett (1981) and work since its appearance has been concerned with either (1) frictional contact (slip) between a belt and a steadily rotating pulley (neglecting vibratory effects) or (2) vibration of belts neglecting frictional contact.

Early research on frictional contact was carried out by Grashof (1883), who studied frictional mechanics of belt drives under steady operating speeds and applied torques. The belt was treated as a string and the mechanism of elastic creep of the belt along the pulley was shown to yield a single *slip arc* on the exit region of the pulley. In this classical creep theory, the transition from low to high tension (or vice-versa) occurs in this slip arc; see also (Johnson, 1985). Later investigators, including Firbank (1970) and Gerbert (1991, 1996), considered the influence of belt thickness as well as its length, and proposed revised estimates of the extent of the slip zone based on belt creep, shear, radial deformation, and seating/unseating behavior. These new analyses predicted tension transitions in the *adhesion* arc as well as in the slip arc. Finally, Townsend and Salisbury (1988) used a control volume approach to calculate the energy efficiency of a two-pulley belt drive based solely on the stiffness per unit length of the belt and the transmitted torque. Again, the operating conditions of the drive were assumed to be steady.

More recently, substantial research has concentrated on serpentine belt drive systems which include an automatic tensioning element, as commonly employed in vehicle front end accessory drives. These investigations, including Barker et al. (1991), Hwang et al. (1994), Beikmann et al. (1996), Kraver et al. (1996), and Leamy and Perkins (1997), have focused on the dynamic response of the belt and the accessory pulleys due to crankshaft excitation. Predictions of transverse and/or longitudinal belt vibration and pulley rotational response employed idealized models of the belt/pulley contact. For instance, in Barker et al. (1991), Hwang et al. (1994), Kraver et al. (1996), and Leamy and Perkins (1997), the belt/pulley contact was modeled by a linear spring representing an equivalent length of belt in an assumed slip arc.

The vibration of the belt spans and pulleys generates tension waves in the belt which influence the belt/pulley contact in a manner yet to be understood. For instance, the amplitude and frequency of these tension fluctuations may influence the number and location of slip arcs at the interface, which in turn may affect the vibrational response. The resulting coupled problem of vibrational response and frictional contact has yet to be investigated.

In this paper, we consider a simplified model of the belt and the belt/pulley interface as a starting point for understanding the coupled response. The model, illustrated in Fig. 1, consists of a length of belt resting partly on a frictional interface. A Coulomb law representation of the dry friction is chosen for its simplicity and ability to capture stick-slip motions, although other friction laws are worthy of future studies (see, for example, Tabor (1981) and Oden and Martins (1985)). In this study, we consider incoming *harmonic* tension waves that impinge upon the support and are reflected back with distorted shape. Quantifying this distortion, the reflected wave energy, and the distance the waves penetrate into the support, are objectives of this study. To this end, a numerical solution method and an approximate closed-form solution are proposed, both of which generalize easily to varied friction laws.

Similar models are also employed to evaluate the response of driven piles (E. Smith, 1960; I. Smith, 1988). Nikitin and Tyurekhodgaev (1990) analyze a semi-infinite rod, resting on a Coulomb surface and subject to various end conditions, using the method of characteristics. Exact solutions for wave propagation are obtained for cases when a constant stress or velocity is applied to the rod's end, or when the end is impacted by a rigid body. These studies, which focus on wave propagation in elastic continuum with friction, also relate to broader topics concerning friction-induced vibration (see, for example, Popp

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Jan. 6, 1998; final revision, June 16, 1998. Associate Technical Editor: L. T. Wheeler.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

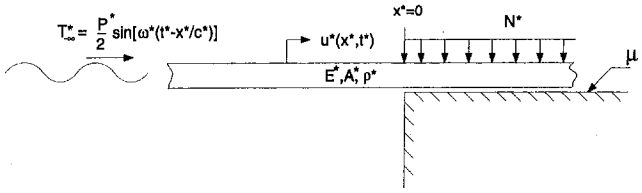


Fig. 1 System model

and Stelter (1990) and the review provided in Ibrahim (1994a, b)).

2 System Model

The displacement $u^*(x^*, t^*)$ of the infinite rod shown in Fig. 1 is governed by the *nonlinear* wave equation,

$$\frac{\partial^2 u^*}{\partial x^{*2}} - \frac{\rho^* A^*}{E^* A^*}\frac{\partial^2 u^*}{\partial t^{*2}} - \frac{\mu N^*}{E^* A^*}\,\mathrm{sgn}\left(\frac{\partial u^*}{\partial t^*}\right)H(x^*) = 0, \quad (1)$$

on the infinite domain $(-\infty < x^* < +\infty)$, where E^*, ρ^*, and A^* represent the Young's modulus, volume density, and cross-sectional area, respectively, of the rod. The coefficient of friction, μ, describes the Coulomb surface supporting the rod in the region $x^* \geq 0$ where N^* denotes the normal force per unit length. In (1), sgn denotes the *signum* operator while $H(x^*)$ is the Heaviside step function. The rod is subject to a steady train of harmonic tension waves of amplitude $P^*/2$ radiating to the right from $x^* = -\infty$ and defined by

$$T^*_{-\infty} = \frac{P^*}{2}\sin\left[\omega^*\left(t^* - \frac{x^*}{c^*}\right)\right] \quad (2)$$

where

$$T^*(x^*, t^*) = E^* A^* \frac{\partial u^*}{\partial x^*} \quad (3)$$

is the tension in the belt, ω^* is the excitation frequency, and $c^* = \sqrt{E^*/\rho^*}$ is the longitudinal wave speed.

Introducing the dimensionless quantities,

$$u = \frac{\mu N^* E^* A^*}{P^{*2}}u^*, \quad x = \frac{\mu N^*}{P^*}x^*, \quad t = \frac{\mu N^* c^*}{P^*}t^*, \quad (4)$$

reduces (1)–(3) to

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} - \mathrm{sgn}\left(\frac{\partial u}{\partial t}\right)H(x) = 0 \quad (5)$$

$$T_{-\infty} = \tfrac{1}{2}\sin[\Omega(t - x)] \quad (6)$$

$$T(x, t) \equiv \frac{T^*}{P^*} = \frac{\partial u}{\partial x} \quad (7)$$

where

$$\Omega = \frac{\omega^* P^*}{\mu N^* c^*}. \quad (8)$$

The dimensionless Eqs. (5)–(7) reveal that the single parameter Ω classifies all problems of the type shown in Fig. 1.

3 Numerical Solution

We first develop a numerical solution for the elastodynamic model of Fig. 1. This requires that the semi-infinite rod of Fig. 1 be replaced by a finite rod, as shown in Fig. 2. Employing a lumped parameter representation and using the dimensionless

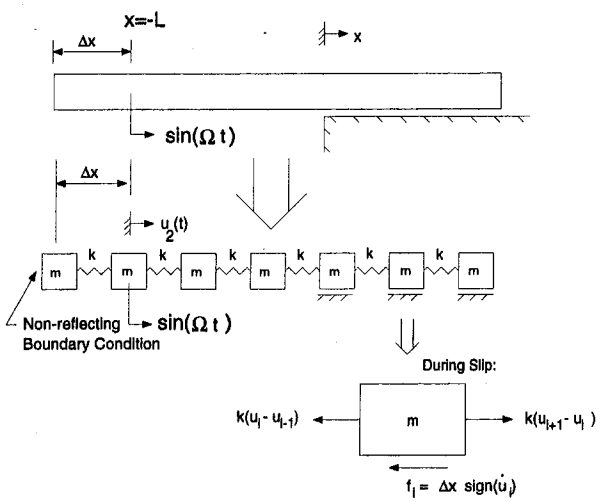


Fig. 2 Lumped parameter model

quantities in (4), the elastic rod is discretized into discrete springs and masses with element (dimensionless) stiffness and mass

$$k = \frac{1}{\Delta x}, \quad m = \Delta x,$$

respectively, where Δx is the discretization length. In Fig. 2, $u_i(t)$ denotes the dimensionless position of the ith element.

Simulation of the radiation condition at $x = -\infty$ of Fig. 1 is accomplished as follows. First, an applied load (sin (Ωt)) on the second mass yields rightward and leftward moving tension waves of amplitude $\frac{1}{2}$. The rightward wave simulates the desired radiation from $x = -\infty$. The total leftward moving tension waves, including tension wave components from the loading on the second mass *and* any reflected wave from the frictional

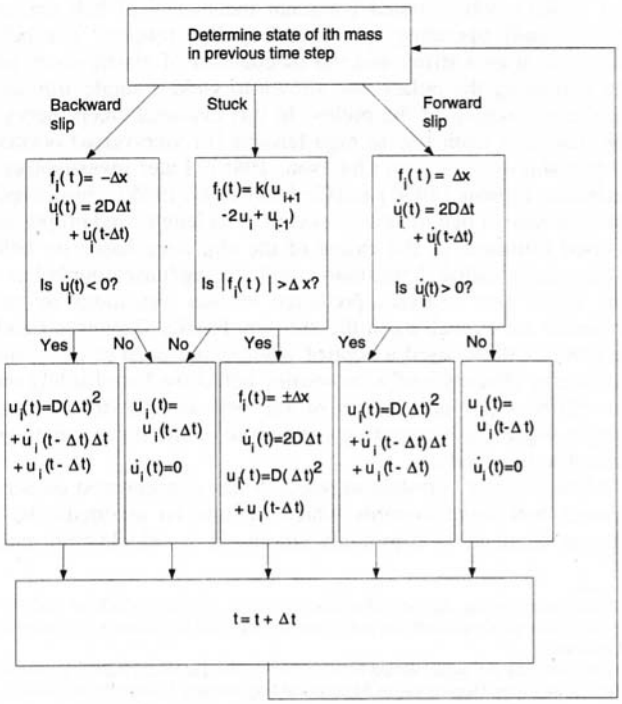


Fig. 3 Flow of calculations in numerical solution for elements in the support region, $x > 0$. The quantity $D = \{k[u_{i-1}(t - \Delta t) - 2u_i(t - \Delta t) + u_{i+1}(t - \Delta t)]\}/(2m)$.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

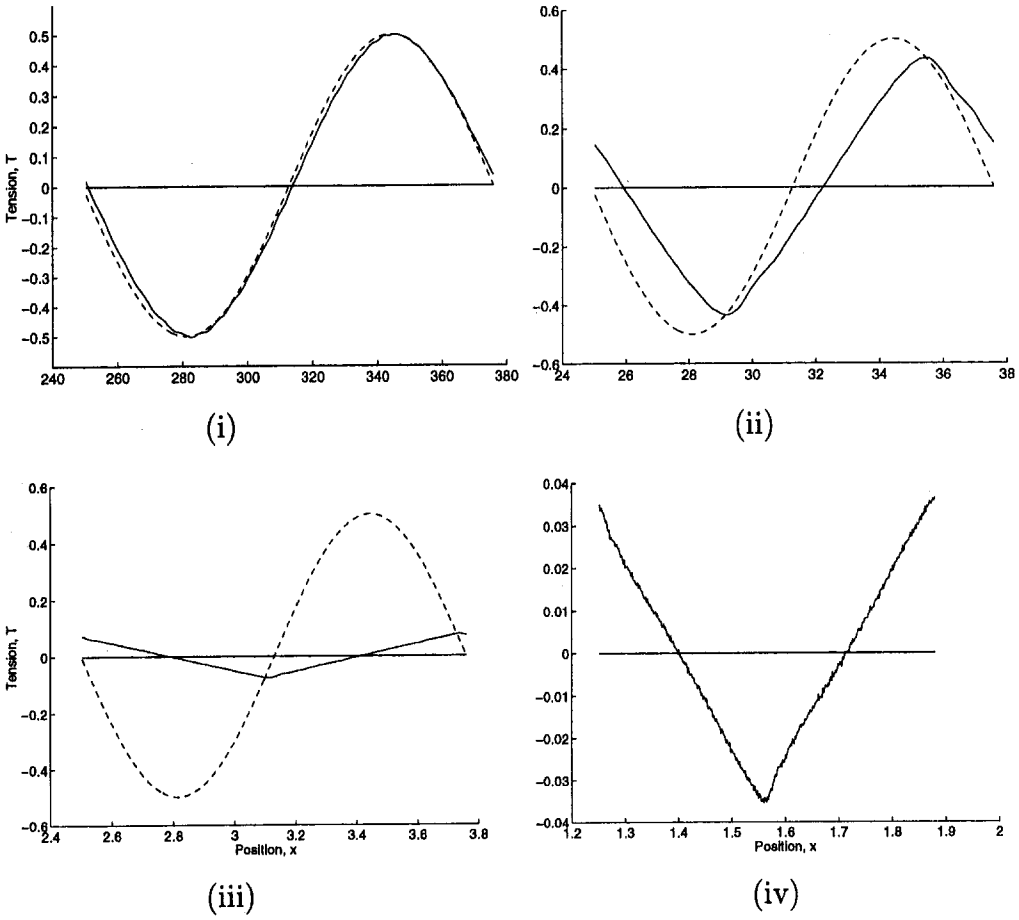


Fig. 4 Steady-state reflected tension waveforms at four values of Ω. Solid curves denote the numerical solution for a frictional support and dashed curves denote the solution for a fixed end. (i) Ω = 0.05; (ii) Ω = 0.5; (iii) Ω = 5.0; (iv) Ω = 10.0.

support, are then absorbed by a nonreflecting boundary condition (Givoli, 1991) at the site of the first element. This boundary condition, given by

$$\left.\frac{\partial u}{\partial x}\right|_{x=-L-\Delta x} = \left.\frac{\partial u}{\partial t}\right|_{x=-L-\Delta x}, \tag{9}$$

leads to a finite difference approximation,

$$\frac{u_2(t-\Delta t) - u_1(t-\Delta t)}{\Delta x} = \frac{u_1(t) - u_1(t-\Delta t)}{\Delta t}, \tag{10}$$

where the spatial derivative has been approximated in a forward $(+x)$ sense (to anticipate the leftward moving wave) and the time derivative is evaluated using u_1. Solving (10) for $u_1(t)$,

$$u_1(t) = \frac{\Delta t}{\Delta x} u_2(t-\Delta t) + \left(1 - \frac{\Delta t}{\Delta x}\right) u_1(t-\Delta t) \tag{11}$$

provides the position of the first element at each time step in the subsequent numerical simulations.

The equation of motion for every other element (i = 2, 3, 4, . . .) is used to simulate the response of the remainder of the rod. Figure 3 provides a flow diagram of the algorithm used to advance the simulation. This figure describes one time-step advancement of the ith element in the support region, $x > 0$, with a simpler algorithm for each element in the unsupported region, $x < 0$. The algorithm begins with calculating the frictional force (f_i) on the element based on the direction of slip at the end of the previous time step. If the element was stuck in the previous time step, the friction force is calculated to be that which is required to maintain equilibrium, and the element is allowed to move only if the magnitude of this force exceeds the maximum frictional force, Δx. If the element was sliding at the end of the previous step, then f_i is set equal to $\pm\Delta x$ and motion is arrested when the element's velocity changes sign. Acceleration, velocity, and position calculations follow from the equations of motion, as detailed in Fig. 3.

An appropriate time-step and spatial discretization is chosen such that the numerical solutions which follow are fully converged with 125 elements per wavelength used in the unsupported region, and at least 130 elements used over the distance that the tension waves penetrate into the support. At small Ω, the penetration distance is much smaller than a wavelength, and this length scale ultimately determines the discretization length.

3.1 Results of Numerical Simulation. The numerical solution is used to investigate what happens to a harmonic disturbance which propagates into a frictional support. In particular, attention focuses on the amplitude and shape of the tension wave which is subsequently reflected from the support.

Figure 4 shows the form of the steady-state reflected tension wave for four sample values of Ω spanning three orders of magnitude. In the first three cases, the waveform reflected from a fixed end (perfect reflection) is also shown for comparison. At small values of Ω, the waveform is nearly harmonic and nearly in phase with the wave reflected from a fixed end. As Ω increases, however, the phase is shifted as the wave penetrates further into the frictional support, and the reflected wave becomes distorted, ultimately conforming to the limiting shape of a nearly triangular

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

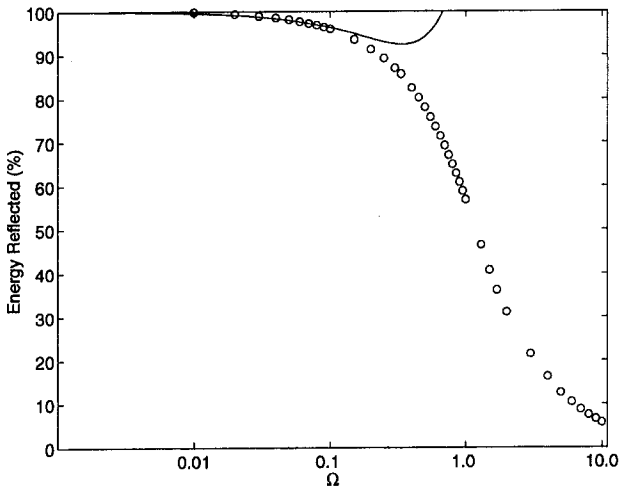


Fig. 5 Percent energy reflected versus Ω for the quasi-static solution (—) and the numerical solution (○)

wave at large Ω. As a result of the increased penetration distance (and associated energy loss), the amplitude of the reflected wave decreases monotonically with increasing Ω.

The loss of energy due to frictional dissipation is apparent in Fig. 5, which shows the energy of the reflected wave as a function of Ω. Less than five percent of the incoming energy is reflected for values of Ω exceeding 10.

The gradual distortion of the harmonic wave with Ω is further evaluated by examining the Fourier content of the reflected wave form. Figure 6 illustrates the magnitude of the first three nonzero Fourier coefficients (first, third, and fifth harmonics) of the reflected wave form, as well as the phase of the first harmonic, as functions of Ω. This phase defines a lag between the first harmonic of the reflected wave and the reference wave reflected from a fixed end. The Fourier coefficients associated with the numerical solution are obtained by numerical quadrature. For small values of Ω in Fig. 6, the wave reflected from the frictional support deviates little in either amplitude or phase from the pure harmonic reflected from the fixed end. As Ω increases, the amplitude of the reflected wave decreases, and the relative contributions of the higher harmonics increase. As the value of Ω increases still further, the higher harmonics achieve a maximum magnitude (near $\Omega = 1$) and then decay with increasing Ω. The phase lag of the first harmonic approaches the limiting value of $\pi/2$. As Ω approaches infinity, all coefficients approach zero and the frictional support acts as an anechoic termination.

Consider next the steady-state tension developed in the support region. Figure 7 illustrates the envelope of the tension wave that penetrates into the support region. The envelope is generated by noting the maximum and minimum tension of each element as a train of tension waves impinge upon the support region. In particular, an average envelope is calculated from the envelopes produced by initially compressive loading ($+\sin \Omega t$) and initially tensile loading ($-\sin \Omega t$). For small Ω, the support creates near perfect reflections and the magnitude of the tension at the entrance to the frictional support is nearly twice the amplitude of the incoming wave, as shown in Fig. 7. As Ω increases, the penetration distance (dimensionless) stretches slightly and the magnitude of the tension envelope at the entrance approaches $\pm\frac{1}{2}$ as the waves pass through with little reflection.

4 Limiting Cases

Two limiting regimes are identified based on limiting values of Ω: (1) a regime for small Ω wherein the rod responds quasi-statically and (2) a regime for large Ω wherein the frictional support approaches an anechoic termination. These limits are discussed next.

4.1 Quasi-Static Regime (Small Ω). Consider again the rod of Fig. 2 subject to the applied load $\sin \Omega t$ near the limit $\Omega \to 0$. This limit may realized as the excitation frequency $\omega^* \to 0$ or as the wave speed $c^* \to \infty$, or as the excitation wavelength $\lambda^* = 2\pi c^*/\omega^*$ approaches infinity. In all cases the rod responds in a quasi-static manner and the inertia term in (5) is negligibly small. The quasi-static tension distribution may then be evaluated from a free-body diagram of the rod. The displacement $u(x, t)$ follows upon integrating the constitutive law $(\partial u/\partial x) = T$.

Figure 8 illustrates the resulting tension distribution for the half cycle during which the load decreases. The shaded trough develops from the fully loaded tension distribution (denoted by the dashed line) as the tension applied to the end of the rod decreases. As this load decreases, a slip zone of opposite slip direction (as compared to the fully loaded slip direction) begins at $x = 0$ and grows towards increasing x, eventually occupying the entire penetration distance when $\sin(\Omega t) = -1$. Note that since the displacement at $x = 1$ is zero, the displacement at $x = -L$ is simply proportional to the shaded area in Fig. 8, namely

$$u(-L, t) = [L + \tfrac{1}{2}] \sin(\Omega t) - \tfrac{1}{4} \sin^2(\Omega t) + \tfrac{1}{4},$$

$$\frac{(4m + 1)\pi}{2} < \Omega t < \frac{(4m + 3)\pi}{2}, \quad (12)$$

where m is an integer. By similar arguments, during the half cycle when the load is increasing,

$$u(-L, t) = [L + \tfrac{1}{2}] \sin(\Omega t) + \tfrac{1}{4} \sin^2(\Omega t) - \tfrac{1}{4},$$

$$\frac{(4m + 3)\pi}{2} < \Omega t < \frac{(4m + 5)\pi}{2}. \quad (13)$$

The quasi-static stretching assumption provides a solution for the displacement and tension which are spatially uniform in the unsupported region, and parameterized by time. The wave-like behavior is lost, but the time parameterization can still be exploited to yield an approximate waveform by solving an associated *elastodynamic* problem for the subdomain to the left of the loading.

To this end, consider a semi-infinite bar $-\infty < x < -L$ whose end $x = -L$ is subjected to the displacement $u(-L, t)$ obtained above. The resulting waveform of this elastodynamic problem can then be compared to that obtained from the numerical simulation. This approximate elastodynamic solution follows directly from (12) and (13) upon replacing Ωt with $\phi = \Omega(x + L + t)$ and is valid for $x < -L$:

$$u(x, t) = \begin{cases} [L + \tfrac{1}{2}] \sin\phi - \tfrac{1}{4} \sin^2\phi + \tfrac{1}{4}, \\ \dfrac{(4m + 1)\pi}{2} < \phi < \dfrac{(4m + 3)\pi}{2} \\ [L + \tfrac{1}{2}] \sin\phi + \tfrac{1}{4} \sin^2\phi - \tfrac{1}{4}, \\ \dfrac{(4m + 3)\pi}{2} < \phi < \dfrac{(4m + 5)\pi}{2}. \end{cases} \quad (14)$$

The associated tension distribution is

$$T(x, t) = \begin{cases} \Omega[L + \tfrac{1}{2}] \cos\phi - \tfrac{1}{2}\Omega \sin\phi \cos\phi, \\ \dfrac{(4m + 1)\pi}{2} < \phi < \dfrac{(4m + 3)\pi}{2} \\ \Omega[L + \tfrac{1}{2}] \cos\phi + \tfrac{1}{2}\Omega \sin\phi \cos\phi, \\ \dfrac{(4m + 3)\pi}{2} < \phi < \dfrac{(4m + 5)\pi}{2}. \end{cases} \quad (15)$$

An approximate solution for the rod subjected to harmonic

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

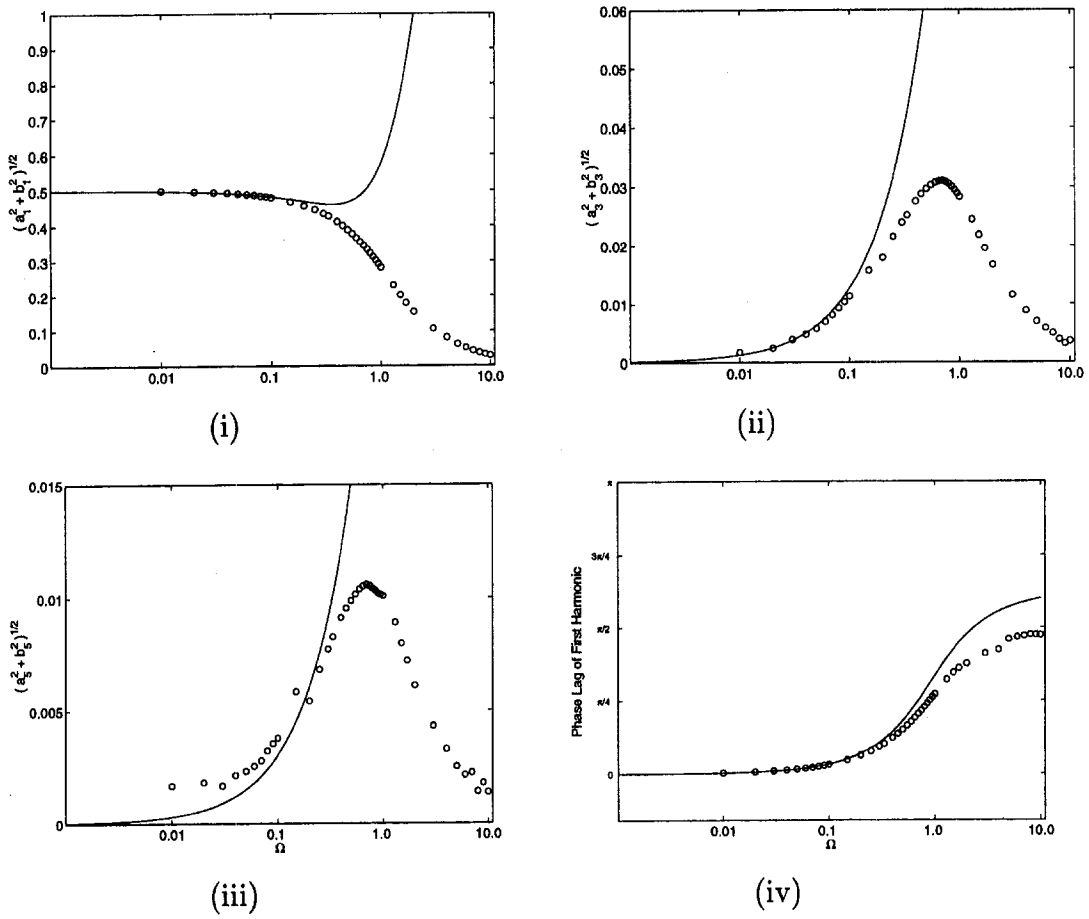


Fig. 6 Fourier coefficients versus Ω. The magnitudes of the first, third, and fifth harmonic are shown in (i)–(iii), respectively, and the phase of the first harmonic is shown in (iv). Numerical solution denoted by ○; quasi-static solution denoted by —.

radiation from $x = -\infty$, Fig. 1, can be recovered from (15) by adding the incoming rightward moving tension wave

$$\tfrac{1}{2}\sin\left[\Omega(x + L - t)\right]$$

in $x < -L$ and cancelling the leftward moving tension wave

$$\tfrac{1}{2}\sin\left[\Omega(x + L + t)\right]$$

in $x < -L$. Additionally, the choice of L is arbitrary and selecting $L = 0$, simplifies the expression further. The resulting solution, valid for $x < 0$, is

$$T(x, t) = \begin{cases} T^{-} = \frac{1}{2}\Omega\cos\phi - \frac{1}{4}\Omega\sin(2\phi) - \frac{1}{2}\sin\phi \\ \quad + \frac{1}{2}\sin\theta, \quad \dfrac{(4m+1)\pi}{2} < \phi < \dfrac{(4m+3)\pi}{2} \\ T^{+} = \frac{1}{2}\Omega\cos\phi + \frac{1}{4}\Omega\sin(2\phi) - \frac{1}{2}\sin\phi \\ \quad + \frac{1}{2}\sin\theta, \quad \dfrac{(4m+3)\pi}{2} < \phi < \dfrac{(4m+5)\pi}{2} \end{cases} \tag{16}$$

with now

$$\phi = \Omega(x + t)$$

$$\theta = \Omega(x - t)$$

where the region $x < 0$ now lies in front of the source. Having a solution valid in front of the source makes comparisons to the numerical results more convenient, and is the motivation for this step. The superscripts + and − distinguish the half cycles in which the loading is increasing + (decreasing −).

The accuracy of the quasi-static solution is now assessed. To this end, consider a Fourier series decomposition of (16) as the means to compare the waveform based on quasi-statics to that

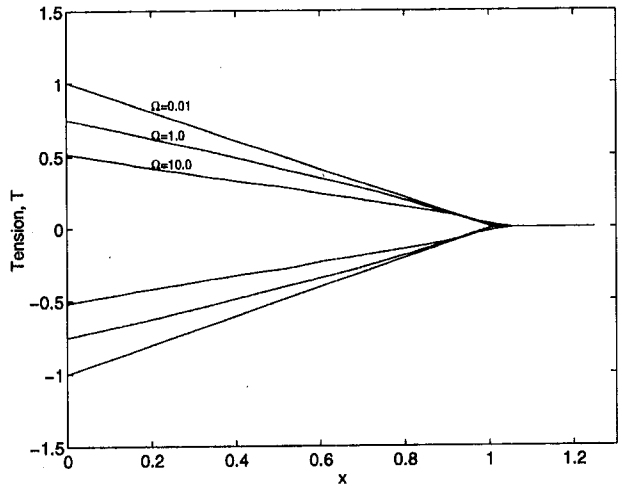


Fig. 7 Envelope of tension distribution in support region. Plotted are envelopes of dimensionless tension (T) versus dimensionless position (x) for Ω = 0.01, 1.0, and 10.0. The amplitude of the envelope monotonically decreases with increasing Ω.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

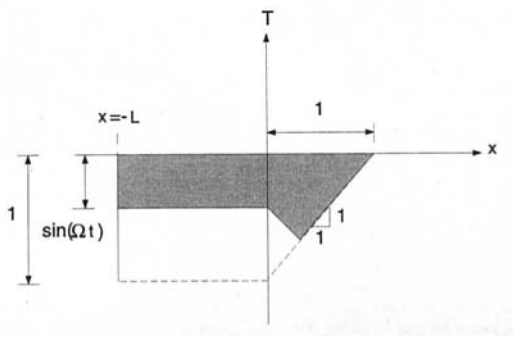


Fig. 8 Quasi-static tension distribution (decreasing excitation force). Dashed line represents the tension distribution for the fully loaded rod.

obtained from the numerical solution. Defining a Fourier series for the tension field,

$$T(x, t) = a_0(t) + \sum_{n=1}^{\infty} [a_n(t) \cos(n\Omega x) + b_n(t) \sin(n\Omega x)], \quad (17)$$

the Fourier coefficients ($a_n(t)$, $b_n(t)$) are evaluated over a single wavelength ($m = 0$ in 16)[1]:

$$a_0(t) = \frac{\Omega}{2\pi} \int_{(\pi/2\Omega)-t}^{(3\pi/2\Omega)-t} T^-(x, t)dx + \frac{\Omega}{2\pi} \int_{(3\pi/2\Omega)-t}^{(5\pi/2\Omega)-t} T^+(x, t)dx \quad (18)$$

$$a_n(t) = \frac{\Omega}{\pi} \int_{(\pi/2\Omega)-t}^{(3\pi/2\Omega)-t} T^-(x, t) \cos(n\Omega x)dx + \frac{\Omega}{\pi} \int_{(3\pi/2\Omega)-t}^{(5\pi/2\Omega)-t} T^+(x, t) \cos(n\Omega x)dx \quad (19)$$

$$b_n(t) = \frac{\Omega}{\pi} \int_{(\pi/2\Omega)-t}^{(3\pi/2\Omega)-t} T^-(x, t) \sin(n\Omega x)dx + \frac{\Omega}{\pi} \int_{(3\pi/2\Omega)-t}^{(5\pi/2\Omega)-t} T^+(x, t) \sin(n\Omega x)dx. \quad (20)$$

Evaluation of (18)–(20) reveals that all even integer Fourier coefficients ($a_0, a_2, \ldots, b_2, b_4, \ldots$) vanish, consistent with the fact that the Coulomb law describes an odd distribution (Pierre et al., 1985). The odd-numbered coefficients are

$$a_1 = \left[\frac{2\Omega}{3\pi} - 1\right] \sin(\Omega t) + \frac{\Omega}{2} \cos(\Omega t) \quad (21)$$

$$b_1 = \frac{2\Omega}{3\pi} \cos(\Omega t) - \frac{\Omega}{2} \sin(\Omega t) \quad (22)$$

$$a_n = (-1)^{(n+1)/2} \frac{2}{n^2 - 4} \frac{\Omega}{\pi} \sin(n\Omega t), \; n = 3, 5, \ldots \quad (23)$$

$$b_n = (-1)^{(n+1)/2} \frac{2}{n^2 - 4} \frac{\Omega}{\pi} \cos(n\Omega t), \; n = 3, 5, \ldots. \quad (24)$$

Note that the coefficients for a given order (n) depend solely on the dimensionless parameter Ω, as expected.

Finally, note that the *reflected* component of the tension waveform may be extracted from the Fourier series (17) by subtracting the coefficients due to the incoming tension wave. The resulting Fourier series of the reflected wave component is again (17) with

$$a_1 = \left[\frac{2\Omega}{3\pi} - \frac{1}{2}\right] \sin(\Omega t) + \frac{\Omega}{2} \cos(\Omega t) \quad (25)$$

$$b_1 = \left[\frac{2\Omega}{3\pi} - \frac{1}{2}\right] \cos(\Omega t) - \frac{\Omega}{2} \sin(\Omega t) \quad (26)$$

and a_n and b_n ($n = 3, 5, 7, \ldots$) again given by (23) and (24).

4.1.1 Results of Quasi-Static Analysis. The Fourier coefficients of the reflected tension wave for the quasi-static solution are superimposed on those for the full elastodynamic solution in Fig. 6. The coefficients for the quasi-static solution converge to those of the numerical solution as $\Omega \to 0$. The convergence of the coefficients for the higher ($n > 5$) harmonics is ultimately limited by the discretization size. A trend captured by both the numerical and quasi-static solutions at small Ω is the increasing importance of the super-harmonics as Ω increases from zero, and the decreasing importance of the fundamental harmonic. This trend continues in the numerical solution until $\Omega \cong 1.0$, beyond which all harmonics decay. This decay is not captured by the quasi-static solution which ultimately predicts for larger values of Ω that the reflected wave possesses a greater amplitude (and thus energy) than the originating wave (a physical impossibility) (refer also to Fig. 5).

Observe from Fig. 6 that a regime for Ω exists in which the quasi-static solution accurately predicts the reflected waveform and that this regime can be identified by examining the fundamental harmonic alone. The higher harmonics have a negligible affect on the magnitude and phase of the reflected wave. Values for $\Omega < \frac{1}{3}$ ensure that the quasi-static solution predicts the magnitude of the reflected wave to within eight percent and the phase to within two percent, when compared to the elastodynamic solution. Figure 5 also indicates the fidelity of the quasi-static solution. For $\Omega < \frac{1}{3}$, both solutions predict near identical reflected wave energy.

A final comparison of the numerical and quasi-static solution is provided in Figs. 7 and 8. From Fig. 8, the quasi-static solution predicts a triangular tension envelope in the support region (see the dotted line shown in the Fig.). From Fig. 7 it is observed that the tension envelope from the numerical solution approaches that of the quasi-static solution for small Ω, with no observable difference between envelopes at $\Omega = 0.01$. Further comparisons reveal an interesting observation; namely, the dimensionless penetration distance (S) of the quasi-static solution predicts the penetration with errors less than five percent over the large range of Ω considered in Fig. 7.

4.2 Anechoic Termination Regime (Large Ω). From Figs. 4–6, as Ω increases, the amplitude of the reflected wave decreases, and the reflected wave energy decreases as the waveform converges to a nearly triangular profile. At sufficiently high Ω, such as $\Omega \geq 10$, nearly all the energy (~95 percent) of the incoming waves is absorbed by the support and an anechoic termination is effectively realized. Thus, in systems characterized by a large Ω, the frictional support may be adequately modeled by an anechoic boundary condition in the form of (9).

5 Conclusions

A single dimensionless parameter Ω has been identified which governs the dynamic response of an infinite rod subject to harmonic tension waves and interacting with a dry friction surface. Incoming harmonic tension waves partially reflect from the frictional support and both the degree of reflection and the degree of distortion are controlled solely by Ω. This distortion and energy reflection have been quantified as a function of Ω using an approximate closed-form quasi-static solution and a

[1] Note that the coefficients are calculated with integration limits that move with the moving tension regions. Alternatively, the coefficients can be calculated over a *fixed* spatial region corresponding to one excitation wavelength. These approaches yield the same expressions for the coefficients as required.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

numerical solution. The results of this study suggest that a quasi-static model for belt/pulley interaction will remain accurate for belt drive systems for which $\Omega < \frac{1}{3}$. The distance harmonic tension waves penetrate into the frictional support is accurately predicted by the quasi-static solution, even for large Ω. For $\Omega > 10$, the frictional support acts effectively as an anechoic termination.

Acknowledgment

The authors gratefully acknowledge the Ford Motor Company for supporting this research endeavor.

References

Barker, C. R., Oliver, L. R., and Brieg, W. F., 1991, "Dynamic Analysis of Belt Drive Tension Forces During Rapid Engine Acceleration," SAE Congress, Detroit, MI, Paper No. 910687, pp. 239–254.

Beikmann, R. S., Perkins, N. C., and Ulsoy, A. G., 1996, "Free Vibration of Serpentine Belt Drive Systems," ASME *Journal of Vibration and Acoustics,* Vol. 118, pp. 406–413.

Euler, M. Leonard, 1762, "Remarques sur l'effect du frottement dans l'equilibre," *Mém. Acad. Sci.,* pp. 265–278.

Fawcett, J. N., 1981, "Chain and Belt Drives—A Review," *Shock & Vibration Digest,* Vol. 13, No. 5, pp. 5–12.

Firbank, T. C., 1970, "Mechanics of the Belt Drive," *International Journal of Mechanical Sciences,* Vol. 12, pp. 1053–1063.

Gerbert, G. G., 1991, "On Flat Belt Slip," *Vehicle Tribology,* Tribology Series 16, Elsevier, Amsterdam, pp. 333–339.

Gerbert, G. G., 1996, "Belt Slip—A Unified Approach," ASME *Journal of Mechanical Design,* Vol. 118, pp. 432–438.

Givoli, D., 1991, "Review Article: Non-reflecting Boundary Conditions," *Journal of Computational Physics,* Vol. 94, pp. 1–29.

Grashof, B. G., 1883, *Theoretische Maschinenlehre,* Bd 2, Leopold Voss, Hamburg.

Hwang, S.-J., Perkins, N. C., Ulsoy, A. G., and Meckstroth, R. J., 1994, "Rotational Response and Slip Prediction of Serpentine Belt Drive Systems," ASME *Journal of Vibration and Acoustics,* Vol. 116, pp. 71–78.

Ibrahim, R. A., 1994a, "Friction-Induced Vibration, Chatter, Squeal, and Chaos: Part I—Mechanics of Contact and Friction," ASME *Applied Mechanics Reviews,* Vol. 47, pp. 209–226.

Ibrahim, R. A., 1994b, "Friction-Induced Vibration, Chatter, Squeal, and Chaos: Part II—Dynamics and Modeling," ASME *Applied Mechanics Reviews,* Vol. 47, pp. 227–253.

Johnson, K. L., 1985, *Contact Mechanics,* Cambridge University Press, London, Chapter 8.

Kraver, T. C., Fan, G. W., and Shah, J. J., 1996, "Complex Modal Analysis of a Flat Belt Pulley System With Belt Damping and Coulomb-Damped Tensioner," ASME *Journal of Mechanical Design,* Vol. 118, pp. 306–311.

Leamy, M. J., and Perkins, N. C., 1998, "Nonlinear Periodic Response of Engine Accessory Drives With Dry Friction Tensioners," ASME *Journal of Vibration and Acoustics,* Vol. 120, pp. 909–916.

Nikitin, L. V., and Tyurekhodgaev, A. N., 1990, "Wave Propagation and Vibration of Elastic Rods with Interfacial Frictional Slip," *Wave Motion,* Vol. 12, pp. 513–526.

Oden, J. T., and Martins, J. A. C., 1985, "Models and Computational Methods For Dynamic Friction Phenomena," *Computer Methods in Applied Mechanics and Engineering,* Vol. 52, pp. 527–634.

Pierre, C., Ferri, A. A., and Dowell, E. H., 1985, "Multi-Harmonic Analysis of Dry Friction Damped Systems Using an Incremental Harmonic Balance Method," ASME JOURNAL OF APPLIED MECHANICS, Vol. 52, pp. 958–964.

Popp, K., and Stelter, P., 1990, "Stick-Slip Vibrations and Chaos," *Philosophical Transactions of the Royal Society of London,* Vol. A332, pp. 89–105.

Smith, E. A. L., 1960, "Pile-Driving Analysis by the Wave Equation," *ASCE Journal of the Soil Mechanics and Foundations Division,* Vol. 86, No. EM 4, Aug., pp. 35–61.

Smith, I. M., and To, P., 1988, "Numerical Studies of Vibratory Pile Driving," *International Journal for Numerical and Analytical Methods in Geomechanics,* Vol. 12, pp. 513–531.

Tabor, D., 1981, "Friction-The Present State of Our Understanding," ASME *Journal of Lubrication Technology,* Vol. 103, pp. 169–179.

Townsend, W. T., and Salisbury, J. K., 1988, "The Efficiency Limit of Belt and Cable Drives," ASME *Journal of Mechanisms, Transmissions, and Automation in Design,* Vol. 110, pp. 303–307.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

V. A. Lubarda

X. Markenscoff
Fellow ASME

Institute for Mechanics and Materials,
Department of Applied Mechanics and
Engineering Science,
University of California, San Diego,
La Jolla, CA 92093-0404

On the Stress Field in Sliding Ellipsoidal Inclusions With Shear Eigenstrain

A stress distribution in elliptical inclusions with sliding interface and uniform shear eigenstrain is analyzed. The existence of a nonvanishing stress in nearly circular inclusions is demonstrated, and an approximate method for its determination is suggested. An inherent nonlinear dependence of the stress on applied eigenstrain is indicated and discussed.

1 Introduction

There is an unusual dependence of stress on the semi-axes ratio and applied shear eigenstrain in ellipsoidal or elliptical inclusions with sliding interface. Mura and Furuhashi (1984) found that, under certain conditions, the stresses vanish everywhere in the sliding inclusion and the matrix. The result does not apply to spherical or circular inclusions, for which there is a nonvanishing stress in both the inclusion and the matrix. Since the solution for a spherical shape does not follow as a special case from available solution for an ellipsoidal shape, the situation was referred to as an anomaly of sliding inclusions (Kouris, Tsuchida, and Mura, 1986). In this paper, we explain this unusual situation and show by geometric considerations that, for sufficiently large levels of applied shear eigenstrain, there must be a nonvanishing stress field in nearly circular or spherical inclusions. In particular, we introduce for elliptical inclusions the parameter $s = (a - b)/b\gamma$, where a and b are the principal semi-axes of the ellipse, and γ is the applied shear eigenstrain, and show that the stress in a sliding inclusion is proportional to γ if $s \ll 1$, and proportional to γ^2 (thus, not observable at the level of linear elasticity) if $s \gg 1$. The latter occurs at all levels of infinitesimal strain γ for elliptical inclusions whose semi-axes differ distinctly, and for nearly circular inclusions at very small values of γ. If the parameter s is of the order of one, the stress depends on γ in a more involved, nonlinear manner, rapidly decreasing with the ratio a/b for any given value of γ. The inherent nonlinear dependence of the stress on applied shear eigenstrain is a peculiar feature of ellipsoidal sliding inclusions, because neither large overall strain nor large rotation actually take place in the considered problem. An analogous situation occurs in the case of sliding elliptical or ellipsoidal inclusions under remote uniform shear loading.

2 Geometric Considerations

Consider a two-dimensional situation in which an infinite block of elastic material is subjected to uniform infinitesimal shear strain ϵ_{xy}^{*}, denoted in sequel by γ, so that the displacement field in rectangular coordinates is given by $u = \gamma y$, $v = \gamma x$. A point with the coordinates (x, y) on the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{1}$$

moves to the point (X, Y) with the coordinates $X = x + \gamma y$ and $Y = y + \gamma x$, so that (Fig. 1)

$$x = \frac{1}{1 - \gamma^2}(X - \gamma Y), \quad y = \frac{1}{1 - \gamma^2}(Y - \gamma X). \tag{2}$$

Substitution into Eq. (1) thus gives

$$\frac{(X - \gamma Y)^2}{a^2} + \frac{(Y - \gamma X)^2}{b^2} = (1 - \gamma^2)^2. \tag{3}$$

This is an equation of the deformed ellipse, which can be rewritten more conveniently by using the coordinates (ξ, η), parallel to the new principal directions of the ellipse. Thus, denoting by φ the angle between the old and new major axes of the ellipse, we have

$$X = \xi \cos\varphi - \eta \sin\varphi, \quad Y = \xi \sin\varphi + \eta \cos\varphi. \tag{4}$$

Equation (3) consequently becomes

$$\frac{\xi^2}{A^2} + \frac{\eta^2}{B^2} - C\xi\eta = 1, \tag{5}$$

where A and B are defined by

$$A = \frac{(1 - \gamma^2)a}{[(\cos\varphi - \gamma\sin\varphi)^2 + (a^2/b^2)(\sin\varphi - \gamma\cos\varphi)^2]^{1/2}}, \tag{6}$$

$$B = \frac{(1 - \gamma^2)b}{[(\cos\varphi + \gamma\sin\varphi)^2 + (b^2/a^2)(\sin\varphi + \gamma\cos\varphi)^2]^{1/2}}, \tag{7}$$

and

$$C = \frac{1}{1 - \gamma^2}\left[\frac{1}{a^2}\left(\sin 2\varphi + \frac{2\gamma}{1 - \gamma^2}\cos 2\varphi\right) - \frac{1}{b^2}\left(\sin 2\varphi - \frac{2\gamma}{1 - \gamma^2}\cos 2\varphi\right)\right]. \tag{8}$$

The lengths A and B are the lengths of the principal semi-axes of the deformed ellipse, provided that the angle φ is such that $C = 0$, which gives

$$\tan 2\varphi = \frac{a^2 + b^2}{a^2 - b^2}\frac{2\gamma}{1 - \gamma^2}. \tag{9}$$

The variation of A/a versus a/b for two selected values of γ is

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Jan. 27, 1997; final revision, Apr. 30, 1998. Associate Technical Editor: W. J. Drugan.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

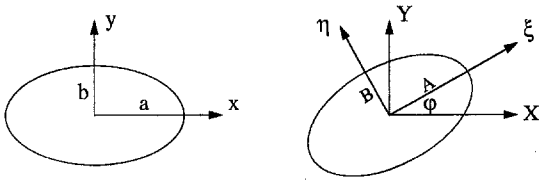


Fig. 1 An ellipse with the principal semi-axes *a* and *b* is deformed under uniform shear into an ellipse with the principal semi-axes *A* and *B*. The new principal axes of the ellipse are at an angle φ relative to the old axes.

shown in Fig. 2. As the ratio a/b increases from the value of 1, A rapidly approaches a, more so the smaller the value of the applied shear strain. Figure 3 shows plots of A versus γ for two values of the ratio a/b. Except for the circular shape ($a = b$), for which A is a linear function of γ, for all elliptical shapes there is at first a range of γ for which A is essentially equal to a, beyond which A begins to notably increase with γ. This range is more extended the larger the ratio a/b, and for $a/b = 1.02$ it already extends throughout the whole elastic strain range of approximately 10^{-3}. Similar remarks apply to the semi-axis B.

To determine which material direction (semi-axis) of the undeformed ellipse becomes a principal semi-axis of the deformed ellipse, consider an arbitrary point of the undeformed ellipse, with the coordinates $x = a \cos\theta$ and $y = b \sin\theta$, where θ is a parameter of the ellipse. The corresponding point on the deformed ellipse is $X = a \cos\theta + \gamma b \sin\theta$ and $Y = b \sin\theta + \gamma a \cos\theta$. If (X, Y) is to be an apex of the deformed ellipse, the extremum condition $d(X^2 + Y^2)/d\theta = 0$ applies, which defines the angle θ to be

$$\tan 2\theta = \frac{2ab}{a^2 - b^2}\frac{2\gamma}{1 + \gamma^2}. \tag{10}$$

The angle ψ, defined as $\tan\psi = (b/a)\tan\theta$, specifies the direction of a semi-axis in the undeformed configuration which becomes the principal semi-axis in the deformed configuration. The angle defining the actual rotation of material lines is $\omega = \varphi - \psi$. As a/b increases, ω rapidly approaches the value of the applied shear strain γ. Thus, although the angle φ may increase up to $\pi/4$, the angle ω remains infinitesimal.

2.1 Limiting Behavior. If the right-hand side of Eq. (9) is sufficiently small for the approximation $\tan 2\varphi \approx 2\varphi$ to accurately apply, we have

$$\varphi = \sin\varphi = \frac{a^2 + b^2}{a^2 - b^2}\gamma, \quad \cos\varphi = 1 - \frac{1}{2}\left(\frac{a^2 + b^2}{a^2 - b^2}\right)^2\gamma^2, \tag{11}$$

neglecting the cubic and higher-order terms in γ. Upon an appropriate expansion, assuming a sufficiently small ratio $b\gamma/(a - b)$, Eqs. (6) and (7) yield

$$A = \left(1 + \frac{1}{2}\frac{a^2 + 3b^2}{a^2 - b^2}\gamma^2\right)a, \tag{12}$$

$$B = \left(1 - \frac{1}{2}\frac{3a^2 + b^2}{a^2 - b^2}\gamma^2\right)b. \tag{13}$$

Thus, the undeformed and deformed principal semi-axes differ by terms proportional to γ^2. This difference cannot be observed at the level of linear elasticity, and the two ellipses are identical at that level ($A = a$, $B = b$).

In the other limiting case, $(a - b)/b\gamma$ is sufficiently small and the approximation $\cot 2\varphi = (a - b)/2b\gamma$ holds, i.e.,

$$\sin\varphi = \frac{\sqrt{2}}{2}\left(1 - \frac{a - b}{4b\gamma}\right), \quad \cos\varphi = \frac{\sqrt{2}}{2}\left(1 + \frac{a - b}{4b\gamma}\right). \tag{14}$$

The subsequent expansions in Eqs. (6) and (7) yield

$$A = (1 + \gamma)a - \frac{a - b}{2}, \tag{15}$$

$$B = (1 - \gamma)b + \frac{a - b}{2}. \tag{16}$$

Therefore, for a given nonzero γ, if $(a - b) \ll b\gamma$ the deformed ellipse differs in size distinctly from the undeformed ellipse.

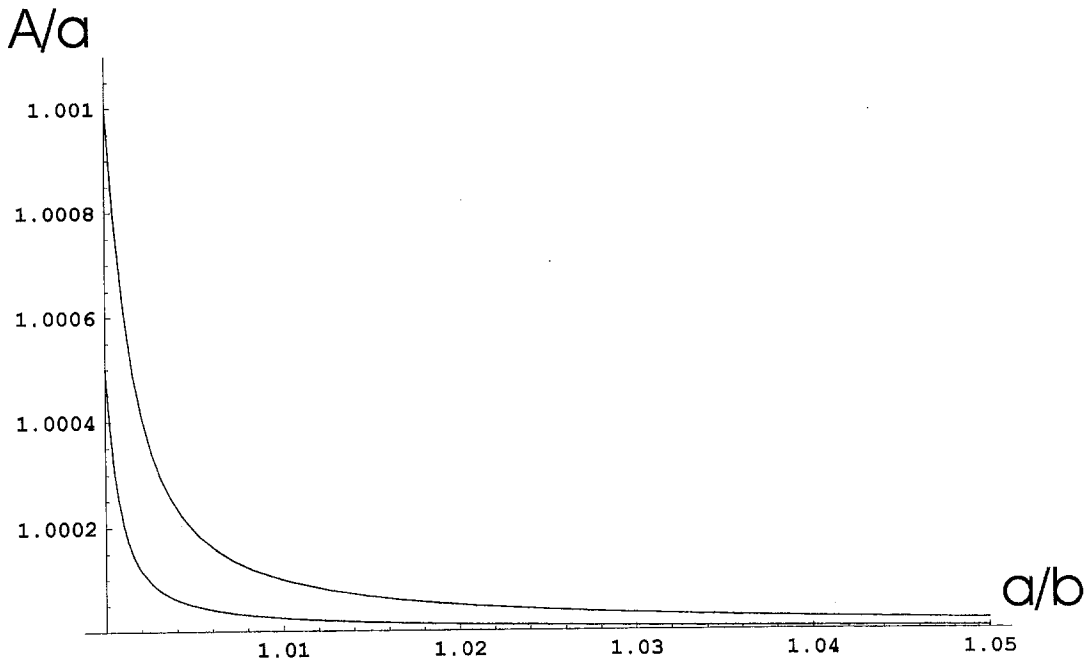


Fig. 2 The ratio *A*/*a* of the principal semi-axes of the ellipse in the deformed and undeformed configuration rapidly decreases with the ratio *a*/*b*. The upper curve corresponds to strain $\gamma = 10^{-3}$ and the lower to $\gamma = 5 \cdot 10^{-4}$.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

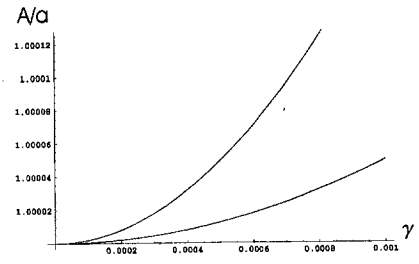


Fig. 3 The variation of A/a with the applied shear strain γ. Except for the circular shape, there is initially a range of γ for which A is essentially equal to a.

Note that for a given level of the shear strain γ, an ellipse with a small ratio $(a - b)/2b\gamma$ has the major semi-axis in the deformed configuration at an angle φ nearly equal to 45 deg, relative to the direction of the major semi-axis in the undeformed configuration. This, of course, does not imply any large rotations taking place, since the major semi-axis in the deformed configuration consists of the material points that were aligned in the undeformed configuration along the direction at an angle ψ, also nearly equal to 45 deg.

2.2 Approximate Expressions. Expressions (6) and (7) have simple but accurate approximations throughout the whole range of γ and a/b, given by

$$A = \left[1 + \frac{\gamma}{1 + k(a - b)/b\gamma}\right]a, \tag{17}$$

$$B = \left[1 - \frac{\gamma}{1 + k(a - b)/b\gamma}\right]b, \tag{18}$$

where $k = \frac{3}{4}$. Evidently, if $(a - b)/b\gamma$ is large, A and B become, within terms proportional to γ^2, equal to a and b. For small values of $(a - b)/b\gamma$, $A/a = 1 + \gamma - k(a - b)/b$, and $B/b = 1 - \gamma + k(a - b)/b$. Numerical evaluations confirm close agreement with exact results. Equations (17) and (18) will be conveniently utilized later in the text.

3 Sliding Inclusion

If the uniform eigenstrain in the elliptical inclusion is pure shear ϵ^*_{xy}, the normal and shear component of the traction at the bonded interface between the inclusion and the surrounding matrix are obtained from the Eshelby (1957) solution as

$$\sigma^E_n = \sigma^E_{xy}\frac{2a^2b^2xy}{b^4x^2 + a^4y^2}, \tag{19}$$

$$\tau^E_n = \sigma^E_{xy}\frac{b^4x^2 - a^4y^2}{b^4x^2 + a^4y^2}. \tag{20}$$

The shear stress $\sigma^E_{xy} = -2k\mu\epsilon^*_{xy}$, with $k = (1 - \nu)^{-1}ab/(a + b)^2$, is the Eshelby uniform stress in the inclusion, and μ and ν are the shear modulus and Poisson ratio. Mura and Furuhashi (1984) considered the sliding ellipsoidal inclusion, in which the shear stress at the interface is relaxed to zero by allowing interface sliding or slipping to occur, preserving the normal components of displacement and traction continuous across the interface. The solution can in principle be obtained by superimposing on the Eshelby solution the Volterra solution for a Somigliana dislocation, associated with an appropriate slip variation $\hat{b}_i$ along the interface, i.e.,

$$u^V_i(\mathbf{x}) = \int_S G_{ij,k}(\mathbf{x} - \mathbf{x}')C_{jklm}\hat{b}_l(\mathbf{x}')n_m(\mathbf{x}')dS. \tag{21}$$

The two-dimensional Green's function of an infinite elastic medium is denoted by G_{ij}, a comma denotes differentiation with respect to unprimed coordinates, and C_{ijkl} are the elastic moduli of isotropic material. The slip (or the variable Burgers vector along the interface S) has to be such that $\hat{b}_i n_i = 0$ for the continuity of normal displacement across the interface, and $\tau^E_n + \tau^V_n = 0$ for the vanishing of the shear traction at the interface. It is not easy to analytically determine, or to predict in advance, the structure of $\hat{b}_i$ for an arbitrary ellipsoidal shape of the inclusion, since the problem reduces to integral equations which are difficult to solve (e.g., Wong and Barnett, 1985). Mura and Furuhashi (1984), however, observed that a Somigliana dislocation based on the slip

$$\hat{b}_x = (\epsilon^*_{xy} + \omega_{xy})y, \quad \hat{b}_y = (\epsilon^*_{xy} - \omega_{xy})x, \tag{22}$$

cancels the stress field of the Eshelby inclusion, provided that

$$\omega_{xy} = \frac{a^2 + b^2}{a^2 - b^2}\epsilon^*_{xy}. \tag{23}$$

The right-hand side of Eq. (23) is precisely the angle φ in Eq. (11) of the previous section between the principal semi-axes of the undeformed and deformed ellipse under pure shear, in the case when the deformed ellipse, within the accuracy of linear elasticity, has the same size as the undeformed ellipse ($A = a, B = b$). There is no stress left in such an inclusion, because the sheared inclusion can be simply rotated by the angle ω_{xy} and inserted into the matrix without any stress (Mura, 1985). This, of course, cannot be done for the inclusion with bonded interface, where no slip is allowed and where each point on the surface of the inclusion has to be in contact with the point of the surrounding matrix to which it was attached before the eigenstrain took place.

In Section 2 it was found that an ellipse is transformed into an identical ellipse if $b\gamma/(a - b)$ is sufficiently small. For example, $b\gamma/(a - b)$ is less than $5 \cdot 10^{-2}$ if $a/b > 1 + 20\gamma$, which for the elastic shear strain of 10^{-3} gives $a/b > 1.02$. For smaller values of the ratio a/b, the principal semi-axes of the deformed ellipse are increasingly more distinct from the undeformed semi-axes, and the inclusion cannot be placed back into the matrix simply by a rotation. The stress must set up to overcome the size difference, which is, for a given eigenstrain, more pronounced the closer the elliptical shape is to the circular shape. Alternatively, in order that terms proportional to ω_{xy} in Eq. (22) represent displacements due to rotation, ω_{xy} in Eq. (23) must be infinitesimal. This is the case if the ratio $\epsilon^*_{xy}b/(a - b)$ is sufficiently small. For larger values of this ratio, significant stresses arise, and no rotation (however large) can by itself place the sliding inclusion back into the matrix.

The slip variation along the boundary of a circular sliding inclusion with uniform shear eigenstrain γ is $\hat{b} = \gamma(x^2 - y^2)/2a$, i.e.,

$$\hat{b}_x = \frac{\gamma}{2}y - \frac{\gamma}{a^2}x^2y, \quad \hat{b}_y = \frac{\gamma}{2}x - \frac{\gamma}{a^2}xy^2, \tag{24}$$

where a is the radius of the inclusion. Note that $\hat{b}$ is proportional to τ^E_n on the boundary of the inclusion. The stress distribution within the inclusion is given by quadratic polynomials in the coordinates x/a and y/a, as

$$\sigma_{xx} = \sigma_{yy} = -\frac{3\mu\gamma}{2(1 - \nu)}\frac{xy}{a^2}, \tag{25}$$

$$\sigma_{xy} = -\frac{3\mu\gamma}{4(1 - \nu)}\left(1 - \frac{x^2 + y^2}{a^2}\right). \tag{26}$$

This was obtained from the Eshelby and Volterra solutions. The average normal stress in the sliding inclusion is zero ($\bar{\sigma}_{xx} = \bar{\sigma}_{yy} = 0$), while the average shear stress is $\bar{\sigma}_{xy} = 3\sigma^E_{xy}/4$, where

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$\sigma^E_{xy} = -\mu\gamma/2(1-\nu)$ is the uniform Eshelby stress in the inclusion with bonded interface.

The fact that the stress within the inclusion is given by a polynomial of second degree was expected, since the slip $\hat{b}$ in Eq. (24) gives rise to a quadratic "eigenstrain" $\hat{\epsilon}_{ij} = (\hat{b}_{i,j} + \hat{b}_{j,i})/2$, introduced by formally extending the functions $\hat{b}_i$ from the boundary inside the inclusion. Such a procedure is often utilized in the analysis of Somigliana dislocations (Asaro, 1975).

While in the case of bonded interface, where only uniform shear stress σ^E_{xy} sets up in the inclusion, in the case of sliding inclusion normal and shear stresses arise, both being nonuniformly distributed within the inclusion. That this is so can also be recognized directly from Eqs. (19) and (20). For a circular inclusion, the shear stress τ^E_n equilibrates itself (this, incidentally, is not the case for an elliptical inclusion, for which τ^E_n gives a net moment), and the normal traction given by Eq. (19), applied alone on the boundary, clearly gives a non-uniform stress distribution within the inclusion. Of course, there is also an additional normal traction that must be supplied at the boundary of the inclusion to insure the continuity of normal displacement across the interface, upon the removal of the shear traction τ^E_n there. Indeed, it can be shown (Lubarda and Markenscoff, 1997) that the normal traction at the bonded interface increases by a factor $\frac{3}{2}$, when the shear stress at the interface is relaxed to zero, allowing for the slip to take place.

4 Approximate Stress Expressions

A nonvanishing stress in elliptical inclusions with shear eigenstrain γ can exist only in nearly circular inclusions, whose semi-axes differ by at most a few percent. Such inclusions are normally modeled as circular at the level of small deformation theory. Indeed, although a deformed configuration in infinitesimal theory can differ from the undeformed configuration by up to about two percent, the governing equilibrium equations and boundary conditions are commonly written with respect to the undeformed configuration. Consequently, the stress field in a nearly circular inclusion will be assumed to be as in an average circular inclusion of radius $a_0 = (a+b)/2$, but with an incorporated effect of the length scale $(a-b)$, to which the stress is strongly sensitive when this scale is of the order of $b\gamma$. This will be taken into account by scaling the magnitude of stress by the factor $(A-a)/a$, because it is this size factor that is responsible for the build-up of the stress in sliding elliptical inclusions, when $(a-b) \sim b\gamma$. Thus, from Eqs. (25) and (26),

$$\sigma_{xx} = \sigma_{yy} = -\frac{3\mu\gamma^*}{2(1-\nu)}\frac{xy}{a_0^2}, \tag{27}$$

$$\sigma_{xy} = -\frac{3\mu\gamma^*}{4(1-\nu)}\left(1 - \frac{x^2+y^2}{a_0^2}\right). \tag{28}$$

The apparent strain $\gamma^* = (A-a)/a$ can be calculated from the exact expression (6), or, more conveniently, from the approximate but accurate expression given by Eq. (17). The latter gives

$$\gamma^* = \frac{\gamma}{1 + k(a-b)/b\gamma}. \tag{29}$$

Note that γ^* is also equal to $(b-B)/b$. The average stress in the sliding inclusion can be obtained by integration from Eqs. (27) and (28). The result is

$$\bar{\sigma}_{xx} = \bar{\sigma}_{yy} = 0, \quad \bar{\sigma}_{xy} = -\frac{3\mu\gamma^*}{8(1-\nu)}. \tag{30}$$

For nearly circular inclusions, there is at first a range of γ for which there is no observable stress in the inclusion (the stress being proportional to γ^2). For these very small values of γ, a nearly circular inclusion actually behaves as a very elliptical inclusion, carrying no stress, in accordance with the Mura and Furuhashi result. Beyond this initial range of γ, the stress begins to substantially increase and approach the linear dependence on the eigenstrain γ, faster so smaller the ratio a/b. For the circular inclusion the nonlinear domain shrinks to a point (since $(a-b) \sim b\gamma$ implies $\gamma = 0$ when $a = b$), and the stress becomes a linear function of γ throughout the whole range of γ. Regarding the variation of the stress with the ratio a/b for a given eigenstrain γ, the stress rapidly diminishes to zero as the ratio a/b increases.

5 Discussion

The result that the stress in elliptical sliding inclusions due to shear eigenstrain γ vanishes if the angle φ is small, applies not only to elliptical inclusions with distinctly different semi-axes a and b, but also to nearly circular inclusions at sufficiently small values of γ, which make the parameter $s = (a-b)/b\gamma$ large. Therefore, the initial slope of the stress versus γ curve for nearly circular inclusions is essentially equal to zero. But, as γ increases (still staying within the infinitesimal range), s decreases and the stress eventually begins to increase with a substantial slope relative to γ. Consequently, the stress dependence on γ is inherently nonlinear. When the parameter s becomes small enough, the stress is simply proportional to γ^2, thus not observable at the level of linear elasticity. This explains why the attempts to calculate the stress in nearly circular sliding inclusions by the equations of linear elasticity (e.g., Kouris et al., 1986) were not successful. The point is subtle, because the linear equations of elasticity usually suffice in problems where no large overall strain or rotation take place. However, in the considered problem of sliding elliptical inclusions, the stress strongly depends on the length $(a-b)$, if this is of the same order as $b\gamma$. Therefore, although γ is an infinitesimal strain and $b\gamma$ is an infinitesimal length change relative to the overall geometry and the lengths a and b, the length change $b\gamma$ is large relative to the length scale $(a-b)$. From a combined kinematic-kinetic point of view, there is an unusual interplay of the contributions from the rotation and stress in fitting the sliding inclusion into the matrix. The stress is taken to be proportional to $(A-a)/a$, because the rotation cannot overcome this size barrier and insert by itself the inclusion into the matrix. For bonded inclusions, on the other hand, the stress is not proportional to $(A-a)/a$, because sliding is not available to help the accommodation of the inclusion, and the stress does exist even when $A = a$.

The unusual behavior, caused by an interplay of the rotation and stress in fitting the sliding inclusion into the matrix, is present only when the shear eigenstrain is "parallel" to the principal directions (x and y) of the ellipse (i.e., $\gamma = \epsilon^*_{xy}$). The effect is absent in the case of biaxial eigenstrain, parallel to the principal directions of the ellipse, since the ellipse deformed in this way always differs in size from the undeformed ellipse. There is, however, a particular situation in which the stress in the inclusion and the matrix can vanish. This happens for a nearly circular inclusion with the biaxial eigenstrain $\epsilon_x = b/a - 1$ and $\epsilon_y = a/b - 1$, which deforms a into b, and b into a. A rotation by 90 deg of an inclusion so strained would place it back into the matrix, without any stress. Such a configuration would correspond to the minimum (zero) total energy of the inclusion and the matrix.

Although the derivation presented in this paper was mainly devoted to elliptical inclusions in plane strain, the conclusion regarding the existence of nonvanishing stress in nearly circular inclusions can be extended to nearly spherical inclusions, as well. Thus, for a given level of applied shear eigenstrain, the stress in nearly spherical inclusions rapidly diminishes as the inclusion becomes less spherical, for the same reasons as in the

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

case of nearly circular inclusions. This rapid decrease of stress, characteristic for a sliding inclusion, is not a feature of the Eshelby inclusion at all, since the stress there decreases far more gradually with the ratio a/b, being practically equal in spherical and nearly spherical inclusions.

Finally, an analogous situation to the one considered in this paper in the case of shear eigenstrain, also occurs in the case of sliding elliptical or ellipsoidal inclusions under remote uniform shear loading. Here, the stress in distinctly elliptical inclusions is of the order τ^2/μ, where τ is the remote shear stress and μ the shear modulus of the material. Such inclusions behave as voids from the standpoint of linear elasticity, since they accommodate deformation of the surrounding matrix simply by an infinitesimal rotation. A nonlinear analysis is required to determine the stress in nearly circular or nearly spherical sliding inclusions. Only at very small values of the remote shear stress do such inclusions (or inhomogeneities) act as voids. The overall stiffness and the resistance of the system to applied shear stress become quite different for larger values of τ. The inclusion is unable to further adjust to the deforming matrix by a rigid-body rotation, and the normal traction builds up at the interface, causing a stress field in the inclusion, and a change in the overall stiffness of the considered matrix-inclusion composite.

Acknowledgments

Suggestions and comments made by the reviewers are kindly acknowledged. V.A.L. is also grateful for research support from the Alcoa Center, and for facilities provided by the Institute for Mechanics and Materials at UCSD.

References

Asaro, R. J., 1975, "Somigliana Dislocations and Internal Stresses; With Application to Second Phase Hardening," *Int. J. Engng. Sci.*, Vol. 13, pp. 271–286.

Eshelby, J. D., 1957, "The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems," *Proc. Roy. Soc.*, Vol. A241, pp. 376–396.

Furuhashi, R., Huang, J. H., and Mura, T., 1992, Sliding Inclusions and Inhomogeneities With Frictional Interfaces," ASME JOURNAL OF APPLIED MECHANICS, Vol. 59, pp. 783–788.

Kouris, D. A, Tsuchida, E., and Mura, T., 1986, "An Anomaly of Sliding Inclusions," ASME JOURNAL OF APPLIED MECHANICS, Vol. 53, pp. 724–726.

Lubarda, V. A., and Markenscoff, X., 1999, "Energy Analysis of Circular Inclusions: Sliding vs. Bonded Interface," *Proc. Roy. Soc. Lond. A,* in press.

Mura, T., 1985, "General Theory of Inclusions," *Fundamentals of Deformation and Fracture: Eshelby Memorial Symposium,* B. A. Bilby, K. J. Miller, and J. R. Willis, eds., Cambridge University Press, Cambridge, pp. 75–89.

Mura, T., and Furuhashi, R., 1984, "The Elastic Inclusion with a Sliding Interface," ASME JOURNAL OF APPLIED MECHANICS, Vol. 51, pp. 308–310.

Wong, G. K., and Barnett, D. M., 1985, "A Dislocation Method for Solving 3-D Crack and Inclusion Problems in Linear Elastic Solids," *Fundamentals of Deformation and Fracture: Eshelby Memorial Symposium,* B. A. Bilby, K. J. Miller and J. R. Willis, eds., Cambridge University Press, Cambridge, pp. 417–437.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Inclusions of Uniform Eigenstrains and Constant or Other Stress Dependence

X. Markenscoff
Professor,
Department of Applied Mechanics
and Engineering Sciences,
University of California, San Diego
La Jolla, CA 92093-0411
Fellow ASME

It is shown that the Eshelby inclusions, that is inclusions of uniform eigenstrains and constant eigenstress, form in three dimensions a connected nine-dimensional manifold, and, as a consequence, the only perturbations of an ellipsoid that preserve the Eshelby property are into another ellipsoid. The method is illustrated in both two and three dimensions, and can be also used to treat similar questions of other than constant dependence of the stress.

1 Inclusions With Constant Eigenstress

In a celebrated paper, Eshelby (1957) proved that the stress field inside an ellipsoidal inclusion of uniform eigenstrain is constant. He also conjectured (Eshelby, 1961) that the ellipsoidal inclusions are the only ones to have this remarkable property. Since the publication of Eshelby's papers this question has remained unanswered and some controversy has risen on that matter, with claims being made that eigenstresses are constant inside inclusions of other shapes. Recently Rodin (1996) has shown by an elaborate calculation that constant eigenstresses are impossible in inclusions with corners, while Markenscoff (1998), by a simple mathematical proof, has shown that polyhedra are prohibited. In a recent paper Markenscoff (1998) has proved that the set of inclusions that have constant eigenstress produced by uniform eigenstrains is a nine-dimensional manifold (each ellipsoid is characterized by nine parameters, namely the three coordinates of the center, the three Euler angles that determine the directions of the axes in space, and the lengths of the three axes). Here we provide an alternate proof and demonstrate that the only perturbation of an ellipsoid that preserves the constancy of the stress is into another ellipsoid. We illustrate these perturbations both in two dimensions and three dimensions, which make the final arguments clearer.

The methodology presented here can be extended also to inclusions of uniform eigenstrains and other than constant, e.g., linear, stress inside the inclusion. This approach may also be applicable to inverse problems in mechanics, such as those treated by Vidergaut (1983), Wheeler (1976) and others, and to contact problems (Willis, 1966).

Suppose that the elastic material with elasticity tensor C_{ijkl} contains an inclusion which occupies a domain Ω_0 (that contains the origin $x = 0$) with boundary $\partial\Omega_0$. The displacement field inside Ω_0 generated by constant eigenstrains ϵ^*_{mn} is given by (see, Mura, 1982, p. 33).

$$u_i(\mathbf{x}) = -C_{jkmu}\epsilon^*_{mn}\int_{\Omega_0} G_{ij,k}(\mathbf{x}-\mathbf{x}')d\mathbf{x}' \qquad (1)$$

where $G_{ij}(\mathbf{x}-\mathbf{x}')$ denotes the Green's function for an infinite space and C_{ijkl} the elasticity tensor for a homogeneous anisotropic elastic body. The domain Ω_0 will have the Eshelby property of constant stress if the derivatives $u_{i,l_1,\ldots,l_m}$ vanish identically in Ω_0 for every $m \geq 2$. Notice that since the field $u_i(\mathbf{x})$ is real analytic the above derivatives will be vanishing everywhere if they vanish at just one point of Ω_0, say for definiteness at the origin $x = 0$. The reader should be alerted here that in computing the derivatives $u_{1,l_1,\ldots,l_m}$ we are not allowed to pass the derivatives onto the Green's function G_{ij} under the integral sign, because the derivatives of the Green's function have nonintegrable singularities at $x - x' = 0$. This remark seems to have escaped the attention of certain authors in the field.

We now ask the question whether the Eshelby property is being preserved under smooth perturbations of the shape of the inclusion. We embed Ω_0 in a smooth family of domains $\Omega(t)$ varying with a real parameter t, say $\Omega_0 = \Omega(0)$. We ask whether, if Ω_0 exhibits the Eshelby property, $\Omega(t)$ will also exhibit the Eshelby property for t near 0. The displacement $u_i(\mathbf{x}, t)$ in $\Omega(t)$ generated by constant eigenstrains ϵ^*_{mn} is given by

$$u_i(\mathbf{x}, t) = -C_{jkmn}\epsilon^*_{mn}\int_{\Omega(t)} G_{ij,k}(\mathbf{x}-\mathbf{x}')d\mathbf{x}'. \qquad (2)$$

Differentiating with respect to t yields at $t = 0$

$$\dot{u}_i(x, 0) = -C_{jkmn}\epsilon^*_{mn}\oint_{\partial\Omega_0} G_{ij,k}(\mathbf{x}-\mathbf{x}')v(\mathbf{x}')dS_{x'} \qquad (3)$$

where v denotes the normal velocity of the boundary $\partial\Omega_0$ under the perturbation of the domain [for example, if the family $\Omega(t)$ is determined by diffeomorphisms $\chi(\cdot, t)$: $\Omega_0 \to \Omega(t)$, then $v(x) = \dot{\chi}(x, 0)\cdot\mathbf{n}$]. We may compute derivatives of $\dot{u}_i(x, 0)$ with respect to x to arbitrary order as follows:

$$\dot{u}_{i,l_1\ldots l_m}(x, 0) = -C_{jkmn}\epsilon^*_{mn}\oint_{\partial\Omega_0} G_{ij,kl_1\ldots l_m}(x-x')v(x')dS_{x'}. \qquad (4)$$

Notice that in arriving at formula (4) it is allowed to pass the differentiation onto the Green's function under the integral sign, because the integration runs over the boundary $\partial\Omega_0$ on which the derivatives of G_{ij} are smooth, since $x - x'$ never vanishes.

A necessary condition for the Eshelby property to hold for $\Omega(t)$, with t sufficiently small is that

$$\dot{u}_{i,l_1\ldots l_m}(0, 0) = 0, \quad \text{for all } m \geq 2. \qquad (5)$$

In view of (4), the ensuing from (5) infinite family of integral equations characterize the class of functions $v(x')$ that define the domains which preserve the Eshelby property. It is obvious that this class is very sparse, since, if, for a $v(x')$ defined on $\partial\Omega_0$, at least one of the integrals

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Dec. 1, 1997; final revision, Jan. 10, 1998. Associate Technical Editor: J. R. Barber.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\int_{\partial\Omega_0} G_{ij,kl_1\ldots l_m}\, v(x')dS_{x'} \neq 0, \tag{6}$$

then the Eshelby property is destroyed. Obviously if the Eshelby property fails for a domain Ω_0, it also fails for a perturbation of this domain because of the continuous dependence of the value of (6) on the perturbation of the domain.

2 The Two-Dimensional Case

We will now characterize the perturbations that preserve the Eshelby property when the domain of the inclusion is a circle in an isotropic material in plane elastic deformation, and the eigenstrain $\epsilon^*_{mn} = \delta_{mn}\epsilon^*$ is a uniform dilatation. In this case the Green's function is (see, Mura, 1982, p. 22)

$$G_{ij}(x) = \frac{1}{8\pi(1-\nu)\mu}\frac{x_i x_j}{R^2} - \frac{3-4\nu}{2}\delta_{ij}\log R^2 \tag{7}$$

where $R^2 = x_k x_k$, $C_{jkmn}\epsilon^*_{mn} = 2(\lambda+\mu)\delta_{jk}\epsilon^*$ (in two dimensions), and (3) reduces to

$$\dot{u}_i(x) = \epsilon^* \int_{|x'|=1} G_{ij,j}(x-x')v(x')dS_{x'} \tag{8}$$

with

$$G_{ij,j}(x-x') = (4\nu-2)\frac{x_i - x_i'}{R^2}\frac{1}{8\pi(1-\nu)\mu}. \tag{9}$$

In order to apply formula (8), it will be convenient but not necessary to use complex variables notation (the analysis was also performed with real variables). We set $z = x_1 + ix_2$ and let $A \equiv G_{1j,j} - iG_{2j,j}$, so that in view of (9)

$$A(x) = 2(2\nu-1)\frac{x_1 - ix_2}{R^2} = 2(2\nu-1)\frac{\bar{z}}{z\bar{z}}$$

$$= 2(2\nu-1)\frac{1}{z}. \tag{10}$$

Hence,

$$\frac{\partial^{k+l}A}{\partial x_1^k \partial x_2^l} = i^l \frac{d^{k+l}}{dz^{k+l}}\left(\frac{1}{z}\right) = i(k+l)!z^{-(k+l+1)}. \tag{11}$$

If $\partial\Omega_0$ is the unit circle $z = e^{i\theta}$ and (11) is written

$$\frac{\partial^{k+l}A}{\partial x_1^k \partial x_2^l} = i^l(k+l)!e^{-i(k+l+1)\theta}, \tag{12}$$

it follows that the condition (5) becomes

$$\int_0^{2\pi} e^{-im\theta}v(\theta)d\theta = 0 \quad m = 3, 4, \ldots. \tag{13}$$

In real terms (13) is equivalent to

$$\int_0^{2\pi}\cos(m\theta)v(\theta)d\theta = 0, \quad \int_0^{2\pi}\sin(m\theta)v(\theta)d\theta = 0,$$

$$m = 3, 4, \ldots \tag{14}$$

which by completeness and orthogonality properties of the trigometric functions yields that $v(\theta)$ must be of the form

$$v(\theta) = a_0 + a_1\cos\theta + b_1\sin\theta$$

$$+ a_2\cos 2\theta + b_2\sin 2\theta. \tag{15}$$

Let us return to the visualization of the Eshelby set as a manifold embedded in the infinite dimensional space of all inclusions. The function family (15) determines the tangent space to the Eshelby set. Notice that this tangent space is five dimensional. Therefore, assuming that the conditions of an implicit function theorem hold, the Eshelby set itself must have, locally at least, the structure of a five-dimensional manifold (thus one derivative with respect to time suffices to determine the number of parameters of the Eshelby manifold). We know, of course, through Eshelby's solution that we can generate such a five-parameter family through ellipses (two parameters corresponding to the two principal axes, their rigid translations (two parameters) and their rigid rotation (one parameter)). Hence this family of ellipses exhausts the family of shapes, in the vicinity of the circle, that satisfies the Eshelby property. It is instructive to verify here independently how the circle is transformed into an ellipse. Indeed, if

$$x_1(t) = a(t)\cos(\theta + \varphi(t)) + c_1(t)$$
$$x_2(t) = b(t)\sin(\theta + \varphi(t)) + c_2(t) \tag{16}$$

with

$$a(0) = b(0) = 1, \quad c_1(0) = c_2(0) = 0,$$

is a family of ellipses emanating at $t = 0$ from the unit circle, then

$$\dot{x}_1(0) = \dot{a}\cos(\theta+\varphi) - \dot{\varphi}\sin(\theta+\varphi) + \dot{c}_1$$
$$\dot{x}_2(\theta) = \dot{b}\sin(\theta+\varphi) + \dot{\varphi}\cos(\theta+\varphi) + \dot{c}_2. \tag{17}$$

Since $\mathbf{n} = (\cos\theta, \sin\theta)$

$$v = \dot{x}\cdot\mathbf{n} = \dot{a}\cos(\theta+\varphi)\cos\theta - \dot{\varphi}\sin(\theta+\varphi)\cos\theta$$
$$+ \dot{c}_1\cos\theta + \dot{b}\sin(\theta+\varphi)\sin\theta + \dot{\varphi}\cos(\theta+\varphi)\sin\theta$$
$$+ \dot{c}_2\sin\theta = \dot{a}\cos\varphi\cos^2\theta - \dot{a}\sin\varphi\sin\theta\cos\theta$$
$$+ \dot{c}_1\cos\theta + \dot{b}\cos\varphi\sin^2\theta + \dot{b}\sin\varphi\sin\theta\cos\theta$$
$$+ \dot{c}_2\sin\theta + \dot{\varphi}[\sin\theta\cos(\theta+\varphi) - \cos\theta\sin(\theta+\varphi)]$$
$$= \tfrac{1}{2}\dot{a}\cos\varphi[1+\cos 2\theta] - \tfrac{1}{2}\dot{a}\sin\varphi\sin 2\theta + \dot{c}_1\cos 2\theta$$
$$+ \tfrac{1}{2}\dot{b}\cos\varphi[1-\cos 2\theta] + \tfrac{1}{2}\dot{b}\sin\varphi\sin 2\theta + \dot{c}_2\sin\theta$$
$$- \dot{\varphi}\sin\varphi = \tfrac{1}{2}(\dot{a}-\dot{b})\cos\varphi\cos 2\theta$$
$$+ \tfrac{1}{2}(\dot{b}-\dot{a})\sin\varphi\sin 2\theta + \dot{c}_1\cos\theta + \dot{c}_2\sin\theta$$
$$+ [\tfrac{1}{2}(\dot{a}+\dot{b})\cos\varphi - \dot{\varphi}\sin\varphi] \tag{18}$$

which indeed satisfies (14).

The above analysis shows in the simple case of the circle that the shape of the inclusion must be orthogonal to the set $(\sin m\theta, \cos m\theta, m = 3, 4\ldots,)$. Since the set of functions $\sin m\theta$, $\cos m\theta$, $m = 1, 2\ldots$, is complete, $v(\theta)$ must necessarily be in the orthogonal complement of $\sin m\theta$, $\cos m\theta$, $m = 3, 4\ldots$, and this determines the dimension of the Eshelby manifold. Since, in addition here the set $\sin m\theta$, $\cos m\theta$ is orthogonal, $v(\theta)$ must be a linear combination of the functions spanning the orthogonal complement, i.e., (15). In the three-dimensional case treated by Markenscoff (1997) in place of the trigonometric functions $\sin m\theta$, $\cos m\theta$ are the functions $(1/|y|)_{,i_1\ldots i_m}$ for all $i_1, \ldots, i_m$ and $m \geq 0$, which will also proven here to be complete (but not orthogonal). We proceed to the proof as follows:

3 The Three-Dimensional Case

In three-dimensional isotropic elasticity, we have

$$C_{jkmn} = \frac{2\mu\nu}{1-2\nu}\delta_{jk}\delta_{mn} + \mu(\delta_{jm}\delta_{kn} + \delta_{jn}\delta_{km}), \tag{19}$$

$$G_{ij,j}(x) = \frac{1}{16n\pi\mu(1-\nu)}\left[(3-4\nu)\frac{\delta_{ij}}{|x|} + \frac{x_i x_j}{|x|^3}\right] \tag{20}$$

so that after a short calculation, (3), with $\epsilon^*_{mn} = \delta_{mn}\epsilon^*$, becomes

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\dot{u}_i(x, 0) = -\frac{\epsilon^*}{4\pi}\frac{1+\nu}{1-\nu}\oint_{\partial\Omega_0}\left(\frac{1}{|x-y|}\right)_{,i} v(y)dS_y. \quad (21)$$

Therefore, (5) reduces to

$$\oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{j_1\ldots j_m} v(y)dS_y = 0, \quad (22)$$

for all $_{j_1\ldots j_m}$ with $m \geq 3$.

For $w(y)$ continuous, we define the function

$$\omega(x) = \oint_{\partial\Omega_0}\frac{1}{|x-y|}\, w(y)dS_y. \quad (23)$$

As is well known (e.g., Courant-Hilbert, Vol. II, Chap. IV, Sections 2, 4) ω is continuous on R^3, harmonic

$$\nabla^2\omega = 0 \quad (24)$$

on Ω_o, as well as on the complement Ω_o^c of $\Omega_o \cup \partial\Omega_o$, and

$$4\pi w = \left(\frac{\partial\omega}{\partial n}\right)_- + \left(\frac{\partial\omega}{\partial n}\right)_+. \quad (25)$$

We identify the subset (linear manifold) H of Lipschitz continuous functions w defined on $\partial\Omega_o$ which induce ω lying in the Sobolev space $W^{1,2}(\Omega_o^c)$. Notice that, since the stress field induced by the inclusion has zero resultant, $\dot{u}$ decays at infinity as $0(R^{-2})$ and so, in particular, $v(y)$ lies in H.

We now show that if $w(y)$ in H satisfies

$$\oint_{\partial\Omega_o}\left(\frac{1}{|y|}\right)_{,j_1\ldots j_m} w(y)dS_y = 0 \quad (26)$$

for all $_{j_1\ldots j_m}$ with $m \geq 0$, then $w(y) \equiv 0$. Indeed, from (23) we deduce

$$\omega_{,j_1\ldots j_m}(0) = (-1)^m \oint_{\partial\Omega_o}\left(\frac{1}{|y|}\right)_{,j_1\ldots j_m} w(y)dS_y = 0. \quad (27)$$

Since ω is real analytic in Ω_0, (27) implies that ω vanishes identically in Ω_o and hence, by continuity, also on $\partial\Omega_o$. Now ω is a solution of (24) in $W^{1,2}(\Omega_o^c)$, with zero boundary conditions on $\partial\Omega_o$ and hence, by standard uniqueness, ω vanishes identically also on Ω_o^c. Then (25) implies $w(y) \equiv 0$, as asserted above.

In view of the above, the set of $v(y)$ which satisfy (22), for all $_{j_1\ldots j_m}$, with $m \geq 3$, forms a finite dimensional subspace of H whose dimension cannot exceed the dimension of the subspace spanned by the 13 functions

$$\frac{1}{|y|}, \quad \left(\frac{1}{|y|}\right)_{,i}, \quad \left(\frac{1}{|y|}\right)_{,ij} \quad i, j = 1, 2, 3. \quad (28)$$

Note that

$$\left(\frac{1}{|y|}\right)_{,i} = -\frac{y_i}{|y|^3} \quad (29)$$

$$\left(\frac{1}{|y|}\right)_{,ij} = \frac{3y_iy_j}{|y|^5} - \frac{\delta_{ij}}{|y|^3}. \quad (30)$$

Hence the above 13 functions satisfy the four linear relations

$$\left(\frac{1}{|y|}\right)_{,ij} = \left(\frac{1}{|y|}\right)_{,ji} \quad 1 \leq i < j \leq 3 \quad (31)$$

$$\left(\frac{1}{|y|}\right)_{,ii} = 0 \quad (32)$$

but are otherwise linearly independent. Therefore, the space of $v(y)$ that satisfy (22), that is, the tangent space to the Eshelby manifold, is necessarily nine dimensional. Therefore, the Eshelby set itself is a nine-dimensional manifold.

4 Perturbations of the Ellipsoid

For illustration, let us determine explicitly the tangent space of the Eshelby set when Ω_0 is an ellipsoid, say,

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \frac{x_3^2}{a_3^2} = 1. \quad (33)$$

The diffeomorphisms $\chi(., t)$: $\Omega_0 \mapsto \Omega(t)$ which deform the ellipsoid into another ellipsoid may be parametrized as

$$\chi_i(x, t) = tc_i + x_i + tA_{ij}x_j \quad (34)$$

where c_i is a constant vector inducing rigid translations, and A_{ij} is some constant matrix which induces the change in the orientation and length of the axes of the ellipsoid. The corresponding normal speed of the boundary at $t = 0$ is

$$\tilde{v}(y) = \dot{\chi}(y, 0)\cdot n = c_in_i + A_{ij}n_jy_i \quad (35)$$

where

$$n_i = \frac{1}{a_i}\frac{y_i}{\sigma(y)}, \quad \sigma(y) = \sqrt{\frac{y_1^2}{a_1^2} + \frac{y_2^2}{a_2^2} + \frac{y_3^2}{a_3^2}}. \quad (36)$$

Assuming that $v(y)$ satisfies (22), we proceed to determine c_i and A_{ij} such that the $\tilde{v}(y)$ induced by (35) satisfies

$$\oint_{\partial\Omega_0}\frac{1}{|y|}\tilde{v}(y)dS_y = \oint_{\partial\Omega_0}\frac{1}{|y|}v(y)dS_y \quad (37)$$

$$\oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,i}\tilde{v}(y)dS_y = \oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,i}v(y)dS_y$$

$$i = 1, 2, 3 \quad (38)$$

$$\oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,ij}\tilde{v}(y)dS_y = \oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,ij}v(y)dS_y$$

$$i, j = 1, 2, 3. \quad (39)$$

Using (35), (36), (28), and the obvious symmetries of the ellipsoid, we easily see that (37) yields

$$A_{11}\oint_{\partial\Omega_0}\frac{y_1^2}{a_1|y|\sigma(y)}dS_y + A_{22}\oint_{\partial\Omega_0}\frac{y_2^2}{a_2|y|\sigma(y)}dS_y + A_{33}\oint_{\partial\Omega_0}\frac{y_3^2}{a_3|y|\sigma(y)}dS_y = \oint_{\partial\Omega_0}\frac{v(y)}{|y|}dS_y. \quad (40)$$

Similarly, (38) reduces to

$$-c_i\oint_{\partial\Omega_0}\frac{y_1^2}{|y|^3\sigma(y)a_i}dS_y = \oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,i}v(y)dS_y$$

$$i = 1, 2, 3 \text{ (no summation on } i). \quad (41)$$

For (39) for $i \neq j$ gives

$$3A_{ij}\oint_{\partial\Omega_0}\frac{y_i^2y_j^2}{|y|^5a_i\sigma(y)}dS_y = \oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,ij}v(y)dS_y$$

$$i \neq j,\ i, j = 1, 2, 3 \text{ (no summation on } i, j). \quad (42)$$

Finally (39) for $i = j = 1, 2, 3$ gives

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$A_{11}\left\{3\oint_{\partial\Omega_0}\frac{y_1^4}{|y|^5 a_1\sigma(y)}\,dS_y-\oint_{\partial\Omega_0}\frac{y_1^2}{|y|^3 a_1\sigma(y)}\,dS_y\right\}$$
$$+A_{22}\left\{3\oint_{\partial\Omega_0}\frac{y_1^2y_2^2}{|y|^5 a_2\sigma(y)}\,dS_y-\oint_{\partial\Omega_0}\frac{y_2^2}{|y|^3 a_2\sigma(y)}\,dS_y\right\}$$
$$+A_{33}\left\{3\oint_{\partial\Omega_0}\frac{y_1^2y_3^2}{|y|^5 a_3\sigma(y)}\,dS_y-\oint_{\partial\Omega_0}\frac{y_3^2}{|y|^3 a_3\sigma(y)}\,dS_y\right\}$$
$$=\oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,11}v(y)dS_y \quad (43)$$

$$A_{11}\left\{3\oint_{\partial\Omega_0}\frac{y_1^2y_2^2}{|y|^5 a_1\sigma(y)}\,dS_y-\oint_{\partial\Omega_0}\frac{y_1^2}{|y|^3 a_1\sigma(y)}\,dS_y\right\}$$
$$+A_{22}\left\{3\oint_{\partial\Omega_0}\frac{y_2^4}{|y|^5 a_2\sigma(y)}\,dS_y-\oint_{\partial\Omega_0}\frac{y_2^2}{|y|^3 a_2\sigma(y)}\,dS_y\right\}$$
$$+A_{33}\left\{3\oint_{\partial\Omega_0}\frac{y_1^2y_3^2}{|y|^5 a_3\sigma(y)}\,dS_y-\oint_{\partial\Omega_0}\frac{y_3^2}{|y|^3 a_3\sigma(y)}\,dS_y\right\}$$
$$=\oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,22}v(y)dS_y \quad (44)$$

$$A_{11}\left\{3\oint_{\partial\Omega_0}\frac{y_1^2y_3^2}{|y|^5 a_1\sigma(y)}\,dS_y-\oint_{\partial\Omega_0}\frac{y_1^2}{|y|^3 a_1\sigma(y)}\,dS_y\right\}$$
$$+A_{22}\left\{3\oint_{\partial\Omega_0}\frac{y_2^2y_3^2}{|y|^5 a_2\sigma(y)}\,dS_y-\oint_{\partial\Omega_0}\frac{y_2^2}{|y|^3 a_2\sigma(y)}\,dS_y\right\}$$
$$+A_{33}\left\{3\oint_{\partial\Omega_0}\frac{y_3^4}{|y|^5 a_3\sigma(y)}\,dS_y-\oint_{\partial\Omega_0}\frac{y_3^2}{|y|^3 a_3\sigma(y)}\,dS_y\right\}$$
$$=\oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,33}v(y)dS_y. \quad (45)$$

Clearly, (41) determine the c_i, $i = 1, 2, 3$ and Eqs. (19) determine the A_{ij} for $i \neq j$. Equations (40), (43), (44), (45) constitute four linear equations in three unknowns A_{11}, A_{22}, A_{33}. Notice, however, that the four equations are compatible, because adding up the three equations (33), (34), (35), we end up with an identity. Once the c_i and A_{ij} have thus determined, we claim that $v(y) \equiv \tilde{v}(y)$. Indeed, upon setting $w(y) = v(y) - \tilde{v}(y)$, we have

$$\oint_{\partial\Omega_0}\left(\frac{1}{|y|}\right)_{,i_1\ldots i_m}w(y)dS_y=0 \quad (46)$$

for all $i_1, \ldots, i_m$ with $m \geq 0$. By completeness proved above, we conclude that $w(y) = 0$. This completes the proof.

5 Other Than Constant Stress in the Inclusion

This approach allows us to ask questions whether there are, in the neighborhood of the circle, shapes of inclusions inside which the stress is linear or higher-order function of position. For instance, in the case of linear stress inside the inclusion, (14) would hold but for $m = 4$, 5 and the dimension of the tangent space would be seven, namely v would be

$$v(\theta) = a_0 + a_1\cos\theta + b_1\sin\theta + a_2\cos 2\theta + b_2\sin 2\theta + a_3\cos 3\theta + b_3\sin 3\theta. \quad (47)$$

This is based of course, on the assumption that the implicit function theorem holds, which would affirm that the dimension of the manifold is the same as the dimension of the tangent. While (47) would determine the shape of the inclusion in the neighborhood of the circle (which, has constant stress and hence is trivially a manifold of linear stress), this approach does not allow the determination of shapes of manifolds of seven dimension. Similar argument apply to three dimensions.

The approach that has been developed here may, in principle, also be applicable to contact problems (e.g., Gladwell, 1980) and to inverse problems, such as those treated by Vigdergauz (1983), Wheeler (1976), and others.

References

Courant, R., and Hilbert, D., 1962, *Methods of Mathematical Physics,* John Wiley and Sons, New York.

Eshelby, J. D., 1957, "The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems," *Proc. Roy. Soc.,* Vol. A241, p. 376.

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.

Gladwell, G. M. L., 1980, *Contact Problems in the Classical Theory of Elasticity,* Sijthoff and Noordhoff, Alpen aan den Jijn, The Netherlands.

Markenscoff, X., 1998, "On the Shape of the Eshelby Inclusions," *J. Elasticity,* Vol. 44, pp. 163–166.

Markenscoff, X., 1998, "Inclusions with Constant Eigenstress," *J. Mech. Phys. Sol.,* Vol. 46, pp. 2297–2301.

Mura, T., 1982, *Micromechanics of Defects in Solids,* Martinus Nijhoff, The Hague, The Netherlands.

Rodin, G. J., 1996, "Eshelby's Inclusion Problem for Polygons and Polyhedra," *Mech. Phys. Sol.,* Vol. 44, pp. 1977–1995.

Vigdergauz, S. B., 1983, "Inverse Problem of Three-Dimensional Elasticity," *MTT,* Vol. 18, No. 2, pp. 83–86.

Wheeler, L., 1976, "On the Role of Constant-Stress Surfaces in the Problem of Minimizing Elastic Stress Concentration," *Int. J. Solids Structures,* Vol. 12, pp. 779–789.

Willis, J. R., 1966, "Hertzian Contact of Anisotropic Bodies," *J. Mech. Phys. Sol.,* Vol. 14, pp. 163–176.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

A Nonuniform Cell Partition for the Analysis of Nonlinear Stochastic Systems

J. Q. Sun
Associate Professor,
Department of Mechanical Engineering,
University of Delaware,
Newark, DE 19716
Mem. ASME

This paper presents a study of nonuniform cell partition for analyzing the response of nonlinear stochastic systems by using the generalized cell mapping (GCM) method. The necessity of nonuniform cell partition for nonlinear systems is discussed first. An ad hoc scheme is then presented for determining optimal cell sizes based on the statistical analysis of the GCM method. The proposed nonuniform cell partition provides a roughly uniform accuracy for the estimate of the one-step transition probability density function over a large region in the state space where the system varies significantly from being linear to being strongly nonlinear. The nonuniform cell partition is shown to lead to more accurate steady-state solutions and enhance the computational efficiency of the GCM method.

1 Introduction

The generalized cell mapping (GCM) method is a method for global analysis of deterministic dynamical systems (Hsu, 1987), and has been extended to stochastic systems (Hsu and Chiu, 1986; Sun and Hsu, 1987, 1988a; Sun, 1988; Sun and Hsu, 1990; Sun and Hsu, 1991; Sun, 1995). The extension is made possible by the fact that the GCM method, mathematically built upon the theory of Markov chains, is a probabilistic description of dynamical systems. There are many other references on the GCM method and its applications. Most of the published studies have used uniform cell partitions of the state space. However, the nonlinear systems having very nonuniform dynamic behavior in the state space really call for nonuniform cell partitions. This paper presents such a study. An ad hoc nonuniform cell partition scheme is presented that is based on the statistical analysis of the GCM method (Sun and Hsu, 1988b). A study of this kind within the context of interpolated cell mapping for deterministic systems is presented in (Tongue and Gu, 1989).

In Section 2, we briefly describe the nonuniform cell partition scheme. In Section 3, we present a numerical example of a nonlinear stochastic system to demonstrate the effectiveness of the proposed partition scheme.

2 Optimal Cell Partition

The mean square error of the simulated transition probability density function for a one-degree-of-freedom nonlinear system is given by (Sun 1988; Sun and Hsu 1988b)

$$MES(\hat{p}_{ij}) = \frac{1}{576}\left[\Delta x^2\alpha(\mathbf{x}_i) + \Delta\dot{x}^2\beta(\mathbf{x}_i)\right]^2 + \frac{p(\mathbf{x}_i, t_0 + T|\mathbf{x}_j, t_0)}{M\Delta x\Delta\dot{x}}, \quad (1)$$

where

$$\alpha(\mathbf{x}) = \frac{\partial^2 p(\mathbf{x}, t_0 + T|\mathbf{x}_j, t_0)}{\partial x^2},$$

$$\beta(\mathbf{x}) = \frac{\partial^2 p(\mathbf{x}, t_0 + T|\mathbf{x}_j, t_0)}{\partial \dot{x}^2}. \quad (2)$$

$\hat{p}_{ij}$ is the estimate of the density function $p(\mathbf{x}_i, t_0 + T|\mathbf{x}_j, t_0)$, and is used to construct the transition probability matrix of the GCM method. M is the number of samples used in the simulation, T is the mapping time step, t_0 is an initial time, $\mathbf{x}_j$ represents the center of the jth cell in the state space spanned by $(x, \dot{x})$, and Δx and $\Delta\dot{x}$ are the cell sizes in the corresponding directions. The first part of the mean square error is the bias squared, and the second part is the variance of the statistical estimation.

Note that for a given number of samples M, there is a trade-off between the bias and the variance with respect to the cell dimensions Δx and $\Delta\dot{x}$. With smaller cells, the bias error is decreased while the variance is increased. With larger cells, the opposite occurs. This trade-off feature of the mean square error of $\hat{p}_{ij}$ suggests that there may be an optimal cell dimension that compromises the bias and variance in order to minimize the mean square error.

By minimizing the mean square error in Eq. (1) with respect to Δx and $\Delta\dot{x}$, we obtain the optimal cell sizes:

$$\Delta x_{\text{opt}} = |\alpha(\mathbf{x}_i)|^{-5/12}|\beta(\mathbf{x}_i)|^{1/12}\left(\frac{72p(\mathbf{x}_i, t_0 + T|\mathbf{x}_j, t_0)}{M}\right)^{1/6},$$

$$\Delta\dot{x}_{\text{opt}} = |\alpha(\mathbf{x}_i)|^{1/12}|\beta(\mathbf{x}_i)|^{-5/12}\left(\frac{72p(\mathbf{x}_i, t_0 + T|\mathbf{x}_j, t_0)}{M}\right)^{1/6}. \quad (3)$$

The above optimal cell dimensions are in general functions of both the initial cell location $\mathbf{x}_j$ and the image cell location $\mathbf{x}_i$ for nonlinear systems. This is a reflection of the nonuniform dynamic behavior of nonlinear systems in the state space. As discussed in (Sun and Hsu 1990), when the mapping time step T is small, the probability density function is nearly a Gaussian distribution with a small standard derivation. Hence, for a given initial cell $\mathbf{x}_j$, $p(\mathbf{x}, t_0 + T|\mathbf{x}_j, t_0)$ is a very sharp function of $\mathbf{x}$ near its mean value. More than 90 percent transition probabilities may be captured by a few cells around the mean of the density function $p(\mathbf{x}, t_0 + T|\mathbf{x}_j, t_0)$. It can be shown that the smallest cell dimensions given by Eq. (3) among all the image cells $\mathbf{x}_i$ of $\mathbf{x}_j$ within a region to which $\mathbf{x}_j$ is mapped in one step

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Nov. 11, 1997; final revision, May 11, 1998. Associate Technical Editor: A. A. Ferri.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

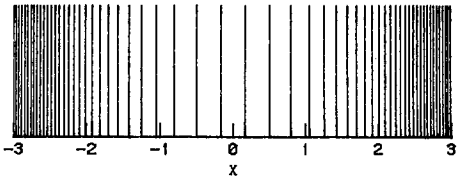


Fig. 1 The nonuniform cell partition for the example

occurs at a cell, denoted by $\hat{\mathbf{x}}_i$, that contains the mean value location of this density function. We shall take these smallest cell dimensions as the optimal sizes at the cell $\hat{\mathbf{x}}_i$ associated with an initial cell $\mathbf{x}_j$.

A procedure to create a nonuniform cell partition for nonlinear systems by using the above optimal cells can be stated as follows:

1 Pick an arbitrary initial state in the domain of interest, and calculate the one-step transition probability density function.
2 Find the smallest cell sizes and the corresponding cell location $\hat{\mathbf{x}}_i$.
3 From an outer edge of the so-determined cells $\hat{\mathbf{x}}_i$, integrate the nonlinear system without stochastic terms backward in time for one mapping step to pick a new initial state.
4 Determine the optimal cell sizes and the cell location for this new initial state. Modify the initial state, if necessary, to ensure that the new image cell is exactly adjacent to the previously determined optimal cells.
5 Repeat from step 3 until the whole domain of interest is completely covered.

It should be noted that the nonuniform cell structure determined by the this procedure does not guarantee that each cell exactly matches the optimal cell associated with the center of a pre-image cell. Nevertheless, this ad hoc procedure creates the necessary distribution density of cells in different regions of the state space such that the accuracy of the estimate of one-step transition probability density function is roughly same in these regions.

3 An Example

Consider a one-dimensional nonlinear system in the sense of Stratonovich (Wehner and Wolfer 1983a, b):

$$\dot{x} + \frac{1}{4}\tanh(x) = \frac{w(t)}{\sqrt{2}\cosh(x)}, \tag{4}$$

where $w(t)$ is the physical Gaussian white noise such that

$$E[w(t)] = 0, \quad E[w(t)w(t_1)] = 2\delta(t - t_1). \tag{5}$$

The system has an exact transition probability density function given by

$$p(x, t \mid x_0, t_0) = \frac{\cosh(x)}{\sqrt{4\pi(1 - z^2)}} \exp\left[-\frac{(\sinh(x) - z\cdot\sinh(x_0))^2}{4(1 - z^2)}\right], \tag{6}$$

where $p(x, t_0 \mid x_0, t_0) = \delta(x - x_0)$ and $z = \exp[-(t - t_0)/4]$. We shall focus our attention in the region $|x| \leq 3$. In this example, $T = 0.2$ and $M = 2000$. Figure 1 shows the nonuniform cell structure generated by following the procedure presented earlier. There is a total of 53 cells with the smallest cell size 0.04 and the largest one 0.3. It takes 0.05 seconds to generate this mesh. Figure 2 shows the one-step transition probability density functions of the system for different initial conditions x_0, and the generalized cell mapping approximation of these functions. This result clearly demonstrates that a nonuniform cell state partition will provide the GCM method a uniform accuracy in a large region of the state space in which the system varies significantly from being nearly linear ($|x| \sim 0$) to being highly nonlinear ($|x| \gg 0$).

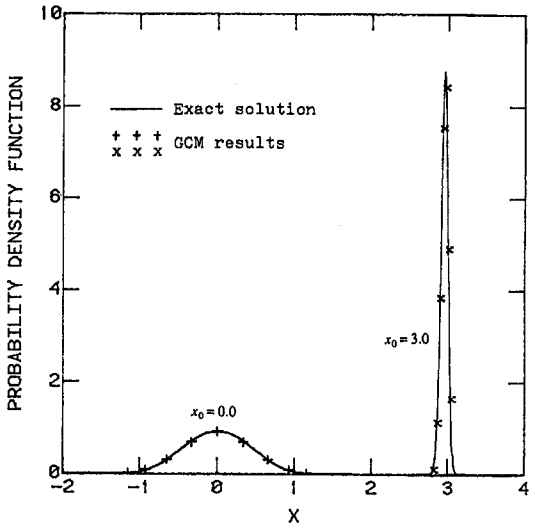


Fig. 2 The GCM estimate of one-step transition probability density function with the nonuniform cell partition for different initial conditions: $x_0 = 0$ and $x_0 = 3.0$. $T = 0.2$, $M = 2000$.

We consider the steady-state solutions of the system. We first present the results using uniform partition of the state space. Table 1 shows the steady-state mean-square values $E[x^2]$ and their errors as compared to the exact solutions computed by using the probability density function in Eq. (6).

Although the error in the computed mean-square values decreases monotonically with the cell size, the approximation of the probability density function becomes worse as the cell size becomes smaller (Sun and Hsu, 1988b). It is interesting to note that although the one-step transition probability density estimation is very poor, the steady-state mean-square values and probability density function are quite accurately calculated by the GCM method (see Fig. 3(*a*)). It seems that the steady-state distribution of a stationary Markov process is quite insensitive to the errors in the one-step transition probabilities. However, the transient responses such as the first-passage time probabilities of the system are highly sensitive to these errors (Sun and Hsu 1988a, 1990).

Next, we present the results using the nonuniform cell partition shown in Fig. 1. The steady-state mean-square response $E[x^2] = 1.0216$. Its error is only 1.02 percent. The computed steady-state probability density function is shown in Fig. 3(*b*), which does not have the visible random error as in Fig. 3(*a*), and is in closer agreement with the exact solution than the results computed with uniform cell partitions. The computation using nonuniform partition involves 53 cells and takes 50 seconds to obtain all the solutions, while the study using uniform partition involves 601 cells and takes 589 seconds for the same solutions. Apparently the nonuniform cell partition enhances the computational efficiency of the GCM method.

Table 1 The steady-state mean-square values computed by the GCM method with uniform cell partitions

Δx	0.3	0.15	0.1	0.05	0.02	0.01
$E[x^2]$	1.1146	1.0484	1.0347	1.0279	1.0263	1.0259
Error%	10.21	3.67	2.31	1.64	1.48	1.44

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

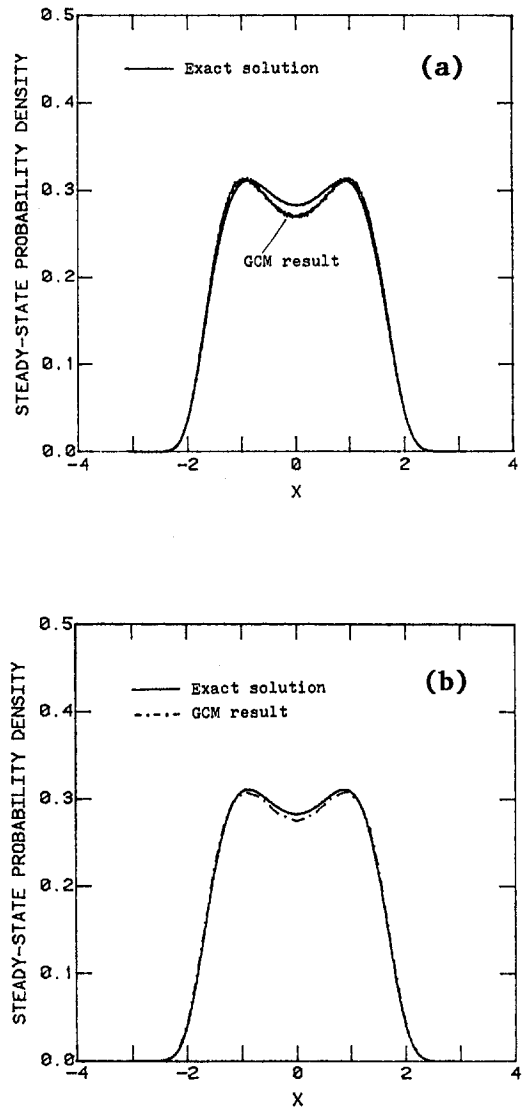


Fig. 3 The steady-state probability density function; (*a*) uniform cell partition ($\Delta x = 0.01$, $T = 0.2$, $M = 2000$), (*b*) nonuniform cell partition ($T = 0.2$, $M = 2000$)

One can expect that the nonuniform cell partition would also render the accurate prediction of the transient solutions. This is an issue being studied.

4 Conclusions

We have presented a study of nonuniform partition of the cell state space for analyzing nonlinear stochastic systems. Because of the nonlinearity, the one-step transition probability density function varies significantly with the initial conditions in the state space. The nonuniform cell partition considered in this paper offers a roughly uniform accuracy of the approximation of the one step transition probability density function over a large region in the state space, and leads to more accurate steady-state solutions. The example studied in the paper also shows that the nonuniform cell partition enhances the computational efficiency of the GCM method by allowing to use a smaller number of cells while maintaining a high accuracy.

Acknowledgment

This work is supported by a grant (CMS-9713338) from the National Science Foundation.

References

Hsu, C. S., 1987, *Cell-to-Cell Mapping: A Method of Global Analysis for Nonlinear Systems,* Springer-Verlag, Berlin.

Hsu, C. S., and Chiu, H. M., 1986, "A Cell Mapping Method for Nonlinear Deterministic and Stochastic Systems—Part I: Method of Analysis and Part II: Examples of Application. ASME JOURNAL OF APPLIED MECHANICS, Vol. 53, pp. 695–710.

Sun, J. Q., 1988, "Random Vibrations of Nonlinear Systems Based Upon the Cell State Space Concept," Ph.D. thesis, Department of Mechanical Engineering, University of California at Berkeley, Berkeley, CA.

Sun, J. Q., 1995, "Random vibration analysis of a nonlinear system with dry friction damping by the short-time gaussian cell mapping method," *Journal of Sound and Vibration,* Vol. 180, pp. 785–795.

Sun, J. Q., and Hsu, C. S., 1987, "Cumulant-Neglect Closure Method for Nonlinear Systems Under Random Excitations," ASME JOURNAL OF APPLIED MECHANICS, Vol. 54, pp. 649–655.

Sun, J. Q., and Hsu, C. S., 1988a, "First-passage time probability of non-linear stochastic systems by generalized cell mapping method," *Journal of Sound and Vibration,* Vol. 124, pp. 233–248.

Sun, J. Q., and Hsu, C. S., 1988b, "A Statistical Study of Generalized Cell Mapping," ASME JOURNAL OF APPLIED MECHANICS, Vol. 55, pp. 694–701.

Sun, J. Q., and Hsu, C. S., 1990, "The Generalized Cell Mapping Method in Nonlinear Random Vibration Based Upon Short-Time Gaussian Approximation," ASME JOURNAL OF APPLIED MECHANICS, Vol. 57, pp. 1018–1025.

Sun, J. Q., and Hsu, C. S., 1991, "Effects of Small Random Uncertainties on Nonlinear Systems Studied by the Generalized Cell Mapping Method," *Journal of Sound and Vibration,* Vol. 147, pp. 185–201.

Tongue, B. H., and Gu, K., 1989, "Adaptive mesh strategies for interpolated maping procedures," *International Journal of Engineering Science,* Vol. 27, pp. 1143–1154.

Wehner, M. F., and Wolfer, W. G., 1983a, "Numerical evaluation of path-integral solutions to Fokker-Planck equations," *Physical Review A,* Vol. 27, pp. 2663–2670.

Wehner, M. F., and Wolfer, W. G., 1983b, "Numerical evaluation of path-integral solutions to Fokker-Planck equations II. Restricted stochastic processes," *Physical Review A,* Vol. 28, pp. 3003–3011.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Stiff Elastic Tape Wrapped Onto a Drum

R. C. Benson
Department of Mechanical and Nuclear Engineering,
The Pennsylvania State University,
137 Reber Building,
University Park, PA 16802-1412
Fellow ASME

A nonlinear analysis is presented to describe the static equilibrium of a long elastic tape wrapped onto a rigid drum. Closed-form results are obtained for the case when the tape can be modeled as an inextensional beam with finite bending rigidity. Strains are assumed to be small, but rotations of the beam may be arbitrarily large. The tape's stiffness in bending requires the abandonment of a common modeling assumption—that the tape approaches the drum along a perfectly straight tangent. After calculating the actual shape of the elastica, a number of changes are seen to occur in the loading and geometry of the system. Among these are a lower tape tension at the point of tape/drum contact, an increased torque on the drum, and the introduction of an elastic compliance between the drum and distant locations in the tape.

Introduction

In countless machines, there is a long elastic tape wrapped around a rigid drum. Figure 1 shows an illustration. The range of materials includes polymers of only a few microns in thickness, to metals, several centimeters thick. In between are numerous applications involving magnetic tape, photographic film, dye-carrying webs (used in digital imaging), cloth, paper, photoreceptor belts (used in xerography), drive belts, extendible tape measures, etc. In the analysis of such systems, it is common to assume that the tape approaches the drum along a straight-line tangent.[1] The error of this assumption is measured by how much of a bending moment exists in the tape where it lies flush with the drum, compared to the zero bending moment in the straight-line tangent. A "stiff" tape, with finite bending rigidity, cannot go immediately from a straight shape to a curved shape without the application of a concentrated bending moment—something the drum cannot impart without gluing or clamping the tape to the drum. Rather, a continuous change of curvature, as illustrated in Fig. 1, is needed to achieve an equilibrated contact.

Providing that equilibrium analysis is the purpose of this paper. The tape will be modeled as an elastic beam that can deform in bending, but permits no appreciable stretching of its neutral axis. The nonlinear theory of the elastica (Frisch-Fay, 1962) will be used to calculate the small-strain large-rotation quasi-static deformation of the tape. Away from the drum, it will be assumed that the tape is pulled by a known tension, and that the tape asymptotes to the line-of-action of that tension. Closed-form results will be obtained that give the reaction forces between the tape and drum, and define the shape of the tape as it approaches the drum. It will be shown that some important differences arise in the loading and geometry when the assumption of a straight-line tangent is abandoned. The tension in the tape where it makes contact with the drum is lower than "expected," the transverse force is higher, and the torque on the drum is higher. The spacing between the tape and drum is more gradual and extended, and an elastic compliance between the drum and distant portions of the tape is introduced.

The theory of the elastica has often been used to study the transport of flexible webs and sheets. See, for example, the recent work by Mansfield and Simmonds (1987), and Stolte and Benson (1992, 1993). There is little in the open literature, however, that provides closed-form results for the present problem over the same nonlinear range. Eshel (1975) studied a stiff "foil" stretched between two parallel cylinders. He linearized the equation of equilibrium for the tape by assuming that there would be small slope differences between the finite-stiffness tape and the perfectly flexible foil. Consistent with this first assumption, he also assumed that the tension in the foil was everywhere constant. Wang (1981) studied the winding of a long sheet, but he permitted the sheet to contact the drum with a radius of curvature different from the radius of the drum. This requires that the sheet be clamped to the drum at the contact point—as happens when a tape is first attached to a reel. This is useful in predicting configurations during the first few revolutions of the tape and reel, but leads to entirely different results from the present analysis.[2] The present analysis treats the tape/drum attachment point as a variable of the analysis, and uses the requirement of matched curvatures between tape and drum as a constraint on the solution.

Fosdick and Villaggio (1986) provide the analysis closest in scope to the present analysis.[3] They studied the influence of bending stiffness on a belt wrapped 180 deg around a pulley. Like the present analysis they neglected the belt's extensibility and transverse shear deformation, the deformation of the hub, the initial curvature in the belt, and gravity. Unlike the present analysis they allowed for inertial effects and a power-law model for the elastic constitutive equation. Compared to Fosdick and Villaggio (1986), the present paper has a greater focus on some of the consequences of geometric nonlinearity, including the tape/drum spacing and the compliance caused by the "slack" in the tape. In a recent work, Müftü and Cole (1998) used the theory of the elastica to model continuous curvature changes in a web supported by an air cushion over an externally pressurized web. Other papers of interest will be identified in the body of this paper.

Mathematical Analysis

Figure 1 shows the problem of interest. An elastic tape of Young's modulus, E, Poisson's ratio, ν, thickness, t, and bending rigidity,

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Oct. 20, 1997; final revision, Jan. 8, 1998. Associate Technical Editor: L. T. Wheeler.

[1] With the reader's indulgence, the word "tape" will generically represent all of these elastic materials, and the cylinder onto which the wraps will always be referred to as a "drum." Further, the tape will always be referred to as "approaching" the drum, even though the transport could be in either direction.

[2] Contrast Fig. 3 by Wang (1981) with Fig. 2 of the present paper.

[3] Fosdick and Villaggio (1986) also provide a nice synopsis of the very early analyses of pulley belts.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

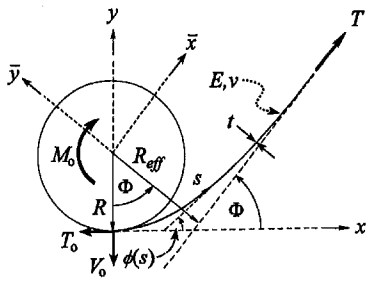


Fig. 1 Geometry

$$D = \frac{Et^3}{12(1-\nu^2)}, \tag{1}$$

is wrapped onto a rigid, cylindrical drum of radius R. Cartesian coordinates, x and y, are positioned at the point where the tape loses contact with the drum. The x-axis points tangent to the drum. Away from the drum, the tape asymptotes to a straight line that crosses the x-axis at an angle Φ. The asymptotic tension in the tape is T, which is equilibrated by the following forces and moment at the drum:

$$T_o = T\cos\Phi, \quad V_o = T\sin\Phi,$$
$$M_o = TR_{\text{eff}} = (T_oR) + (D/R). \tag{2}$$

Here, R_{eff}, is the moment arm, or "effective radius" from the center of the drum to the line-of-action of the asymptotic tension. Note that Φ and R_{eff} are unknowns that must be calculated as part of the analysis. Note, also, that the D/R term appearing in the last equation of (2) is the bending moment in the tape at all points where it lies flush with the drum.

A second pair of Cartesian coordinates, $\bar{x}$ and $\bar{y}$, is centered with the drum, with the $\bar{x}$-axis aligned with the asymptotic angle of the tape:

$$\begin{Bmatrix}\bar{x}\\ \bar{y}\end{Bmatrix} = \begin{bmatrix}\cos\Phi & \sin\Phi\\ -\sin\Phi & \cos\Phi\end{bmatrix}\begin{Bmatrix}x\\ y-R\end{Bmatrix}. \tag{3}$$

It is assumed that all effects are constant across the width of the tape[4] and that the tape may be modeled as a quasi-static, inextensible, linear elastic beam undergoing large rotations. It is also assumed that the deformation caused by the tape's weight and acceleration are small compared to the deformation caused by the applied tension. From the theory of the elastica (Frisch-Fay, 1962) we have the following equation of equilibrium:

$$D\frac{d^2\phi}{ds^2} = -T\sin[\Phi - \phi(s)]. \tag{4}$$

Here, s is the arclength along the tape, measured from the point of separation from the drum, and $\phi(s)$ is the angle, at s, that a tangent to the tape forms with the x-axis. At the separation point, $s = 0$, the angle of the tape is zero, and the curvature matches the curvature of the drum,

$$\phi(0) = 0, \quad \frac{d\phi}{ds}(0) = \frac{1}{R}. \tag{5}$$

As $s \to \infty$, the tape approaches the asymptotic angle Φ (still to be determined), and the curvature approaches zero:

$$\phi(\infty) = \Phi, \quad \frac{d\phi}{ds}(\infty) = 0. \tag{6}$$

In employing the boundary conditions of (6), it is assumed that any other guides for the tape are sufficiently far away so that it is appropriate to use a straight-line asymptote for the tape's geometry. After multiplying through Eq. (4) by $d\phi/ds$, a closed-form integration can be made to produce the following:

$$\left(\frac{d\phi}{ds}\right)^2 = \frac{2T}{D}[1-\cos(\Phi-\phi)]. \tag{7}$$

The boundary conditions of (6) have been used to determine the constant of integration arising in (7). Substitution of (5) into (7) gives the asymptotic angle of the tape:

$$\Phi = \arccos\left(1 - \frac{D}{2TR^2}\right). \tag{8}$$

Substituting this result into Eqs. (2) gives the force and moment reactions at the drum:

$$\frac{T_o}{T} = 1 - \frac{D}{2TR^2}, \quad \frac{V_o}{T} = \frac{D}{TR^2}\left(1 - \frac{D}{4TR^2}\right),$$
$$\frac{M_o}{TR} = \frac{R_{\text{eff}}}{R} = 1 + \frac{D}{2TR^2}. \tag{9}$$

Equation (7) along with the following two geometrical relationships,

$$\frac{dx}{ds} = \cos\phi(s), \quad \frac{dy}{ds} = \sin\phi(s), \tag{10}$$

can be numerically integrated to produce the geometry of the deflected tape. Alternatively, the arclength, s, can be swapped with the angle, ϕ, as the independent variable, thereby leading to the following elliptic integrals:

$$\frac{s}{R} = \frac{1}{R}\int_0^s ds = \frac{1}{R}\int_0^\phi \frac{ds}{d\phi}\,d\phi$$
$$= \left(\frac{D}{2TR^2}\right)^{1/2}\int_0^\phi \frac{d\phi}{[1-\cos(\Phi-\phi)]^{1/2}}, \tag{11}$$

$$\frac{x}{R} = \frac{1}{R}\int_0^s \cos[\phi(s)]ds = \frac{1}{R}\int_0^\phi (\cos\phi)\frac{ds}{d\phi}\,d\phi$$
$$= \left(\frac{D}{2TR^2}\right)^{1/2}\int_0^\phi \frac{(\cos\phi)d\phi}{[1-\cos(\Phi-\phi)]^{1/2}}, \tag{12}$$

$$\frac{y}{R} = \frac{1}{R}\int_0^s \sin[\phi(s)]ds = \frac{1}{R}\int_0^\phi (\sin\phi)\frac{ds}{d\phi}\,d\phi$$
$$= \left(\frac{D}{2TR^2}\right)^{1/2}\int_0^\phi \frac{(\sin\phi)d\phi}{[1-\cos(\Phi-\phi)]^{1/2}}. \tag{13}$$

The tape coordinates in the alternative frame are found with the help of Eq. (3) and trigonometric identities:

$$\frac{\bar{x}}{R} = \frac{1}{R}\int_0^s \cos[\Phi-\phi(s)]ds - \sin\Phi,$$
$$\frac{\bar{y}}{R} = -\frac{1}{R}\int_0^s \sin[\Phi-\phi(s)]ds - \cos\Phi. \tag{14}$$

The gap, h, between the tape and drum is measured radially from the drum center:

$$\frac{h}{R} = [\bar{x}^2 + \bar{y}^2]^{1/2} - 1. \tag{15}$$

The compliance of the tape in the "machine direction," C, is calculated by measuring how much a material particle changes its $\bar{x}$ coordinate in proportion to a change in the tension,

[4] Thus all loads will have conventional units *per unit width.*

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

T. To make this calculation it is first necessary to be able to locate the spatial location of a particular particle in the tape. It will be shown in the next section that the asymptotic angle, Φ, will always be less than or equal to 180 deg.[5] Referring to Fig. 1, it can thus be guaranteed that the tape will be in contact with the drum at location $\bar{x} = 0$, $\bar{y} = R$ (as well as at all other locations in the clockwise direction). This conveniently sets the location, or "reference point," of one known material particle. Then, because the tape is modeled as an inextensible beam, all other material particles can be located by measuring along the tape's arclength from the reference point. In particular, consider a second material particle that is at an arclength L from the reference point (moving counterclockwise along the tape in Fig. 1). Between the reference point at $\bar{x} = 0$, $\bar{y} = R$, and the separation point at $x = 0$, $y = 0$ the tape will lie flush with the drum. The arclength of this contact is

$$S_o = R(\pi - \Phi). \tag{16}$$

The remaining distance for which the tape is *not* in contact with the drum is, therefore

$$S = L - S_o. \tag{17}$$

This value is used as the upper limit in the integrals of Eq. (14) to locate the Cartesian coordinates, $\bar{X}$ and $\bar{Y}$, of the desired material particle,

$$\frac{\bar{X}}{R} = \frac{1}{R}\int_0^S \cos\left[\Phi - \phi(s)\right]ds - \sin\Phi,$$

$$\frac{\bar{Y}}{R} = -\frac{1}{R}\int_0^S \sin\left[\Phi - \phi(s)\right]ds - \cos\Phi. \tag{18}$$

This calculation can be repeated for varying amounts of tape tension, T, to get a functional dependence, $\bar{X}(T)$. The slope of this curve is the "machine direction" compliance, C, and the reciprocal of the compliance is the stiffness, K:

$$C = \frac{d\bar{X}}{dT}, \quad K = \left[\frac{d\bar{X}}{dT}\right]^{-1}. \tag{19}$$

A second compliance/stiffness pair is also of interest—that felt by the drum. If a distant material particle in the tape is held at a fixed $\bar{x}$ location, how much change is there in the angular orientation of the drum when there is a change in the moment M_o? Recall Fig. 1 and Eq. (9). Straightforward algebra leads to the following simple relationships between the machine direction compliance and stiffness, C and K, and the torsional compliance and stiffness, C_o and K_o:

$$C_o = C/R^2, \quad K_o = KR^2. \tag{20}$$

Recall Eq. (8). In many practical applications the quantity D/TR^2 will be small enough so that the asymptotic angle, Φ, will also be small. This allows Taylor series expansions through terms of Φ^3 to be employed to gain closed-form solutions to some of the key results of this section. Following are what might be thought of as "high tension" approximations—marked with an asterisk—for Eqs. (8), (7), (14), and (19):

$$\Phi^* \approx \left(\frac{D}{TR^2}\right)^{1/2}, \tag{21}$$

$$\phi^*(s) \approx \Phi^*\left[1 - \exp\left(\frac{-s}{R\Phi^*}\right)\right], \tag{22}$$

[5] Otherwise the tape unwinds completely from the drum.

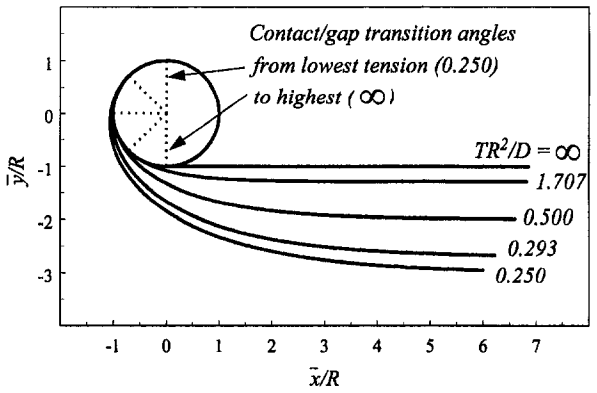


Fig. 2 Range of tape wrapping possibilities

$$\frac{\bar{x}^*}{R} \approx \left(\frac{s}{R} - \Phi^*\right) - \frac{\Phi^{*3}}{4}\left[1 - \exp\left(\frac{-2s}{R\Phi^*}\right)\right],$$

$$\frac{\bar{y}^*}{R} \approx -1 - \frac{\Phi^{*2}}{2}\left[1 - 2\exp\left(\frac{-s}{R\Phi^*}\right)\right], \tag{23}$$

$$\frac{C^*D}{R^3} \approx \frac{\Phi^{*5}}{8}, \quad \frac{K^*R^3}{D} \approx \frac{8}{\Phi^{*5}}. \tag{24}$$

Discussion of Results

Figure 2 shows the full range of tape wrapping possibilities as a function of the nondimensionalized tension,[6] TR^2/D. The figure is set in a Cartesian frame, $\bar{x}/R : \bar{y}/R$, that is nondimensionalized to the drum radius, and fixed, as in Fig. 1, so that the $\bar{x}$-axis is parallel to the asymptotic angle of the tape. To aid in interpretation, one may imagine that Fig. 2 shows tape winding under different tensions. The drum turns clockwise and the tape flows from right to left. Five tape contours are shown for the tensions:

$$\frac{TR^2}{D} = \frac{1}{4}, \frac{1}{2+\sqrt{2}}, \frac{1}{2}, \frac{1}{2-\sqrt{2}}, \infty, \tag{25}$$

as these give the following, respective angles according to Eq. (8):

$$\Phi = \pi, 3\pi/4, \pi/2, \pi/4, 0. \tag{26}$$

This angle has heretofore been called the "asymptotic angle," but from the perspective of the approaching tape, it might also be called the "entry angle," as the tape needs to bend through this much of an angle before making smooth, equilibrated contact with the drum.

Note that Eq. (8) does not have a real solution for $TR^2/D < \frac{1}{4}$. Physically, the tape would completely uncoil from the drum if the nondimensional tension dropped below this value. To explain; an entry angle of $\Phi = \pi$ puts the tape contact on the far side of the drum from the approaching tape. This gives the largest possible moment arm from the line-of-action of the asymptotic tension, T. If, for this maximum moment arm, there is an insufficient bending moment to produce the $1/R$ curvature needed to match the curvature of the drum, then the tape cannot wind onto the drum at any point.

Figure 2 can also be studied in regard to tape/drum spacing. Obviously, a longer and more gradual convergence between the tape and drum occurs when the bending rigidity of the tape is substantial. The entrance spacing between a tape and drum is

[6] Or, if the reader prefers, the square of the "nondimensional drum radius," or the reciprocal of the "nondimensional bending rigidity."

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

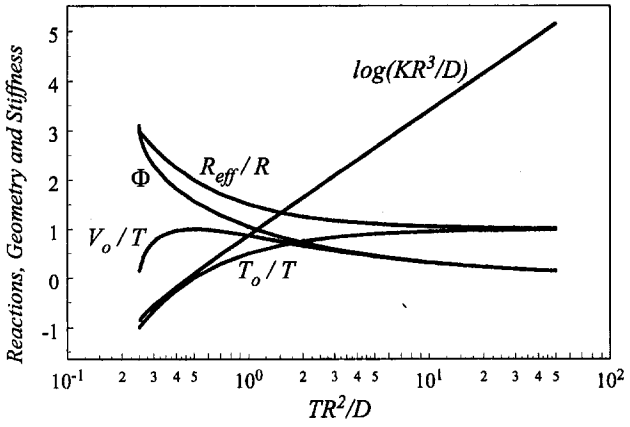


Fig. 3 Reactions, geometry, and stiffness

known to be important for air entrainment. This creates design advantages, such as when air is used as a lubricant in magnetic tape recording (Gross, 1980); and design disadvantages, such as when air entering into a wound roll causes uneven stacking of the layers (Keshavan and Wickert, 1997). It behooves the "foil bearing" designer to make a calculation of the angle Φ using Eq. (8) or (21), and if it is more than a few degrees to use the arc of the elastica curve in place of the straight-line approximation.

Turning again to Fig. 2, each of the tape profiles has been plotted for the same physical arclength. From the previously defined "reference point" at $\bar{x}/R = 0$ and $\bar{y}/R = 1$, an arclength of $L/R = 10$ was felt sufficient to bring the tape close to its horizontal asymptote, thereby leading to accurate numerical calculations for compliance and stiffness. Recall Eqs. (16)–(19). The curve for $TR^2/D = \infty$ shows what is typically—and to some degree erroneously—assumed for tape/drum contact. For infinite tension, the material point at $L/R = 10$ has the Cartesian coordinates: $\bar{X}/R = 10 - \pi$, $\bar{Y}/R = 0$. As the nondimensional tension is lowered, the tape bends through a greater entry angle and the material point at $L/R = 10$ moves leftward and downward. The leftward movement is what contributes to the machine direction stiffness, K (recall Eq. (19)), and the downward movement is what contributes to the effective radius, R_{eff} (recall Eq. (2)). Both of these effects will be discussed with Fig. 3. A table of results showing the movement of the material particle at $L/R = 10$ due to a tension change is given below.

TR^2/D	Φ	$\bar{X}/R$	$\bar{Y}/R$
0.250	180 deg	6.000	−2.946
0.293	135 deg	6.226	−2.674
0.500	90 deg	6.601	−1.994
1.707	45 deg	6.820	−1.293
∞	0 deg	6.858	−1.000

$L/R = 10$ for all cases.

Figure 3 is a nondimensionalized graph of some of the key variables of this problem. It is seen that all dependent variables can be expressed as a function of the nondimensional ratio, TR^2/D. As discussed for Fig. 2, the entry angle, Φ, decreases from π at $TR^2/D = \frac{1}{4}$ to 0 as $TR^2/D \to \infty$.

The tension in the tape, where the tape makes first contact with the drum, T_o, will be less than the asymptotic tension, T. It has been noted in a number of papers on wound roll stresses (Good et al., 1992; Benson, 1995) that a slight reduction in the "winding tension" can lead to a very large reduction in the radial stress in the wound roll. Although the reduction of T_o/T from unity may be quite small in most wound roll applications, it may nevertheless be important to account for this difference in order to have a good estimate of the interlayer pressure in the roll.

The interfacial reaction, V_o, where the tape makes first contact with the drum, also depends on the nondimensional parameter TR^2/D. For $TR^2/D \to \infty$ the drum experiences no reaction from the tape in the radial direction. The maximum value of $V_o/T = 1$ occurs when $TR^2/D = 0.5$. This is an obvious equilibrium requirement when one realizes that for $TR^2/D = 0.5$, the entry angle is $\Phi = \pi/2$, making V_o equal and opposite to T. As TR^2/D reduces further to the minimum value of $\frac{1}{4}$, the interfacial reaction between the tape and drum decreases to zero.

As $TR^2/D \to \infty$ the moment at the center of the drum (Recall Eq. (9) and Fig. 1) has the "expected" value of $M_o/TR = R_{\text{eff}}/R = 1$. However, for smaller values of TR^2/D the effective radius increases. Designers of tape handling equipment will need to account for this increased torque requirement—above the "expected" value—when winding a stiff tape under light tension onto a small diameter drum.

Figure 3 also shows results for the nondimensionalized machine direction stiffness,[7] KR^3/D. This is calculated numerically using Eq. (16)–(19). For values of $TR^2/D > \frac{1}{2}$ the graph for the closed-form approximation of Eq. (24) would be indistinguishable from the numerically calculated graph in Fig. 3. As TR^2/D grows large, KR^3/D grows by the 5/2 power. For many practical applications the machine direction stiffness will be great, but it will still be smaller than the stiffness associated with straining of the tape's midsurface. Further, even small amounts of compliance can have important design consequences in applications such as color, digital imaging, where pixel densities are often in the range of 10 to 20 per millimeter.

Summary

The manner in which an elastic tape winds onto a rigid drum has been studied, particularly in regard to the role of the tape's bending rigidity. The nondimensional ratio of TR^2/D has been shown to be the key parameter of the system. When the asymptotic tension is great, or the drum radius is large, or the tape's bending rigidity is small, then the tape will approach the drum nearly along a straight-line tangent to the drum as is often assumed in tape handing analyses. However, as TR^2/D becomes smaller, the tape bends further around the drum—albeit never more than 180 deg—before making an equilibrated contact. The reactions on the drum deviate from the "expected" values. The tension in the tape at the point of contact is lower, the transverse force at the point of contact is higher, and the torque on the drum is higher. The bend of the tape around the drum also introduces an elastic compliance between the drum and distant locations in the tape. The spacing between the tape and drum is more gradual when the equilibrated elastic arc is used instead of a straight-line tangent. When TR^2/D drops below $\frac{1}{4}$ the tape uncoils completely from the drum.

Acknowledgments

Funding received from the Eastman Kodak Company in support of the Mechanics of Flexible Structures Project (MFSP) at Penn State University is gratefully acknowledged. The author thanks his colleagues of the MFSP for many helpful conversations as this analysis progressed.

References

Benson, R. C., 1995, "A Nonlinear Wound Roll Model Allowing for Large Deformation," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 853–859.

[7] The logarithm of the stiffness is plotted in order to conveniently fit the graph into Fig. 3, and to show the power-law relationship between the stiffness and tension. Note, by Eq. (20), that this is also the nondimensional graph for the torsional, drum stiffness, $K_o R/D$.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Eshel, A., 1975, "The Slack in a Stretched Thin Foil Bent Against Two Parallel Cylinders," ASME *Journal of Engineering for Industry,* Vol. 97, pp. 1149–1152.

Fosdick, R., and Villaggio, P., 1986, "Influence of the Bending Stiffness on the Shape of a Belt in Steady Motion," ASME JOURNAL OF APPLIED MECHANICS, Vol. 53, pp. 266–270.

Frisch-Fay, R., 1962, *Flexible Bars,* Butterworth, Washington, DC.

Good, J. K., Pfeiffer, J. D., and Giachetto, R. M., 1992, "Losses in Wound-On Tension in Centerwinding of Wound Rolls," *Proceedings of the Web Handling Symposium,* ASME Applied Mechanics Division, AMD-Vol. 149, pp. 1–12.

Gross, W. A., et al., eds., 1980, *Fluid Film Lubrication,* John Wiley and Sons, New York.

Keshavan, M., and Wickert, J. A., 1997, "Air Entrainment During High Speed Web Spooling," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 916–922.

Mansfield, L., and Simmonds, J. G., 1987, "The Reverse Spaghetti Problem: Drooping Motion of an Elastica Issuing from a Horizontal Guide," ASME JOURNAL OF APPLIED MECHANICS, Vol. 54, pp. 147–150.

Müftü, S., and Cole, K. A., 1998, "The Fluid/Structure Interaction in Supporting a Thin Flexible Cylindrical Web with an Air Cushion," Internal Report, MIT Haystack Observatory, Westfield, MA.

Stolte, J., and Benson, R. C., 1992, "Dynamic Deflection of Paper Emerging From a Channel," ASME *Journal of Vibration and Acoustics,* Vol. 114, No. 2, pp. 187–193.

Stolte, J., and Benson, R. C., 1993, "An Extending Dynamic Elastica: Impact With a Surface," ASME *Journal of Vibration and Acoustics,* Vol. 115, No. 3, pp. 303–313.

Wang, C.-Y., 1981, "Winding a Long Elastic Sheet," *Acta Mechanica,* Vol. 39, pp. 297–301.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

B. Ravindra

P. Hagedorn

Institut fur Mechanik II,
Darmstadt University of Technology,
Darmstadt D-64289,
Germany

Invariants of Chaotic Attractor in a Nonlinearly Damped System

The characterization of a chaotic attractor in a driven, Duffing-Holmes oscillator with power-law damping is considered. State space reconstruction of the time series of the attractor is carried out to investigate its structure. The invariants associated with the attractor such as correlation dimension and entropy are computed. Also the maximum-likelihood (ML) estimation of dimension and entropy are carried out. The use of obtained invariants in building models for prediction and control using power-law dampers is discussed.

1 Introduction

Chaos in dissipative systems differs significantly from that in Hamiltonian systems. This is due to the fact that the concept of an attractor is meaningful only in the presence of dissipation. Also dissipation restricts the number of active degrees-of-freedom and makes the study of complicated systems through low-dimensional models feasible. Strictly dissipative nonlinear damping models (velocity to the pth power models) arise in diverse engineering and physical applications such as Josephson junctions, Vibration isolators, ship dynamics etc. It has been shown that the bifurcation structure and structure of the chaotic attractor are not altered by the damping index (Ravindra and Mallik, 1994). However, the critical parameter values at which bifurcations occur have been shown to depend both on the damping index and the damping coefficient. Hence power-law damping can be used as a passive control strategy. The emergence of electrorheological dampers whose damping properties can be tailored suitably for control purposes is another impetus for the present work. In this paper a detailed characterization of the chaotic attractor through the associated invariants such as dimensions and entropies is carried out. One of the objectives of this exercise is the possibility of using the computed invariants to build models for prediction and control of chaos using power-law damping. The procedure of state space reconstruction is the first step in the model building (Gershenfeld and Weigend, 1993; Abarbanel, 1994). Therefore, a comprehensive study of state space reconstruction and invariants of attractors in a nonlinearly damped system is undertaken in this work.

The statistical mechanical aspects of hyperbolic dynamical systems are founded on the theory of Sinai-Ruelle-Bowen measure (Ruelle, 1978). Recent work has addressed the issues such as the calculation of viscosities, diffusion coefficients, etc., and their relation to the Lyapunov exponents and entropies that characterise the chaotic dynamics of the system. While substantial progress has been made in the case of hyperbolic systems, the theory of nonhyperbolic systems is still being developed. The dynamics of even simple systems governing practical engineering situations turn out to be nonhyperbolic. Hence it is important to gain an insight into the behavior of invariants of attractors of nonhyperbolic systems and their relationship to dissipation. The theory of state space reconstruction proposed by Takens (Takens, 1981) and advanced by others (Sauer et al., 1991; Casdagli and Eubank, 1991) is a valuable tool in this endeavor.

2 System Description

Consider the Duffing-Holmes oscillator with nonlinear damping term as

$$x'' + 2\zeta x' |x'|^{p-1} - x + x^3 = f\cos(\omega t). \qquad (1)$$

Here ζ refers to the damping coefficient and p refers to the damping exponent. The case $p = 1$ is the usual viscous damping. Quadratic damping encountered in fluid drag forces is given by $p = 2$. In this paper apart from these two models, cubic damping $p = 3$ is also considered. For $p = 1$ this system has been studied extensively in the past as it governs the response of the first mode of a forced buckled beam (Holmes and Moon, 1983; Abhyankar et al., 1993). The above equation is integrated numerically using the usual Runge-Kutta method for selected parameter values ($\zeta = 0.125$, $\omega = 1.0$, $f = 0.4$). It is known (Ravindra and Mallik, 1994) that for this set of parameter values a chaotic attractor is obtained through the period-doubling route in all the three cases of damping models considered here ($p = 1$, 2, and 3). The time series data sets corresponding to the chaotic attractors of the above three damping models consisting of 20,000 points are obtained after ignoring the transients. Statistics of these data sets are shown in Table 1. In this table, the average cycle time is defined as the average time that is needed to complete a full cycle after the first passage through the average of the signal. Average absolute deviation (computed using the maximum norm) is an estimator of the width of time series data around the mean (Schouten et al., 1994a).

State space reconstruction of these data sets is carried out by the time delay embedding technique and the resulting phase plot of the reconstructed attractor for $p = 2$ is shown in Fig. 1. (The cases with $p = 1$ and $p = 3$ exhibit similar features). It may be noted that in obtaining Fig. 1, the data points are normalized with respect to the average absolute deviation.

3 State Space Reconstruction

In this method one starts with the time series of the observable measured from the experiment or obtained from numerical simulations. Takens (1981) proved that it is possible to imbed this time series in a reconstructed state space such that the asymptotic properties of the original dynamical system and that of the reconstructed dynamics are the same. One novel aspect of this approach is that one can obtain information about a dynamical system by observing the time evolution of just one variable. The central idea of this method involves in constructing time delay vectors by choosing appropriate delay vectors from the time series. Imbedding the dynamics in a reconstructed state

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Oct. 9, 1997; final revision, Mar. 7, 1998. Associate Technical Editor: N. C. Perkins.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 1

Feature	$p = 1$	$p = 2$	$p = 3$
Average	−0.00991	0.02142	0.00586
Average absolute deviation	0.76745	0.80076	0.79719
Number of points per cycle	138	172	179
Average cycle frequency	0.0390	0.0830	0.0895
Average cycle time	14.5	11.6	11.2

space requires a choice of three parameters: sampling time, embedding dimension, and time delay. The dimension of the reconstructed space is known as the imbedding dimension. The extension of Taken's work to the case of driven systems such as Eq. (1) has been carried out in the literature. A comprehensive review of the theory of state space reconstruction and the determination of associated invariant measures of the reconstructed attractor such as entropies and dimensions can be found in Eckmann and Ruelle (1985) and Kumar and Mallik (1995). In this work a partial review of the definitions of these invariants and various methods to compute them are presented with a view to bridge the gap between the researchers in dynamical systems and identification of vibrating systems. Then the results obtained in the case of three types of damping models are presented.

3.1 Mutual Information. Mutual information measure provides a guide line for the choice of time delay needed for state space reconstruction. Mutual information is based on the requirement of the independence of the scalar variables of the reconstructed space as compared to uncorrelatedness. The mutual information between two random variables y and z is a functional defined by their joint probability density,

$$I(y; z) = \langle \log_2 (p_{y,z}/p_y p_z) \rangle \tag{2}$$

where p_y and p_z are the probability density functions of y and z, respectively, $p_{y,z}$ is their joint probability density function and $\langle \cdot \rangle$ denotes the ensemble average. If mutual information is calculated for a range of values of time delay, it will typically decay for increasing but small values of time delay and then reach a base value due to external noise that is no longer dependent on time delay. The value of time delay at which the mutual information first reaches asymptotic value is the best choice. There exist several algorithms to compute mutual information. A histogram based approach is given in (Fraser, 1989). Another based on kernel density estimators is given in (Moon et al., 1995). In this work the method based on histograms is used. The average value of mutual information measure obtained for quadratic damping model (similar pattern is observed for $p = 1$ and $p = 3$) is plotted against delay time in Fig. 2.

4 Dimension Estimation

The attractor dimensions and entropies are defined in terms of the natural invariant probability measure of the dynamical system. It is in general not possible to obtain a theoretical expression for this invariant probability measure from the physical model of the dynamics. However, it is still possible to estimate the invariants directly from the time evolution by following state space reconstruction.

Dimension estimation gives an estimate of the necessary and sufficient number of state variables needed to approximate the steady-state dynamics from which the time series is realized. There exist a variety of dimensions to characterize the complexity of the chaotic attractor. The theory of generalized dimensions gives a unified approach to the various definitions of dimensions and there exists a notion of a multifractal based on the thermodynamic formalism. In this work we determine the correlation dimension and its variations such as ML estimation.

Let $\Pi = [\pi_1, \pi_2, \ldots, \pi_{M(r)}]$ denote a finite partition of resolution r of the attractor M and P_{π_i} be the normalized probability measure of the partition π_i. The correlation dimension D is given by

$$D = \lim_{r \to 0} \log \left(\sum_{i=1}^{M(r)} P_{\pi_i}^2 \right) / \log (r). \tag{3}$$

A direct method of estimation of D would involve the estimation of the probabilities in the above equation. This requires large data length of the time series which increases rapidly with the imbedding dimension. An alternative is the use of the generalized correlation sum.

4.1 Correlation Sum. Consider a time series x_i, $i = 1, \ldots, N'$ obtained from the time evolution of a dynamical system (such as obtained from Eq. (1)). A time delay reconstruction in R^n produces a vector time series X_i^n, $i = 1, \ldots N$, where $N = N' - n + 1$. For appropriate choice of n, the dynamical invariants evaluated from the original dynamical system and the reconstructed time series are the same. Hence one can deter-

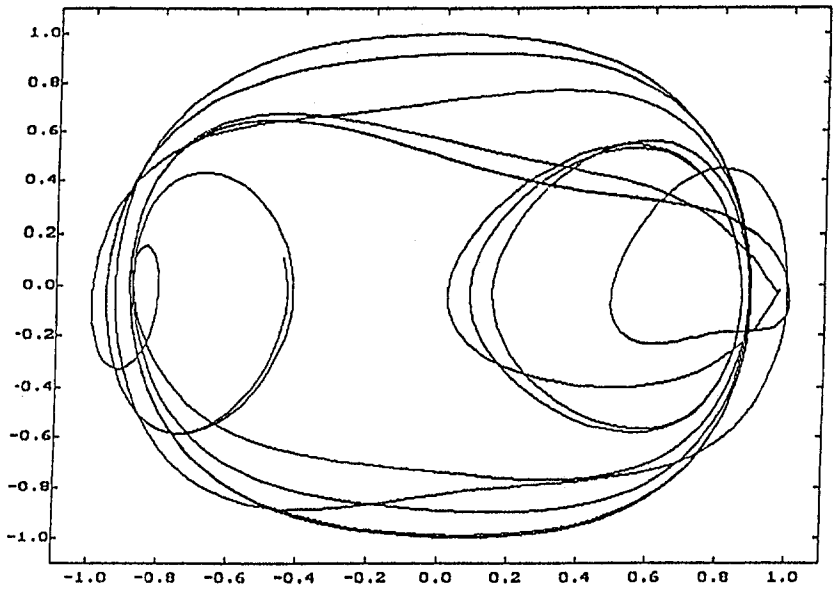


Fig. 1 Phase plots of reconstructed attractor, $p = 2$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

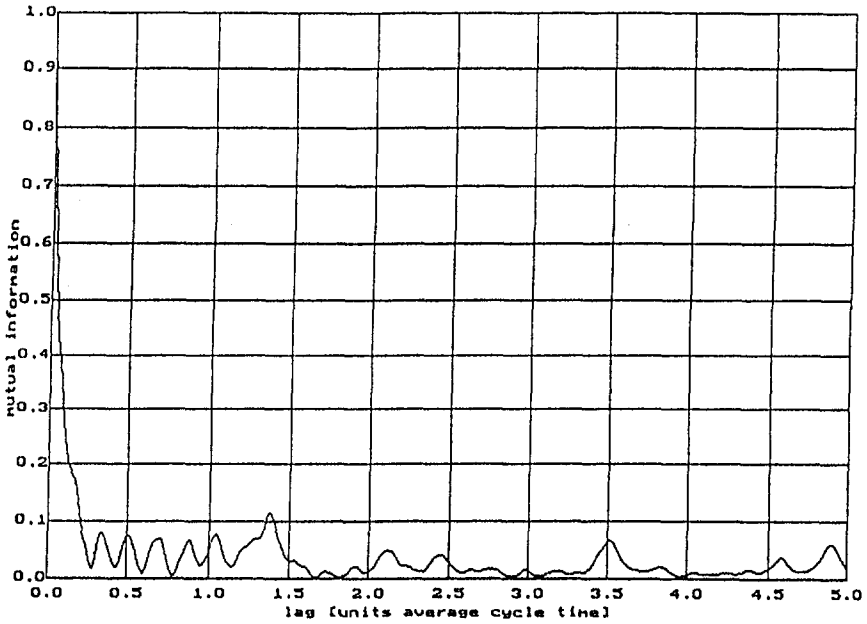


Fig. 2 Mutual information, $p = 2$

mine the dynamical invariants such as dimension and entropy from the reconstructed time series even when the physical model is not available. Consider a sphere of radius r around the point X_j^n on the reconstructed trajectory. The generalized correlation sum is defined as

$$C(r, n, N) = 1/[N(N-1)] \sum_{j=1}^{N} \sum_{i=1, i\neq j}^{N} \Theta(r - \|X_i^n - X_j^n\|), \quad (4)$$

where Θ is the Heaviside step function. The correlation dimension D can then be obtained as

$$D = \lim_{r\to 0} \log (C(r, n, N))/\log (r). \quad (5)$$

Thus the algorithm for computing D involves in plotting log $(C(r, n, N))$ versus log (r) for increasing values of embedding dimension n. A routine method is to carryout a linear least-square fit in the linear region of the plot and find the slope D. We omit the details of these computations and present a brief summary in Table 2 and Fig. 3. These computations are based on Schouten et al. (1994a). Estimation of dimension in the case of viscous damping has been considered previously (Ravindra and Cartmell, 1996). It can be seen that the chaotic attractor in the presence of quadratic damping has a lower value of dimension compared to the viscous and cubic damping models.

4.2 Maximum-Likelihood (ML) Estimation. There are several possible sources of errors in estimating the invariants of the attractor. It is important to have an estimate of the accuracy with which the invariant has been obtained. An alternative approach is to employ ML estimation of the invariant which requires the computation of the invariant only once. In this case one treats the correlation integral as a probability distribution. A ML estimator for correlation dimension (known as Taken's estimator) is given by

$$D_{ML} = [1/M \sum_{i=1}^{M} - \ln (r/r_0)]^{-1} \quad (6)$$

where M is the sample size of the interpoint distances r needed to compute the correlation integral. r_0 indicates the upper bound on the scaling distance r. Table 3 shows the results obtained for the Taken's estimator in the case of three types of damping.

4.3 Effect of Noise. The time series of a chaotic attractor obtained from an experiment is contaminated by noise. It is a normal practice to use some filtering strategy to eliminate the influence of noise. But it has been observed that filtering the time series in order to reduce the noise might increase or decrease the dimension of the attractor (Chennaoui et al., 1990; Badii et al., 1988). To account for the effect of noise, an approach to estimate correlation integral based on rescaled interpoint distances has been developed (Schouten et al., 1994a). This method also yields the noise level present in the time

Table 2

Feature	$p = 1$	$p = 2$	$p = 3$
Number of distances checked	1050087	408419	436591
Number of distances below scaling length	14257	11726	10822
Correlation Dimension (D)	2.37588	2.06786	2.32965
Deviation from Correlation integral	0.21008	0.02819	0.04014

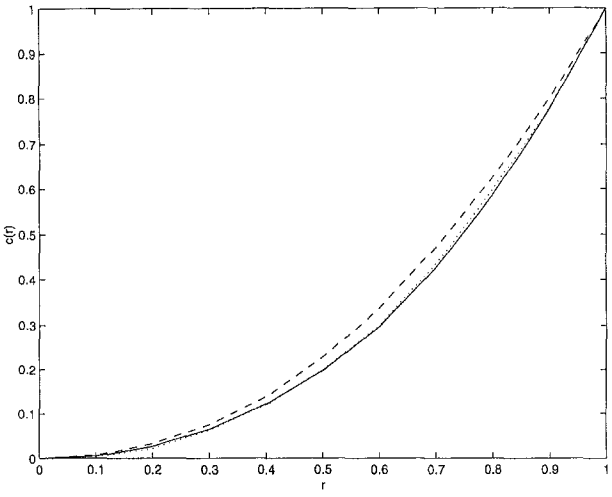


Fig. 3 Correlation integral, — $p = 1$; - - - $p = 2$; · · · $p = 3$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 3

Feature	$p = 1$	$p = 2$	$p = 3$
Number of distances checked	1050087	408419	436591
Number of distances below scaling length	14257	11726	10822
Correlation Dimension (D_{ML})	2.33659	2.13741	2.33281
Deviation from Correlation integral	0.31367	0.06985	0.04145
Relative standard error (percent)	0.83750	92347	0.96127

series. The results obtained in this case are presented in Fig. 4. The results follow a similar pattern observed in the case of correlation dimension estimation.

5 Kolmogorov-Sinai Entropy

The Kolmogorov-Sinai (KS) entropy of an attractor can be viewed as a measure for the rate of loss of information or as a measure of degree of predictability along the attractor given an initial point. A zero entropy refers to periodic, infinite entropy to stochastic and finite entropy to chaotic responses, respectively. The mathematical basis for KS entropy is based on theory of Markov partitions and Shannon's information theoretic entropy. The KS entropy can be considered as an dynamical systems analogue for Shannon's entropy. KS entropy is defined as

$$K = -\lim_{T\to 0}\lim_{r\to 0}\lim_{n\to\infty} 1/(nT)\log\Big(\sum_{i_1,\ldots,i_n} P^2_{i_1,\ldots,i_n}\Big) \qquad (7)$$

where T is the sampling time, r is the precision required to locate the system on a particular trajectory $i_1, \ldots, i_n$, and P's are the joint probabilities (Kumar and Mallik, 1995).

As in the case of correlation dimension, the numerical estimation of probabilities involved in the definition of KS entropy is difficult and an alternative approach based on generalized correlation sum is adopted. KS entropy (in the limit $n \to \infty$) is then given by

$$K = (1/T)\log\left(C(r, n, N)/C(r, n+1, N)\right) \qquad (8)$$

where $C(\cdot)$ is the correlation sum given by Eq. (4). KS entropy (also known as metric entropy) is inversely proportional to the average time over which the state of a dynamical system can be predicted. The predictability time T_p of the evolution of a system from a given initial state of precision r is given by

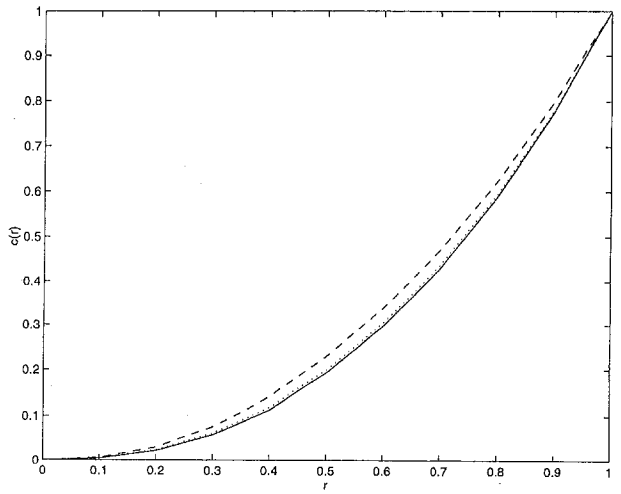


Fig. 4 Rescaled $C(r)$, — $p = 1$; - - - $p = 2$; · · · $p = 3$

Table 4

Feature	$p = 1$	$p = 2$	$p = 3$
Total number of distances checked	1081904	433053	440750
Number of distances found	14812	12505	10791
Average of b	86.32989	96.23016	66.11181
KS entropy K_{ML} in (bits/sec)	0.16809	0.15020	0.21821
Relative standard error (percent)	0.82167	0.89425	0.96266

$$T_p = (1/K)\log(1/r). \qquad (9)$$

Hence a lower value of entropy implies higher predictability.

The algorithm used here to compute entropy is based on correlation integral estimation. This algorithm is based on the assumption that the divergence of nearby points on different orbits of an attractor with time follows an exponential distribution,

$$C(b) = e^{-Kb\tau_s}, \qquad (10)$$

where $b = 1, 2, 3, \ldots$ and τ_s refers to the time step between two sampled data points of the time series. A ML estimate for KS entropy has been derived in literature (Schouten et al., 1994b). The distribution given by the above equation is seen to be valid in all the cases of damping considered in this work. The details of the computed results of ML estimator for entropy are summarized in Table 4. It is interesting to note that the value for quadratic damping ($p = 2$) is lower compared to the case of viscous ($p = 1$) or cubic ($p = 3$) damping cases. This implies that the predictability horizon is highest for quadratic damping among the three models studied here.

It may be noted that, though the values of the dimension and entropy depend also on the damping parameter ζ, the qualitative features presented here persist in the parameter range of interest.

6 Conclusions

A comprehensive characterization of chaotic attractor in Duffing-Holmes system with power-law damping is presented. The dimension and entropy computations carried out indicate that the attractor associated with quadratic damping has qualitatively distinct features as opposed to viscous and cubic damping. The results obtained imply that a quadratic damper has a low model order and high predictability and hence can be employed to quench chaos. It appears that to probe the fine structure of the traces of dissipation one needs to consider elaborate computations of generalized dimensions and Renyi entropies. This task requires intensive computational resources. An alternative paradigm for modeling chaotic responses of vibrating systems based on a black box approach (as opposed to the traditional methods relying on differential equation models (Masri et al., 1987; Wang and Fang, 1986)) is needed. Invariants of the attractor computed in this work provide valuable guidance in such a model building exercise: Mutual information measure guides in the selection of time delay required for state space reconstruction. Dimension estimation provides the number of active degrees-of-freedom and thus helps in selecting the suitable order of the model. Entropy gives a measure of predictability horizon and hence useful in arriving at an estimate of the frequency of the model updating required for online control strategies. Recent advances such as ML estimation of dimension and entropy and the role of noise are also considered in this work.

In the case of linear systems there exist standard techniques of model building in both time and frequency domain (Gersch and Liu, 1976; Wang and Fang, 1986). Also, weakly nonlinear systems can be modeled based on the nonlinear ARMA processes, Volterra-Wiener series, and other approximate methods (Tong, 1990). But the main feature of a nonhyperbolic system

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

is the possibility of multiple attractors in state space. This feature makes it difficult to obtain global models. However, one can still build local linear models (Farmer and Sidorowich, 1987) and then estimate the system parameters following the Bartlett theorem (Braun, 1986). This is the topic for further work.

Acknowledgments

The first author thanks the Alexander von Humboldt foundation for their support.

References

Abarbanel, H. D. I., et al., 1994, "Predicting Physical Variables in Time-Delay Embedding," *Physical Review E,* Vol. 49, pp. 1840–1853.

Abhyankar, N. S., Hall, E. K., and Hanagud, S. V., 1993, "Chaotic Vibrations of Beams: Numerical Solution of Partial Differential Equations," ASME JOURNAL OF APPLIED MECHANICS, Vol. 60, pp. 167–172.

Badii, R., et al., 1988, "Dimension Increase in Filtered Chaotic Signals," *Physical Review Letters,* Vol. 60, pp. 979–982.

Braun, S. G., 1986, *Mechanical Signature Analysis,* Academic Press, London.

Casdagli, M., and Eubank, S., 1991, *Nonlinear Prediction and Modelling,* Addison-Wesley, Reading, MA.

Chennaoui, A., et al., 1990, "Attractor Reconstruction From Filtered Chaotic Time Series," *Physical Review A,* Vol. 41, pp. 4151–4159.

Eckmann, J.-P., and Ruelle, D., 1985, "Ergodic Theory of Chaos and Strange Attractors," *Reviews of Modern Physics,* Vol. 57, pp. 617–656.

Farmer, J. D., and Sidorowich, J. J., 1987, *Physical Review Letters,* Vol. 59, pp. 845–848.

Fraser, A. M., 1989, "Information and Entropy in Strange Attractors," *IEEE transactions on Information theory,* Vol. 35, pp. 245–262.

Gersch, W., and Liu, R. S. Z., 1976, "Time Series Methods for the Synthesis of Random Vibration Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 43, pp. 159–165.

Gershenfeld, N. A., and Weigend, A. S., 1993, "The Future of Time Series," *Time Series Prediction: Forecasting the Future and Understanding the Past,* Addison-Wesley, Reading, MA, 1–70.

Holmes, P. J., and Moon, F. C., 1983, "Strange Attractors and Chaos in Nonlinear Mechanics," ASME JOURNAL OF APPLIED MECHANICS, Vol. 50, pp. 1021–1032.

Kumar, A., and Mullick, S. K., 1995, "A Dynamical Systems Approach to Nonlinear Time Series Analysis for Modelling and Prediction," preprint.

Masri, S. F., Miller, R. K., Saud, A. F., and Caughey, T. K., 1987, "Identification of Nonlinear Vibrating Structures: Part I—Formulation," ASME JOURNAL OF APPLIED MECHANICS, Vol. 54, pp. 918–922.

Moon, F. C., 1987, *Chaotic Vibrations,* John Wiley and Sons, New York.

Moon, Y., et al., 1995, "Estimation of Mutual Information Using Kernel Density Estimators," *Physical Review E,* Vol. 52, pp. 2318–2321.

Ravindra, B., and Mallik, A. K., 1994, "Role of Nonlinear Dissipation in Soft Duffing Oscillators," *Physical Review E,* Vol. 49, pp. 4950–4954.

Ravindra, B., and Cartmell, M., 1996, "Dynamical Systems Approach to Identification of Vibrating Systems," *International Conference on Identification in Engineering Systems,* M. I. Friswell, ed., Swansea, UK, pp. 570–583.

Ruelle, D., 1978, *Thermodynamic Formalism,* Addison-Wesley, Reading, MA.

Sauer, T., et al., 1991, "Embedology," *Journal of Statistical Physics,* Vol. 65, pp. 579–616.

Schouten, J. C., et al., 1994a, "Estimation of the Dimension of a Noisy Attractor," *Physical Review E,* Vol. 50, pp. 1851–1861.

Schouten, J. C., et al., 1994b, "Maximum-Likelihood Estimation of the Entropy of an Attractor," *Physical Review E,* Vol. 49, pp. 126–129.

Takens, F., 1981, "Detecting Strange Attractors in Turbulence," *Lecture Notes in Mathematics,* Vol. 898, Springer-Verlag, New York, pp. 366–381.

Tong, H., 1990, *Nonlinear Time Series: A Dynamical Systems Approach,* Oxford University Press, Oxford, UK.

Wang, Z., and Fang, T., 1986, "A Time Domain Method for Identifying Modal Parameters," ASME JOURNAL OF APPLIED MECHANICS, Vol. 53, pp. 28–32.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Solution of the Displacement Boundary Value Problem of an Interface Between Two Dissimilar Half-Planes and a Rigid Elliptic Inclusion at the Interface

V. Boniface
Visiting Postdoctoral Researcher

N. Hasebe
Professor

Department of Civil Engineering,
Nagoya Institute of Technology,
Gokiso-cho, Showa-ku,
Nagoya 466, Japan

The displacement boundary value problem of a bimaterial interface is solved using the complex stress function technique. A rational mapping function is used to map the two half-planes into unit circles and analysis is carried out in the mapped plane. The symmetric bimaterial problem is considered and the particular case of a rigid elliptic inclusion at the interface is solved. Uniform remote tensions both along and normal to the interface are considered. Stress distributions on the inclusion boundary are shown. Stress concentration factors at the inclusion tips are obtained and are expressed in terms of the radius of curvature using an approximate form of a general expression. These results are used to predict the likelihood of debonding/cracking at the tips. Also, stress concentration factors at the tips of an elliptic inclusion and elliptic void are compared. Stress intensity factors at the tips of a thin rigid elliptic inclusion are also determined.

Introduction

A composite system consists of two or more phases bonded together. Analysis of fracture in such a heterogeneous medium necessarily involves simplifying assumptions. The different models of composite fracture can be broadly classified under two categories—(a) continuum and (b) discrete. In the continuum approach, only gross properties of the composite are considered, i.e., properties of the individual phases are averaged. In the discrete approach, however, focus is on critical domains in a composite, usually the interface between two phases. Interface fracture mechanics deals with the study of interface crack nucleation, propagation, direction of propagation and arrest. Other important areas where interfaces are present include bonded structures, welded joints, explosively clad materials, and thin film coatings.

Earlier work in this field primarily dealt with understanding crack propagation in an interface. It was mainly concerned with stress fields and energy release rates associated with pre-existing cracks. Some of the salient work on interface cracks is found in Sih and Rice (1964), England (1965), Erdogan (1965), Rice and Sih (1965), Hayashi and Nemat-Nasser (1981), Rice (1988), Comninou (1990), and Shih and Asaro (1990). Currently a lot of research is focused on gaining an understanding of crack nucleation. In this direction, the effects of voids and inclusions on the stress field at an interface are of primary importance. Recently, Hasebe et al. (1992) and Okumura et al. (1995) modeled voids as elliptical holes at an interface and obtained stress distributions. Earlier work on a rigid inclusion at a bimaterial interface includes Ballarini (1990) and Asundi and Deng (1995). They considered a thin line-like inclusion at the interface. In this paper, the effect of a rigid inclusion (of arbitrary shape) on the stress field at a bimaterial interface is analyzed. The particular problem of a rigid elliptic inclusion at the interface is solved.

The displacement boundary value problem at a bimaterial interface, where the displacement boundary exists at the interface, is solved first. Then the solution for a bimaterial interface with a symmetric rigid inclusion is obtained. In this paper, the inclusion is assumed to be elliptic in shape. Two types of loading are considered—uniform remote tension (a) along and (b) normal to the interface, i.e., along the x and y-directions, respectively.

Applied Tension in the x-Direction

A planar interface formed by the perfect bonding of two dissimilar half-planes, each with a semi-elliptical notch, is considered. As shown in Fig. 1(a), material 1 is in the upper half-plane ($y \geq 0$) and material 2 is in the lower half-plane ($y \leq 0$). The two materials are assumed to be isotropic and homogeneous exhibiting linear-elastic behavior. Material 1 has elastic properties μ_1 and ν_1 while material 2 has μ_2 and ν_2 (μ_j is the shear modulus and ν_j is the Poisson's ratio where $j = 1$ and 2 correspond to materials 1 and 2, respectively). A rigid elliptic inclusion is located symmetrically with respect to the bimaterial interface. It extends to a depth of a in each material and spans a width of $2b$ along the interface. Uniform remote tension is applied along the x-direction. Debonding of the interface at infinity due to infinite displacements is avoided by relating applied tensions in the two material half-planes by a factor γ_x.

Materials 1 and 2 are considered separately in the z_1 ($=x_1 + iy_1$) and z_2 ($=x_2 + iy_2$) planes, respectively (where $i^2 = -1$), as shown in Figs. 1(b) and (c). The z_1-plane is obtained by rotation of the z ($=x + iy$) plane through π about the x-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Sept. 25, 1997; final revision, Mar. 1, 1998. Associate Technical Editor: J. R. Barber.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

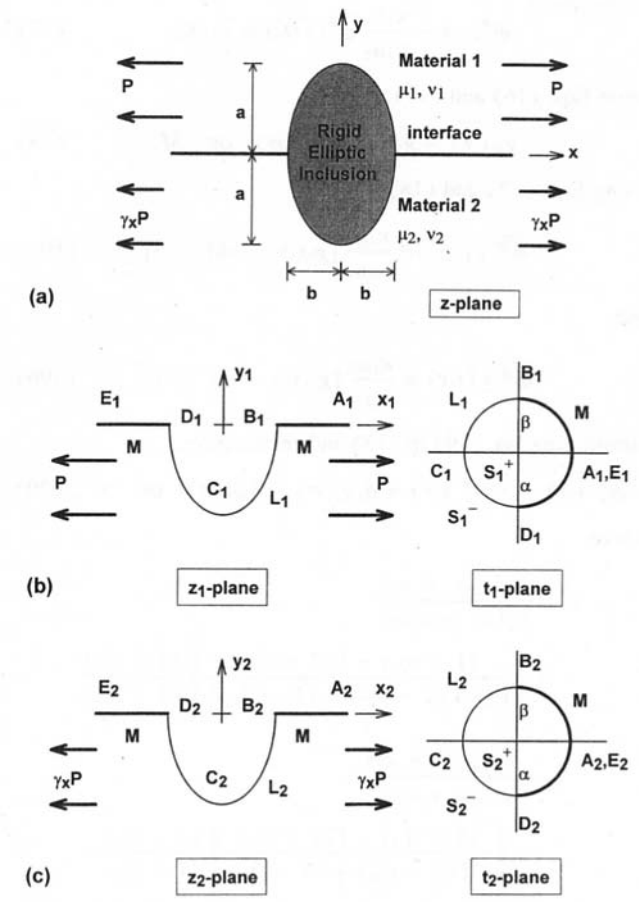


Fig. 1 *(a)* **Rigid elliptic inclusion in a bimaterial interface with remote tension in the *x*-direction, *(b)* and *(c)* mapping of half-planes with semi-ellipses into unit circles**

axis while the z_2-plane coincides with the z-plane. Since the inclusion geometry is symmetric with respect to the interface (x-axis), the shapes in the z_1 and z_2-planes are identical. Therefore, the same mapping function is used to map the half-planes with semi-ellipses in the z_1 and z_2-planes into unit circles in the t_1 and t_2-planes, respectively. In Fig. 1, M denotes the bimaterial interface and L_j the displacement (inclusion) boundary. On the boundary of the unit circle ($M + L_j$), t_j is denoted by σ ($=e^{i\theta}$) and so $\bar{\sigma} = 1/\sigma$. Also, the boundary divides the t_j-plane into two domains, S_j^+ and S_j^- as shown in the figures.

External forces, stresses, and displacements in the z_1 and z_2-planes are related to those in the z-plane (physical plane) as

$$(P_x)_1 = P_{x1} \quad (P_y)_1 = -P_{y1} \quad (P_x)_2 = P_{x2} \quad (P_y)_2 = P_{y2}$$
$$(u_x)_1 = u_{x1} \quad (u_y)_1 = -u_{y1} \quad (u_x)_2 = u_{x2} \quad (u_y)_2 = u_{y2}$$
$$(\sigma_x)_1 = \sigma_{x1} \quad (\sigma_y)_1 = \sigma_{y1} \quad (\tau_{xy})_1 = -\tau_{x1y1}$$
$$(\sigma_x)_2 = \sigma_{x2} \quad (\sigma_y)_2 = \sigma_{y2} \quad (\tau_{xy})_2 = \tau_{x2y2} \qquad (1)$$

where $(\cdot)_j\, j = 1, 2$ denotes the value of the bracketed quantity in the *physical plane* corresponding to material j. A mapping function which maps a half-plane with a semi-ellipse into a unit circle is given as (Hasebe, 1972; Salama and Hasebe, 1996)

$$z_j = -ia\frac{\sqrt{1+t_j^2}}{1-t_j} - ib\frac{1+t_j}{1-t_j} \quad (j = 1, 2).$$

A rational mapping function is formulated from the above equation as

$$z_j = \frac{E_o}{1-t_j} + \sum_{k=1}^{N}\frac{E_k}{\zeta_k - t_j} + E_c \equiv \omega(t_j) \quad (j = 1, 2) \qquad (2)$$

where E_o, E_k, E_c, and ζ_k are complex constants, and $N = 28$ in this analysis. Details of this procedure are described in Hasebe and Inohara (1980) and Salama and Hasebe (1996). For bimaterial problems, it is convenient to use Dundurs' constants given by

$$\alpha_D = \frac{(\kappa_1+1)\Gamma - (\kappa_2+1)}{(\kappa_1+1)\Gamma + (\kappa_2+1)} \quad \text{and}$$
$$\beta_D = \frac{(\kappa_1-1)\Gamma - (\kappa_2-1)}{(\kappa_1+1)\Gamma + (\kappa_2+1)} \qquad (3)$$

where

$$\kappa_j = \begin{cases}(3-4\nu_j)\\ (3-\nu_j)/(1+\nu_j)\end{cases} \quad (j = 1, 2)$$

for plane strain and plane stress, respectively, and

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

The required stress functions $\phi_j(t_j)$ and $\psi_j(t_j)$, which are regular in the unit circles, are expressed in terms of two components as

$$\phi_j(t_j) = \phi_j^A(t_j) + \phi_j^B(t_j) \qquad (4a)$$

and

$$\psi_j(t_j) = \psi_j^A(t_j) + \psi_j^B(t_j) \qquad (4b)$$

where superscript A refers to uniform remote tension. These components are obtained as

$$\phi_1^A(t_1) = \frac{P}{4}\omega(t_1) \quad \psi_1^A(t_1) = -\frac{P}{2}\omega(t_1)$$
$$\phi_2^A(t_2) = \frac{\gamma_x P}{4}\omega(t_2) \quad \psi_2^A(t_2) = -\frac{\gamma_x P}{2}\omega(t_2) \qquad (5)$$

$\phi_j^B(t_j)$ and $\psi_j^B(t_j)$ are complementary functions.

If P is the uniform remote tension acting on material 1, $\gamma_x P$ must be applied on material 2 to ensure continuity of strains in the x-direction, i.e., $(\epsilon_x)_1 = (\epsilon_x)_2$, across the interface (Hasebe et al., 1992). This is necessary to avoid infinite displacements at infinity. Therefore, γ_x is given by

$$\gamma_x = \frac{\mu_2(\kappa_1+1)}{\mu_1(\kappa_2+1)} = \frac{1+\alpha_D}{1-\alpha_D}. \qquad (6)$$

Since the elliptic inclusion is rigid, the displacement boundary condition is written as $(u_x)_j = (u_y)_j = 0$ on L_j $(j = 1, 2)$. Thus the functions $\psi_j(t_j)$ are given by analytic continuation at the rigid boundary as (Muskhelishvili, 1963)

$$\psi_j(t_j) = \kappa_j\overline{\phi_j}(1/t_j) - \frac{\overline{\omega}(1/t_j)}{\omega'(t_j)}\phi_j'(t_j). \qquad (7)$$

Prescribed loads and displacements on the boundary are expressed using Eq. (7) as

$$\phi_j(\sigma) + \frac{\omega(\sigma)}{\overline{\omega'(\sigma)}}\overline{\phi_j'(\sigma)} + \overline{\psi_j(\sigma)}$$
$$= \phi_j^+(\sigma) + \kappa_j\phi_j^-(\sigma) = i\int (P_{xj} + iP_{yj})ds \qquad (8)$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

and

$$\frac{1}{\mu_j}\left[\kappa_j\phi_j(\sigma) - \frac{\omega(\sigma)}{\omega'(\sigma)}\overline{\phi_j'(\sigma)} - \overline{\psi_j(\sigma)}\right]$$

$$= \frac{\kappa_j}{\mu_j}[\phi_j^+(\sigma) - \phi_j^-(\sigma)] = 2(u_{xj} + iu_{yj}). \quad (9)$$

Since the inclusion is rigid, displacements are zero on L_j and Eq. (9) gives

$$\phi_j^+(\sigma) - \phi_j^-(\sigma) = 0 \quad \text{on} \quad L_j.$$

Substituting Eqs. (4) and (5) in the above equation

$$\phi_j^{B+}(\sigma) - \phi_j^{B-}(\sigma) = 0 \quad \text{on} \quad L_j \quad (10)$$

Traction continuity across M in the z-plane is written as

$$i\int((P_x)_1 + i(P_y)_1)ds = i\int((P_x)_2 + i(P_y)_2)ds \quad \text{on} \quad M$$

Using relations (1) this equation is expressed in the z_1 and z_2-planes as

$$i\int(P_{x1} + iP_{y1})ds = \overline{i\int(P_{x2} + iP_{y2})ds} \quad \text{on} \quad M$$

Substituting from Eq. (8), the above equation becomes

$$\phi_1^+(\sigma) + \kappa_1\phi_1^-(\sigma) = \overline{\phi_2^+}(1/\sigma) + \kappa_2\overline{\phi_2^-}(1/\sigma) \quad \text{on} \quad M. \quad (11)$$

Similarly, displacement continuity across M is given by

$$\frac{\kappa_1}{\mu_1}[\phi_1^+(\sigma) - \phi_1^-(\sigma)]$$

$$= \frac{\kappa_2}{\mu_2}[\overline{\phi_2^+}(1/\sigma) - \overline{\phi_2^-}(1/\sigma)] \quad \text{on} \quad M. \quad (12)$$

Equations (11) and (12) represent the boundary conditions on M. Substituting Eqs. (4) and (5) in (11) and (12),

$$\phi_1^{B+}(\sigma) + \kappa_1\phi_1^{B-}(\sigma)$$
$$= \overline{\phi_2^{B+}}(1/\sigma) + \kappa_2\overline{\phi_2^{B-}}(1/\sigma) + D_1(\sigma) \quad \text{on} \quad M \quad (13)$$

and

$$\frac{\kappa_1}{\mu_1}[\phi_1^{B+}(\sigma) - \phi_1^{B-}(\sigma)]$$

$$= \frac{\kappa_2}{\mu_2}[\overline{\phi_2^{B+}}(1/\sigma) - \overline{\phi_2^{B-}}(1/\sigma)] \quad \text{on} \quad M \quad (14)$$

where

$$D_1(\sigma) = \frac{P}{4}[(1 + \kappa_2)\gamma_x - (1 + \kappa_1)]\omega(\sigma). \quad (15)$$

Rewriting Eq. (14),

$$\left[\phi_1^{B+}(\sigma) + \frac{\kappa_2\mu_1}{\kappa_1\mu_2}\overline{\phi_2^{B-}}(1/\sigma)\right]$$

$$-\left[\phi_1^{B-}(\sigma) + \frac{\kappa_2\mu_1}{\kappa_1\mu_2}\overline{\phi_2^{B+}}(1/\sigma)\right] = 0 \quad \text{on} \quad M. \quad (16)$$

Equation (16) represents a Riemann-Hilbert problem. The general solution of Eq. (16) is an arbitrary rational function $g_1(t_1)$ (Muskhelishvili, 1963) which is determined later in the analysis. Therefore,

$$\phi_1^B(t_1) + \frac{\kappa_2\mu_1}{\kappa_1\mu_2}\overline{\phi_2^B}(1/t_1) = g_1(t_1). \quad (17)$$

From Eqs. (16) and (17),

$$g_1^+(\sigma) = g_1^-(\sigma) = g_1(\sigma) \quad \text{on} \quad M. \quad (18)$$

Using Eqs. (17) and (18),

$$\overline{\phi_2^{B-}}(1/\sigma) = \frac{\kappa_1\mu_2}{\kappa_2\mu_1}[g_1(\sigma) - \phi_1^{B+}(\sigma)] \quad (19a)$$

and

$$\overline{\phi_2^{B+}}(1/\sigma) = \frac{\kappa_1\mu_2}{\kappa_2\mu_1}[g_1(\sigma) - \phi_1^{B-}(\sigma)]. \quad (19b)$$

Substituting Eq. (19) in (13) and rearranging,

$$\phi_1^{B+}(\sigma) + \lambda_1\phi_1^{B-}(\sigma) = \delta_1 g_1(\sigma) + h_{M1}(\sigma) \quad \text{on} \quad M \quad (20)$$

where

$$\lambda_1 = \frac{\kappa_1(\kappa_2\mu_1 + \mu_2)}{\kappa_2(\mu_1 + \kappa_1\mu_2)}$$
$$= \left\{\frac{(1 + \alpha_D) - \Gamma(1 + \beta_D)}{\Gamma[-\Gamma(1 - \alpha_D) + (1 - \beta_D)]}\right\}\frac{(1 - \beta_D)}{(1 + \beta_D)}$$

$$\delta_1 = \frac{\kappa_1\mu_2(1 + \kappa_2)}{\kappa_2(\mu_1 + \kappa_1\mu_2)}$$
$$= \left\{\frac{(1 + \alpha_D) - \Gamma(1 + \beta_D)}{-\Gamma(1 - \alpha_D) + (1 - \beta_D)}\right\}\frac{(1 - \alpha_D)}{(1 + \beta_D)}$$

$$h_{M1}(\sigma) = \frac{\mu_1}{\mu_1 + \kappa_1\mu_2}D_1(\sigma)$$
$$= \frac{(\alpha_D - \beta_D)}{(\Gamma - 1)(1 + \beta_D)}D_1(\sigma). \quad (21)$$

From the above equations, it is seen that (unlike the void problem (Hasebe et al., 1992; Okumura et al., 1995)), Dundurs' constants are not sufficient to define the problem. An additional material parameter (e.g., moduli ratio Γ) is also necessary.

Using Eqs. (10) and (20), the Riemann-Hilbert problem for $\phi_1^B(t_1)$ is written as

$$\phi_1^{B+}(\sigma) - \phi_1^{B-}(\sigma) = 0 \quad \text{on} \quad L_1$$

and

$$\phi_1^{B+}(\sigma) + \lambda_1\phi_1^{B-}(\sigma) = \delta_1 g_1(\sigma) + h_{M1}(\sigma) \quad \text{on} \quad M. \quad (22)$$

Proceeding on similar lines, the Riemann-Hilbert problem for $\phi_2^B(t_2)$ is

$$\phi_2^{B+}(\sigma) - \phi_2^{B-}(\sigma) = 0 \quad \text{on} \quad L_2$$

and

$$\phi_2^{B+}(\sigma) + \lambda_2\phi_2^{B-}(\sigma) = \delta_2 g_2(\sigma) + h_{M2}(\sigma) \quad \text{on} \quad M \quad (23)$$

where λ_2, δ_2, and $h_{M2}(\sigma)$ are obtained by interchanging subscripts 1 and 2 in Eq. (21) and

$$D_2(\sigma) = -D_1(\sigma). \quad (24)$$

$g_2(\sigma)$ is an arbitrary rational function which is determined later.

Solution of the Riemann-Hilbert Problem. The unknown function $g_j(t_j)$ in Eqs. (17) and (23) is assumed as a sum of irregular terms inside and outside the unit circle given by

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$g_j(t_j) = \sum_n \frac{a_{nj}}{\xi_{nj} - t_j} + \sum_n \frac{b_{nj}}{\eta_{nj} - t_j} + \frac{d_j}{1 - t_j} \tag{25}$$

where ξ_{nj}, η_{nj}, a_{nj}, b_{nj}, and d_j are complex constants to be evaluated, and $|\xi_{nj}| > 1$ and $|\eta_{nj}| < 1$ are assumed. There is only one pole on the unit circle at $t_j = 1$. Equation (25) does not contain higher-order terms since coefficients of poles of order greater than 1 finally become zero (see Eq. (38)).

A solution of the Riemann-Hilbert problem in Eq. (22) is obtained as

$$\phi_1^B(t_1) = H_1(t_1) + \frac{\delta_1 \chi_1(t_1)}{2\pi i} \int_M \frac{g_1(\sigma) d\sigma}{\chi_1^+(\sigma)(\sigma - t_1)} + \chi_1(t_1) P_1(t_1) \tag{26}$$

where $P_1(t_1)$ is an arbitrary rational function and $\chi_1(t_1)$ is the Plemelj function given by

$$\chi_1(t_1) = (t_1 - \alpha)^{m_1} (t_1 - \beta)^{1 - m_1} \tag{27}$$

with

$$m_1 = \frac{1}{2} + i \ln\left(\frac{\lambda_1}{2\pi}\right). \tag{28}$$

α and β correspond to the junctions between M and L_1 as shown in Fig. 1(b). Since $\chi_1(t_1)$ is a multivalued function, the branch $\chi_1(t_1)/t_1 \to 1$ for $t_1 \to \infty$ is chosen. $\chi_1(t_1)$ on the boundary satisfies

$$\chi_1^+(\sigma) = -\lambda_1 \chi_1^-(\sigma) \quad \text{on} \quad M \tag{29a}$$

and

$$\chi_1^+(\sigma) = \chi_1^-(\sigma) \quad \text{on} \quad L_1. \tag{29b}$$

In Eq. (26), the first term on the right-hand side is transformed to a contour integral around the boundary M as

$$H_1(t_1) = \frac{\chi_1(t_1)}{2\pi i} \int_M \frac{h_{M1}(\sigma) d\sigma}{\chi_1^+(\sigma)(\sigma - t_1)} = \frac{\chi_1(t_1)}{(1 + \lambda_1) 2\pi i} \oint_M \frac{h_{M1}(\sigma) d\sigma}{\chi_1(\sigma)(\sigma - t_1)}.$$

Substituting from Eqs. (2), (15), and (21), the above expression is evaluated using the residue theorem. The pole $t_1 = 1$ lies on M and so residue at this pole is obtained differently. Thus,

$$H_1(t_1) = C_1 \left\{ \left[1 + \frac{(1 - \lambda_1)}{2} \frac{\chi_1(t_1)}{\chi_1(1)} \right] \frac{E_o}{1 - t_1} + \sum_k \left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta_k)} \right] \frac{E_k}{\zeta_k - t_1} \right\} \tag{30}$$

where

$$C_1 = \frac{P}{4} \frac{\mu_1}{(\mu_1 + \kappa_1 \mu_2)} \left[\frac{(1 + \kappa_2)\gamma_x - (1 + \kappa_1)}{1 + \lambda_1} \right] = \frac{P}{4} \left\{ \frac{(1 + \alpha_D)(\Gamma - 1)[\Gamma(1 - \alpha_D) - (1 - \beta_D)]}{\Gamma^2(1 - \alpha_D)(1 + \beta_D) - (1 + \alpha_D)(1 - \beta_D)} \right\}. \tag{31}$$

Similarly the second term in Eq. (26) is evaluated as

$$\frac{\delta_1 \chi_1(t_1)}{2\pi i} \int_M \frac{g_1(\sigma) d\sigma}{\chi_1^+(\sigma)(\sigma - t_1)} = \frac{\delta_1}{(1 + \lambda_1)} \left\{ \sum_n \left[1 - \frac{\chi_1(t_1)}{\chi_1(\xi_{n1})} \right] \frac{a_{n1}}{\xi_{n1} - t_1} + \sum_n \left[1 - \frac{\chi_1(t_1)}{\chi_1(\eta_{n1})} \right] \frac{b_{n1}}{\eta_{n1} - t_1} + \left[1 + \frac{(1 - \lambda_1)}{2} \frac{\chi_1(t_1)}{\chi_1(1)} \right] \frac{d_1}{1 - t_1} \right. \tag{32}$$

Next, $P_1(t_1)$ is determined using the regular characteristics of $\psi_1^B(t_1)$. Substituting Eqs. (4) and (5) in (7),

$$\psi_1^B(t_1) = \kappa_1 \overline{\phi_1^B}(1/t_1) - \frac{\overline{\omega}(1/t_1)}{\omega'(t_1)} \phi_1^{B'}(t_1) + \frac{P}{4} [(\kappa_1 - 1)\overline{\omega}(1/t_1) + 2\omega(t_1)]. \tag{33}$$

The second term on the right-hand side of Eq. (33) is expanded in a Laurent's series around $t_1 = 1/\overline{\zeta_k} \equiv \zeta_k'$ as

$$\frac{\overline{\omega}(1/t_1)}{\omega'(t_1)} \phi_1^{B'}(t_1) = -\sum_k \frac{A_{k1} \overline{B_k} \zeta_k'^2}{\zeta_k' - t_1} + \text{Regular terms in } S_1^+ \tag{34}$$

where

$$A_{k1} \equiv \phi_1^B(\zeta_k') \quad \text{and} \quad B_k \equiv E_k / \overline{\omega'(\zeta_k')}. \tag{35}$$

The first term on the right-hand side of Eq. (33) is written using Eqs. (26), (30), and (32), and the second and third terms using Eqs. (34) and (2), respectively. The right-hand side includes irregular terms in S_1^+ which must be canceled because $\psi_1^B(t_1)$ is regular in S_1^+. Therefore, $P_1(t_1)$ is determined as

$$P_1(t_1) = \sum_k \frac{1}{\kappa_1} \left[\overline{A_{k1}} B_k - \frac{P}{4}(\kappa_1 - 1) E_k \right] \frac{1}{\chi_1(\zeta_k)(\zeta_k - t_1)} + \left\{ \frac{P}{4} \lambda_1 \frac{(\kappa_1 + 1)}{\kappa_1} E_o + \left[C_1 E_o + \frac{\delta_1}{1 + \lambda_1} d_1 \right] \times \frac{(3\lambda_1 - 1)}{2} \right\} \frac{1}{\chi_1(1)(1 - t_1)}. \tag{36}$$

Proceeding on similar lines for material 2, the solution of the Riemann-Hilbert problem in Eq. (23) is obtained by interchanging subscripts 1 and 2, and P and $\gamma_x P$ in Eqs. (26) to (30) and (32) to (36) where

$$C_2 = \frac{P}{4} \frac{\mu_2}{(\mu_2 + \mu_1 \kappa_2)} \frac{[(1 + \kappa_1) - (1 + \kappa_2)\gamma_x]}{1 + \lambda_2} = \frac{P}{4} \left\{ \frac{(1 + \alpha_D)(\Gamma - 1)[(1 + \alpha_D) - \Gamma(1 + \beta_D)]}{\Gamma^2(1 - \alpha_D)(1 + \beta_D) - (1 + \alpha_D)(1 - \beta_D)} \right\}. \tag{37}$$

Next, the coefficients of $g_1(t_1)$ are obtained by substituting expressions for $\phi_1^B(t_1)$ and $\phi_2^B(1/t_1)$ and Eqs. (25) in (17). Noting that the poles must be the same on both sides of the equation and the coefficients at each pole must be equal, coefficients of Eq. (25) are determined. Therefore, $g_1(t_1)$ is expressed as

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$g_1(t_1) = \sum_k \frac{1}{\kappa_1}\left[\overline{A_{k1}}B_k - \frac{P}{4}(\kappa_1 - 1)E_k\right]\frac{1}{\zeta_k - t_1}$$

$$- \sum_k \frac{\mu_1}{\kappa_1\mu_2}\left[A_{k2}\overline{B_k} - \frac{\gamma_x P}{4}(\kappa_2 - 1)\overline{E_k}\right]\frac{\zeta_k'^2}{\zeta_k' - t_1}$$

$$- \frac{P}{4}\frac{(1 + \kappa_1)}{\kappa_1}\frac{E_o}{1 - t_1}. \quad (38)$$

Substituting Eqs. (30), (32), (36), and (38) in (26), $\phi_1^B(t_1)$ is obtained as

$$\phi_1^B(t_1) = -\frac{P}{4}\frac{(\kappa_1 + 1)}{\kappa_1}\frac{\lambda_1}{(1 + \lambda_1)}\left[1 - \frac{\chi_1(t_1)}{\chi_1(1)}\right]\frac{E_o}{1 - t_1}$$

$$+ \frac{P}{4}\frac{(\mu_2 - \mu_1)}{(\mu_1 + \kappa_1\mu_2)}\frac{(\kappa_1 + 1)}{(1 + \lambda_1)}\sum_k\left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta_k)}\right]\frac{E_k}{\zeta_k - t_1}$$

$$+ \frac{\delta_1}{\kappa_1(1 + \lambda_1)}\sum_k\left[1 + \frac{(1 + \lambda_1 - \delta_1)}{\delta_1}\frac{\chi_1(t_1)}{\chi_1(\zeta_k)}\right]$$

$$\times\left[\overline{A_{k1}}B_k - \frac{P}{4}(\kappa_1 - 1)E_k\right]\frac{1}{\zeta_k - t_1}$$

$$- \frac{\mu_1\delta_1}{\kappa_1\mu_2(1 + \lambda_1)}\sum_k\left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta_k')}\right]$$

$$\times\left[A_{k2}\overline{B_k} - \frac{\gamma_x P}{4}(\kappa_2 - 1)\overline{E_k}\right]\frac{\zeta_k'^2}{\zeta_k' - t_1}. \quad (39)$$

Similarly, $g_2(t_2)$ and $\phi_2^B(t_2)$ are obtained by interchanging subscripts 1 and 2, and P and $\gamma_x P$ in Eqs. (38) and (39).

As mentioned earlier, A_{k2} is defined by interchanging subscripts 1 and 2 in Eq. (35). The complex constants A_{k1} and A_{k2} in the above equation are determined using these definitions for $k = 1, \ldots, N$ and solving the resulting system of $4N$ simultaneous equations with respect to the real and imaginary parts of A_{k1} and A_{k2}.

The solution for a thin rigid inclusion at a bimaterial interface can be obtained from the above analysis. In this case, the mapping function is expressed by $E_k = 0$ ($k = 1, 2, \ldots N$) in Eq. (2), which corresponds to a half-plane. Therefore, $\phi_1^A(t_1)$ and $\phi_1^B(t_1)$ are given by the first terms in the right-hand side of Eqs. (5) and (39), respectively.

Applied Tension in the *y*-Direction

The geometry is identical to the previous case with a rigid elliptic inclusion located symmetrically at a bimaterial interface as shown in Fig. 2. Here, however, uniform remote tension is applied along the y-direction. As before, a remote tension $\gamma_y P$ is applied on material 2 in the x-direction to avoid infinite displacements on the interface at infinity. The analysis here is similar to that in the earlier section.

The components of the stress functions corresponding to remote tension are obtained as

$$\phi_1^A(t_1) = \frac{P}{4}\omega(t_1),\ \psi_1^A(t_1) = \frac{P}{2}\omega(t_1)$$

$$\phi_2^A(t_2) = \frac{P(1 + \gamma_y)}{4}\omega(t_2),\ \psi_2^A(t_2) = \frac{P(1 - \gamma_y)}{4}\omega(t_2). \quad (40)$$

Ensuring the continuity of strains in the x-direction along the interface,

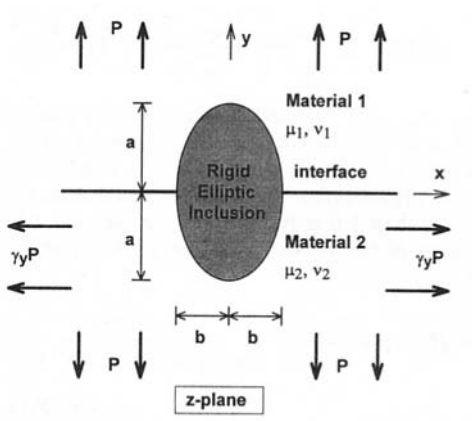


Fig. 2 Rigid elliptic inclusion in a bimaterial interface with remote tension in the *y*-direction

$$\gamma_y = \frac{\mu_2(\kappa_1 - 3) - \mu_1(\kappa_2 - 3)}{\mu_1(\kappa_2 + 1)} = \frac{2(2\beta_D - \alpha_D)}{(1 - \alpha_D)}. \quad (41)$$

Formulating the Riemann-Hilbert problem and solving it,

$$\phi_1^B(t_1) = -\frac{P}{4}\frac{(\kappa_1 - 3)}{\kappa_1}\frac{\lambda_1}{(1 + \lambda_1)}\left[1 - \frac{\chi_1(t_1)}{\chi_1(1)}\right]\frac{E_o}{1 - t_1}$$

$$+ \frac{P}{4}\frac{(\mu_2 - \mu_1)(\kappa_1 - 3)}{(\mu_1 + \kappa_1\mu_2)(1 + \lambda_1)}\sum_k\left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta_k)}\right]\frac{E_k}{\zeta_k - t_1}$$

$$+ \frac{\delta_1}{(1 + \lambda_1)}\frac{1}{\kappa_1}\sum_k\left[\overline{A_{k1}}B_k - \frac{P}{4}(\kappa_1 - 1)E_k\right]$$

$$\times\left[1 + \frac{(1 + \lambda_1 - \delta_1)}{\delta_1}\frac{\chi_1(t_1)}{\chi_1(\zeta_k)}\right]\frac{1}{\zeta_k - t_1} - \frac{\delta_1}{(1 + \lambda_1)}\frac{\mu_1}{\kappa_1\mu_2}$$

$$\times\sum_k\left[A_{k2}\overline{B_k} - \frac{P}{4}(\kappa_2 - 1)(1 + \gamma_y)\overline{E_k}\right]$$

$$\times\left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta_k')}\right]\frac{\zeta_k'^2}{\zeta_k' - t_1} \quad (42)$$

and

$$\phi_2^B(t_2) = -\frac{P}{4}\frac{\mu_2}{\mu_1}\frac{(\kappa_1 - 3)}{\kappa_2}\frac{\lambda_2}{(1 + \lambda_2)}\left[1 - \frac{\chi_2(t_2)}{\chi_2(1)}\right]\frac{E_o}{1 - t_2}$$

$$+ \frac{P}{4}\frac{\mu_2}{\mu_1}\frac{(\mu_1 - \mu_2)(\kappa_1 - 3)}{(\mu_2 + \kappa_2\mu_1)(1 + \lambda_2)}\sum_k\left[1 - \frac{\chi_2(t_2)}{\chi_2(\zeta_k)}\right]\frac{E_k}{\zeta_k - t_2}$$

$$+ \frac{\delta_2}{(1 + \lambda_2)}\frac{1}{\kappa_2}\sum_k\left[\overline{A_{k2}}B_k - \frac{P}{4}(1 + \gamma_y)(\kappa_2 - 1)E_k\right]$$

$$\times\left[1 + \frac{(1 + \lambda_2 - \delta_2)}{\delta_2}\frac{\chi_2(t_2)}{\chi_2(\zeta_k)}\right]\frac{1}{\zeta_k - t_2} - \frac{\delta_2}{(1 + \lambda_2)}\frac{\mu_2}{\kappa_2\mu_1}$$

$$\times\sum_k\left[A_{k1}\overline{B_k} - \frac{P}{4}(\kappa_1 - 1)\overline{E_k}\right]$$

$$\times\left[1 - \frac{\chi_2(t_2)}{\chi_2(\zeta_k')}\right]\frac{\zeta_k'^2}{\zeta_k' - t_2}. \quad (43)$$

As in the previous case, complex constants A_{k1} and A_{k2} are

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

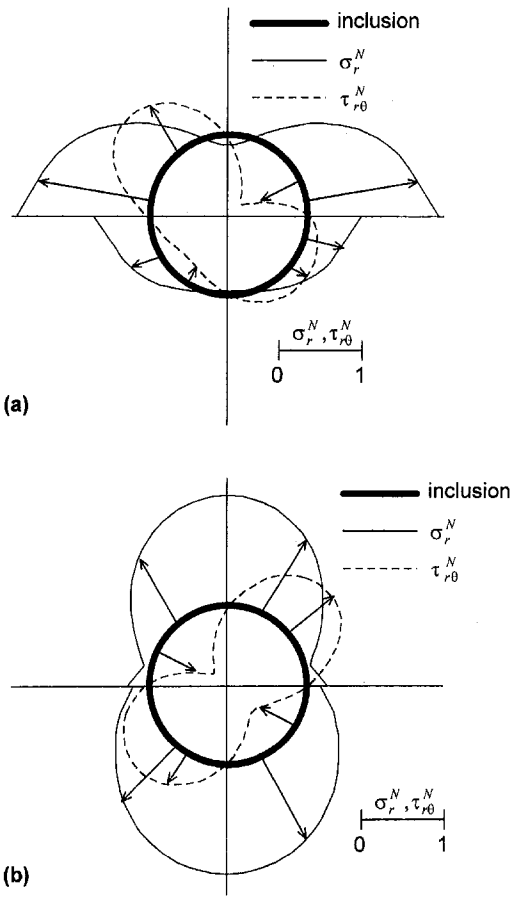


Fig. 3 Normalized curvilinear stress distribution around a rigid circular inclusion for uniform remote tension in (*a*) the *x*-direction and (*b*) the *y*-direction

determined by solving a set of $4N$ simultaneous equations in terms of the real and imaginary parts of A_{k1} and A_{k2}.

Also, the solution for a rigid thin inclusion at the bimaterial interface is given by the first terms on the right-hand side of Eqs. (40), (42), and (43).

Stress Distribution Around the Inclusion

Figures 3(*a*) and (*b*) show the distribution of normalized curvilinear stresses on the inclusion boundary corresponding to remote tension in the x and y-directions, respectively. These stresses are plotted radially from the inclusion boundary as indicated by the arrows. The particular case of $a/b = 1$ is considered. At this value, the ellipse degenerates into a circle and a rigid circular inclusion results as seen in the figures. Since the inclusion boundary is rigid, σ_r and σ_θ on the boundary are related as (Hasebe, 1979)

$$\frac{\sigma_\theta}{\sigma_r} = \frac{3 - \kappa_j}{1 + \kappa_j} \quad (j = 1, 2). \tag{44}$$

Therefore, only σ_r and $\tau_{r\theta}$ plots are shown here. These stresses are normalized as

$$\sigma_r^N = \frac{\sigma_r}{P_o} \quad \text{and} \quad \tau_{r\theta}^N = \frac{\tau_{r\theta}}{P_o} \tag{45}$$

where for material 1 $P_o = P$, and for material 2 $P_o = \gamma_x P$ or P corresponding to remote tension in the x or y-direction, respectively. A Fe-Al bimaterial system is considered which has $\mu_1 = 81.1$ GPa, $\nu_1 = 0.30$, $\mu_2 = 26.7$ GPa, and $\nu_2 = 0.32$. Therefore, the corresponding constants are given by

$$\Gamma = 0.32922 \quad \alpha_D = -0.49376 \quad \beta_D = -0.12538.$$

$$\gamma_x = 0.33890 \quad \gamma_y = 0.32535$$

For remote tension in the x-direction (Fig. 3(*a*)), it is seen that at the two tips of the inclusion (corresponding to points C_1 and C_2 in Figs. 1(*b*) and (*c*)), σ_r^N is compressive while $\tau_{r\theta}^N$ shows a sign change. Therefore, if the direction of the remote loading is reversed, i.e., if uniform remote compression is applied, σ_r^N will be tensile at these points. As seen in the next section, the absolute value of σ_r^N increases with increase in a/b which is equivalent to a decrease in ρ/b (ρ is the radius of curvature at the tip of the inclusion). This corresponds to an increasing stress concentration at the tips with increasing a/b. From Eq. (44), it is seen that the ratio of σ_θ/σ_r has maximum range of 0–1 in plane strain and 0–0.5 in plane stress corresponding to a Poisson's ratio variation of 0–0.5. Therefore σ_r is always greater than or equal to σ_θ. Hence, if the fracture toughness of the inclusion-material interface is assumed to be equal to that of the material, debonding at the tip is more likely than cracking.

Figure 3(*b*) shows similar plots for the case of uniform remote tension in the y-direction. Here also, $\tau_{r\theta}^N$ suffers a sign change at the inclusion tips. However, unlike the previous case, a remote tensile field results in tensile σ_r^N at the tips. The two earlier observations, (i) absolute value of σ_r^N increases with increase in a/b and (ii) debonding rather than cracking is likely (if fracture toughness of the inclusion-material interface is assumed to be equal to that of the material), are valid for this case also.

Comparing Figs. 3(*a*) and (*b*), it is seen that remote loading in the y-direction leads to more stress concentration and hence is more likely to result in debonds.

Stress Concentration Factor

Hasebe et al. (1988) analyzed the problem of a rigid semi-elliptic inclusion on the edge of a half-plane. Following Hasebe et al. (1986), they expressed the stress concentration factor at the inclusion tip in terms of the radius of curvature ρ (given by $\rho = b^2/a$) using an approximate form of a general expression. Extending this concept to the present problem, stress concentration factor (SCF) at an inclusion tip is expressed as

$$\text{SCF} = k_1\sqrt{\frac{a}{\rho}} + k_2 + k_3\sqrt{\frac{\rho}{a}} + k_4\frac{\rho}{a}. \tag{46}$$

Values of the coefficients (k_1, k_2, k_3, and k_4) are obtained by the least-squares method using values of SCF for different ρ/a. Since there are two inclusion tips in this problem, stress concentration factors at C_1 and C_2 (Figs. 1(*b*) and (*c*)) are, respectively, defined as

$$(\text{SCF})_{C1} = \frac{\sigma}{P} \quad \text{and} \quad (\text{SCF})_{C2} = \frac{\sigma}{\gamma P}, \tag{47}$$

where σ is either σ_r or σ_θ, and γ is either γ_x or 1 corresponding to remote loading in the x or y-direction, respectively. However, using Eq. (44), stress concentration factors corresponding to σ_r and σ_θ are related and so only one set of these values is reported in this paper.

Stress concentration factors for the bimaterial elliptic inclusion problem are checked using results of the earlier problem of a semi-elliptic inclusion on a half-plane (Hasebe et al., 1988). This is done by assuming a shear moduli ratio of $\Gamma = 10{,}000$ which corresponds to the near absence of material 1 and hence the problem becomes similar to a half-plane. As in the half-plane problem, remote uniform tension in the x-direction

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 1 Coefficients of stress concentration factor Eq. (46) for tension in the x-direction (Γ = 10,000)

$\kappa_1 = \kappa_2$		k_1	k_2	k_3	k_4
3.0	Material 1	0.00017	0.5699	-0.0661	0.0221
	Material 2	-0.00015	0.1222	0.0399	-0.0116
		(-0.0001)	(0.1222)	(0.0398)	(-0.0115)
2.0	Material 1	-0.2674	0.3631	-0.0665	0.0224
	Material 2	-0.1152	0.0445	0.0336	-0.0094
		(-0.1152)	(0.0447)	(0.0329)	(-0.0089)
1.0	Material 1	-0.7763	0.2004	-0.1450	0.0514
	Material 2	-0.3077	-0.0696	0.0293	-0.0077
		(-0.3077)	(-0.0700)	(0.0300)	(-0.0082)

is assumed. Table 1 shows coefficients k_1 to k_4 for different κ_1 ($=\kappa_2$) values. It is seen that the coefficients in material 2 agree well with values corresponding to the half-plane problem which are shown in brackets. Coefficients corresponding to the near-absent material 1 are not physically relevant and are reported only to provide the full set of results. Next, the error between the actual and approximate stress concentration factors for $\kappa_1 = \kappa_2 = 2$ is obtained. Table 2 gives the actual stress concentration factors and error percentages with values for the half-plane problem in brackets. Again, a good agreement is evident. Here, also, values corresponding to material 1 are reported only to give a complete picture.

Hasebe et al. (1992) and Okumura et al. (1995) solved the problem of an elliptical void in a bimaterial interface subject to remote uniform tension along and normal to the interface. They obtained the stress concentration factors at the void tips and expressed it in terms of an approximate expression as Eq. (46)—however, they used only three coefficients k_1, k_2, and k_3. They considered a bimaterial system with $\alpha_D = -0.50$ and $\beta_D = -0.15$ which compares closely with the Fe-Al system used in this paper for the inclusion problem. Table 3 shows stress concentration factors at the two tips for remote tension in the x-direction. Corresponding values for an elliptic void (from Hasebe et al., 1992) are shown in brackets. It is seen that stress concentration factors at void tips are greater than those at inclusion tips by an order of magnitude. Table 4 shows similar values for remote tension in the y-direction. In this case, however, stress concentration factors at the tips of an inclusion and a void (from Okumura et al., 1995) are comparable.

Stress Intensity Factor at the Tip of a Thin Rigid Inclusion

Ballarini (1990) analyzed the problem of a thin rigid inclusion at a bimaterial interface between two isotropic materials. This solution is obtained using the condition $E_k = 0$ ($k = 1, 2, \ldots N$) in the present paper. He determined the singularity and stress intensity factors at the tip of the inclusion, and compared them with corresponding expressions for an interface crack. The problem of a thin rigid inclusion at the interface between dissimilar anisotropic media was solved by Asundi and Deng (1995). Here the problem of a thin rigid inclusion perpendicular to a bimaterial interface is addressed.

Hasebe and Takeuchi (1985) studied the problem of a thin rigid body in a semi-infinite plate. The tip of this thin rigid body was similar to a crack tip and the stress field in its vicinity was expressed in terms of stress intensity factors defined as

$$K_{IR} - iK_{IIR} = \sqrt{\pi}\,\exp(-i\theta_C/2)\,\frac{\phi'(\sigma_C)}{\sqrt{\omega''(\sigma_C)}} \qquad (48)$$

where σ_C is the point on the unit circle corresponding to the tip of the thin rigid body, θ_C is the angle between the direction of the thin rigid body and the x-axis, and K_{IR} and K_{IIR} are the stress intensity factors in modes I and II, respectively. In this case, $\sigma_C = -1$ and $\theta_C = \pi/2$. The stress field in terms of these stress intensity factors is given by

$$\sigma_x = \frac{K_{IR}}{\sqrt{2\pi r}}\left[(\kappa + 3)\cos\frac{\theta}{2} - \sin\theta\sin\frac{3\theta}{2}\right] - \frac{K_{IIR}}{\sqrt{2\pi r}}\left[(3 - \kappa)\sin\frac{\theta}{2} + \sin\theta\cos\frac{3\theta}{2}\right]$$

$$\sigma_y = -\frac{K_{IR}}{\sqrt{2\pi r}}\left[(\kappa - 1)\cos\frac{\theta}{2} - \sin\theta\sin\frac{3\theta}{2}\right] - \frac{K_{IIR}}{\sqrt{2\pi r}}\left[(\kappa + 1)\sin\frac{\theta}{2} - \sin\theta\cos\frac{3\theta}{2}\right]$$

Table 2 Stress concentration factors at inclusion tips for tension in the x-direction ($\kappa_1 = \kappa_2 = 2$ and Γ = 10,000)

ρ/b	Material 1		Material 2	
	σ_r/P	Error (%)	$\sigma_r/\gamma_x P$	Error (%)
0.01	-2.318	0.007	-1.1039 (-1.1037)	-0.006 (0.006)
0.02	-1.537	-0.010	-0.7654 (-0.7652)	0.006 (-0.007)
0.04	-0.986	-0.013	-0.5251 (-0.5250)	0.006 (-0.010)
0.08	-0.599	0.012	-0.3539 (-0.3539)	-0.008 (0.003)
0.10	-0.501	0.026	-0.3100 (-0.3100)	-0.015 (0.012)
0.20	-0.260	0.071	-0.1998 (-0.1998)	-0.031 (0.033)
0.40	-0.093	-0.033	-0.1201 (-0.1201)	0.005 (0.016)
0.80	0.023	0.348	-0.0618 (-0.0618)	0.037 (-0.040)
1.00	0.052	-0.372	-0.0464 (-0.0464)	-0.141 (0.077)
2.00	0.121		-0.0069 (-0.0069)	
4.00	0.168		0.0230 (0.0229)	
8.00	0.199		0.0459 (0.0458)	
10.00	0.207		0.0520 (0.0520)	

Table 3 Stress concentration factors at inclusion tips for tension in the x-direction (Fe-Al bimaterial system)

ρ/b	Material 1 (σ_θ/P)	Material 2 ($\sigma_\theta/\gamma_x P$)
0.001	-81.837 (2027.000)	-135.63
0.005	-16.593 (406.400)	-27.475
0.010	-8.292 (203.700)	-13.716
0.020	-4.129 (102.300)	-6.815
0.040	-2.045 (51.630)	-3.360
0.060	-1.350 (34.730)	-2.208
0.080	-1.002 (26.280)	-1.632
0.100	-0.793 (21.220)	-1.286
0.200	-0.375 (11.090)	-0.595
0.400	-0.165 (6.026)	-0.251
0.600	-0.095 (4.344)	-0.137
0.800	-0.060 (3.504)	-0.080
1.000	-0.038 (3.001)	-0.047

Table 4 Stress concentration factors at inclusion tips for tension in the y-direction (Fe-Al bimaterial system)

ρ/b	Material 1 (σ_θ/P)	Material 2 (σ_θ/P)
0.001	222.83 (-169.004)	215.22
0.005	45.594 (-34.475)	44.085
0.010	23.046 (-17.646)	22.313
0.100	2.676 (-2.501)	2.644
0.200	1.542 (-1.664)	1.550
0.400	0.975 (-1.250)	1.003
0.600	0.786 (-1.115)	0.821
0.800	0.691 (-1.049)	0.730
1.000	0.635 (-1.011)	0.675
1.250	0.589 (-0.981)	0.632
1.500	0.559 (-0.962)	0.603
2.000	0.521 (-0.939)	0.567
2.500	0.498 (-0.926)	0.545

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\tau_{xy} = \frac{K_{IR}}{\sqrt{2\pi r}}\left[(\kappa + 1)\sin\frac{\theta}{2} + \sin\theta\cos\frac{3\theta}{2}\right] - \frac{K_{IIR}}{\sqrt{2\pi r}}\left[(\kappa - 1)\cos\frac{\theta}{2} + \sin\theta\sin\frac{3\theta}{2}\right]. \quad (49)$$

Extending this concept to the present problem, two sets of stress intensity factors exist at the two inclusion tips (C_1 and C_2 in Figs. 1(b) and (c)). These stress intensity factors are normalized as

$$(K^N_{IR} - iK^N_{IIR})_j = \frac{(K_{IR} - iK_{IIR})_j}{P_o\sqrt{\pi a}} \quad (j = 1, 2) \quad (50)$$

where for C_1 (in material 1) $P_o = P$, and for C_2 (in material 2) $P_o = \gamma_x P$ or P corresponding to remote loading in the x or y-direction, respectively. Table 5 shows normalized stress intensity factors at C_1 and C_2 for the Fe-Al bimaterial system used earlier. Results for the two cases of remote tension (in the x and y-directions) are shown. Since the shear stress is zero at the tips (Figs. 3(a) and (b)), $(K^N_{IIR})_j$ is always zero. $(K^N_{IR})_j$ is negative at the two tips for tension in the x-direction because σ_θ is compressive at these points (from Fig. 3(a) and Eq. (44)).

These stress intensity factors can also be determined from the expression for stress concentration factor (Eq. (46)) as explained in Iida et al. (1987). Using a similar procedure, $(K^N_{IR})_j$ is obtained as

$$(K^N_{IR})_j = -\frac{(C_\theta)_j(k_1)_j\sqrt{2}}{(\kappa_j - 1)} \quad (j = 1, 2) \quad (51)$$

Here C_θ corresponds to the constant given in the above-mentioned paper and k_1 is the first coefficient in the expression for stress concentration factor (Eq. (46)) corresponding to σ_θ. A similar relation can be written for $(K^N_{IIR})_j$ in terms of $\tau_{r\theta}$. But $\tau_{r\theta}$ is zero at the tips (Figs. 3(a) and (b)), and so $(K^N_{IIR})_j$ is also zero. Table 6 shows the coefficients of Eq. (46) corresponding to σ_θ for the Fe-Al system used earlier. Tensions in both x and y-directions are considered. The values of C_θ corresponding to the two materials is obtained from Iida et al. (1987) as

$$(C_\theta)_1 = -0.47211 \quad (C_\theta)_2 = -0.39819.$$

Substituting these values of C_θ and k_1 from Table 6 into Eq. (51), the normalized stress intensity factors are obtained as

Tension in x-direction: $(K^N_{IR})_1 = -0.0697$ $(K^N_{IR})_2 = -0.1082$

Tension in y-direction: $(K^N_{IR})_1 = 0.1892$ $(K^N_{IR})_2 = 0.1712$.

As mentioned earlier, $(K^N_{IIR})_j = 0.0$ for all these cases. It is seen that these values compare very well with those in Table 5.

Conclusions

The displacement boundary value problem of a bimaterial interface was solved. The particular case of a rigid elliptic inclusion at a bimaterial interface was considered. However, it is not difficult to solve other displacement boundary conditions. Other rigid inclusion geometries can be solved by changing coefficients of the mapping function. Uniform remote tensions both along and normal to the interface (i.e., in x and y-directions, respectively) were considered separately. It was seen that Dundurs' constants alone are not sufficient to define in the displacement boundary value problem; another material parameter (e.g., moduli ratio) is also required. Stresses on the inclusion boundary were obtained in a curvilinear coordinate system aligned with the tangent and normal to this boundary. It was seen that a remote tension in the x-direction led to compressive normal stresses (σ_r and σ_θ) at the tips while remote tension in the y-direction resulted in tensile stresses at the tips. Further, $\tau_{r\theta}$ was zero at the tips for both the cases of loading due to symmetry.

Since the inclusion boundary is rigid, stresses σ_r and σ_θ on the boundary are related in terms of the Poisson's ratio (see Eq. (44)). Using this relation it was seen that the magnitude of σ_r was always greater than or equal to σ_θ. This indicated that, at an inclusion tip, debonding is more likely than cracking (assuming that the fracture toughness of the inclusion-material interface is the same as that of the material). Further it was seen that stress concentration at the inclusion tips increases with increase in the a/b ratio. Stress concentration factors at the tips were expressed in terms of the radius of curvature using an approximate form of a general expression. These values were checked against results of a half-plane problem reported earlier.

Stress concentration factors at an inclusion tip were compared with those at a void tip from earlier papers. It was seen that y-direction loading leads to comparable values. For loading in the x-direction, however, the void tips have values which are an order of magnitude higher. Therefore voids are more susceptible to cause failure when loading is in x-direction, while for loads in the y-direction both voids and inclusions are equally capable of initiating failure.

When the inclusion is thin (i.e., b/a is zero), stress field at the tips can be expressed in terms of stress intensity factors as in the case of crack tips. These values are shown in Table 5 corresponding to both loading conditions. It was seen that the mode II stress intensity factors are always zero. These stress intensity factors were also be obtained from the approximate expression for stress concentration factors (Eq. (46)). These values compared well with earlier results in Table 5.

Table 5 Normalized stress intensity factors at C_1 and C_2 (Fe-Al bimaterial system)

	at the tip C_1 (Material 1)		at the tip C_2 (Material 2)	
	$(K^N_{IR})_1$	$(K^N_{IIR})_1$	$(K^N_{IR})_2$	$(K^N_{IIR})_2$
Tension in x direction	-0.0696	0.000	-0.1081	0.000
Tension in y direction	0.1890	0.000	0.1711	0.000

Table 6 Coefficients of stress concentration factor Eq. (46) (Fe-Al bimaterial system)

		k_1	k_2	k_3	k_4
Tension in x direction	Material 1	-0.0835	0.0413	0.0052	-0.0014
	Material 2	-0.1383	0.0980	-0.0088	0.0026
Tension in y direction	Material 1	0.2267	0.4089	-0.0015	0.0005
	Material 2	0.2189	0.4548	0.0023	-0.0006

References

Asundi, A., and Deng, W., 1995, "Rigid inclusions on the interface between dissimilar anisotropic media," *Journal of Mechanics and Physics of Solids,* Vol. 43, pp. 1045–1058.

Ballarini, R., 1990, "A rigid line inclusion at a bimaterial interface," *Engineering Fracture Mechanics,* Vol. 37, pp. 1–5.

Comninou, M., 1990, "An Overview of Interface Cracks," *Engineering Fracture Mechanics,* Vol. 37, pp. 197–208.

England, A.H., 1965, "A Crack Between Dissimilar Media," ASME JOURNAL OF APPLIED MECHANICS, Vol. 32, pp. 400–402.

Erdogan, F., 1965, "Stress Distribution in Bonded Dissimilar Materials With Cracks," ASME JOURNAL OF APPLIED MECHANICS, Vol. 32, pp. 403–410.

Hasebe, N., 1972, "Stress Analysis of a Semi-Infinite Plate with a Semi-Elliptic Notch," *Bulletin of Nagoya Institute of Technology,* Vol. 24, pp. 295–301.

Hasebe, N., and Inohara, S., 1980, "Stress Analysis of a Semi-Infinite Plate with an Oblique Edge Crack," *Ingenieur-Archiv,* Vol. 49, pp. 51–62.

Hasebe, N., Okumura, M., and Nakamura, T., 1992, "Bonded Bi-material Half-Planes With Semi-elliptical Notch Under Tension Along the Interface," ASME JOURNAL OF APPLIED MECHANICS, Vol. 59, pp. 77–83.

Hasebe, N., Sugimoto, T., and Nakamura, T., 1986, "Stress Analysis of a Blunted Notch in a Clamped Edge," *ASCE Journal of Engineering Mechanics,* Vol. 112, pp. 142–153.

Hasebe, N., and Takeuchi, T., 1985, "Stress Analysis of a Semi-infinite Plate with Thin Rigid Body," *International Journal of Engineering Science,* Vol. 23, pp. 531–539.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Hasebe, N., Tsutsui, S., and Nakamura, T., 1988, "Debondings at a Semielliptic Rigid Inclusion on the Rim of a Half Plane," ASME JOURNAL OF APPLIED MECHANICS, Vol. 55, pp. 574–579.

Hayashi, K., and Nemat-Nasser, S., 1981, "On Branched, Interface Cracks," ASME JOURNAL OF APPLIED MECHANICS, Vol. 48, pp. 529–533.

Iida, J., Hasebe, N., and Matuura, S., 1987, "Intensity of Corner in Fixed Edge of Plane Problem," *ASCE Journal of Engineering Mechanics,* Vol. 113, pp. 1194–1207.

Muskhelishvili, N., 1963, *Some Basic Problems of the Mathematical Theory of Elasticity,* Noordhoff, Groningen, The Netherlands.

Okumura, M., Hasebe, N., and Nakamura, T., 1995, "Bimaterial plane with elliptic hole under uniform tension normal to the interface," *International Journal of Fracture,* Vol. 71, pp. 293–310.

Rice, J.R., 1988, "Elastic Fracture Mechanics Concepts for Interfacial Cracks," ASME JOURNAL OF APPLIED MECHANICS, Vol. 55, pp. 98–103.

Rice, J.R., and Sih, G.C., 1965, "Plane problems of cracks in dissimilar media," ASME JOURNAL OF APPLIED MECHANICS, Vol. 32, pp. 418–423.

Salama, M., and Hasebe, N., 1996, "Stress Concentration Factors at an Elliptical Hole on the Interface Between Bonded Dissimilar Half-Planes Under Bending Moment," ASME JOURNAL OF APPLIED MECHANICS, Vol. 63, pp. 7–14.

Shih, C.F., and Asaro, R.J., 1990, "Elastic-plastic and asymptotic fields of interface cracks," *International Journal of Fracture,* Vol. 42, pp. 101–116.

Sih, G.C., and Rice, J.R., 1964, "The Bending of Plates of Dissimilar Materials With Cracks," ASME JOURNAL OF APPLIED MECHANICS, Vol. 31, pp. 477–482.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Effect of Material Properties in Certain Thermoelastic Contact Problems

D. Joachim-Ajao

J. R. Barber
Mem. ASME

Department of Mechanical Engineering
and Applied Mechanics,
University of Michigan,
Ann Arbor, MI 48109-2125

When two conforming bodies slide against each other, frictional heating and thermoelastic distortion generally causes the contacting surfaces to become convex and hence leads to a reduction in the size of the contact area. It is shown that, under such circumstances, the contact area is independent of the applied contact loads and the thermal and mechanical fields are linearly proportional to these loads. For two-dimensional problems involving a plane boundary, it is shown that there is a reduced dependence on material properties and in the case of a single material, the solution depends only on a single parameter which can be interpreted as a dimensionless sliding speed. These results extend to both steady-state and transient problems and therefore also characterize the critical sliding speed above which the system is unstable.

1 Introduction

In an interesting series of papers, Dundurs and his co-workers (Dundurs and Stippes, 1970; Dundurs, 1975; Keer et al., 1972) established some unexpectedly simple results for *receding* unilateral frictionless contact problems; i.e., problems in which the extent of the contact area under load is contained within or identical to that in the unloaded state. In particular, they showed that the contact area may decrease discontinuously as soon as the load is applied, but that further increase of load causes no further change in the contact area. It then follows that the stress and displacement fields are proportional to the applied boundary tractions or displacements as long as these retain the same sign. Furthermore, in plane elasticity, if the boundary conditions are defined in terms of tractions, the stresses are independent of the elastic constants, including Poisson's ratio and there is a reduced dependence on material properties for cases involving dissimilar materials.

Recent numerical studies by the present authors of thermoelastic contact problems involving frictional heating have exhibited a related reduced dependence on material properties. In the present paper, we shall establish the conditions under which such reduced dependence can be expected and draw some interesting conclusions for steady-state and transient thermoelastic contact and for thermoelastic contact stability.

2 Problem Statement

We consider the conforming contact problem of Fig. 1(*a*) in which two plane elastic bodies make contact over an extended region Γ_0 when there is no applied load. We now apply some boundary tractions, as shown in Fig. 1(*b*), and allow relative sliding to occur between the bodies in the out-of-plane direction, resulting in an out-of-plane shear stress $|\sigma_{nz}| = fp$, where n is the local normal to the contact surface, p is the contact pressure, and f is the coefficient of friction. If the sliding velocity is V, heat will be generated at the interface according to the relation

$$q = fVp. \tag{1}$$

The contact area may change as a result of the thermal and mechanical loading and the instantaneous value is denoted by Γ_c. Heat input to either body produces a locally convex thermoelastic distortion (Dundurs, 1974) and hence the boundary condition (1) generally results in receding contact ($\Gamma_c \subseteq \Gamma_0$).

Problems of this class are of importance because of the application to brakes, clutches, and seals. Burton et al. (1973) and Barber (1976) obtained steady-state solutions for the case of two sliding half-planes and related transient problems have been investigated numerically by Azarkhin and Barber (1986) and Zagrodzki (1990). If the sliding speed V is sufficiently high, both steady-state and transient solutions can be unstable in the sense that an arbitrarily small perturbation in the initial conditions can cause large changes in the subsequent behavior (Dow and Burton, 1972; Du et al., 1997). This phenomenon is known as frictionally excited thermoelastic instability or TEI (Barber, 1969).

2.1 Thermal conditions. The temperature field T^i in the bodies must satisfy the heat conduction equation

$$\nabla^2 T^i = \frac{1}{k_i}\frac{\partial T^i}{\partial t}, \tag{2}$$

where $i = 1, 2$ denotes bodies 1, 2, respectively, t is time, and k_i is the thermal diffusivity of material i.

In the contact region, Γ_c, we assume continuity of temperature

$$T^1 = T^2 \quad \text{in} \quad \Gamma_c \tag{3}$$

and a heat balance yields

$$q^1 - q^2 = -K_1\frac{\partial T^1}{\partial n} + K_2\frac{\partial T^2}{\partial n} = fVp \quad \text{in} \quad \Gamma_c, \tag{4}$$

from Eq. (1), where K_i is the thermal conductivity of material i and we have adopted the convention that the common normal n points into body 1.

At the free boundaries of the body $\Gamma - \Gamma_0$, we assume that there is radiation to a medium at a reference temperature of zero and hence

$$q_n = -K_i\frac{\partial T^i}{\partial n} = hT^i \quad \text{in} \quad \Gamma - \Gamma_0, \tag{5}$$

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Sept. 16, 1997; final revision, Apr. 16, 1998. Associate Technical Editor: L. T. Wheeler.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

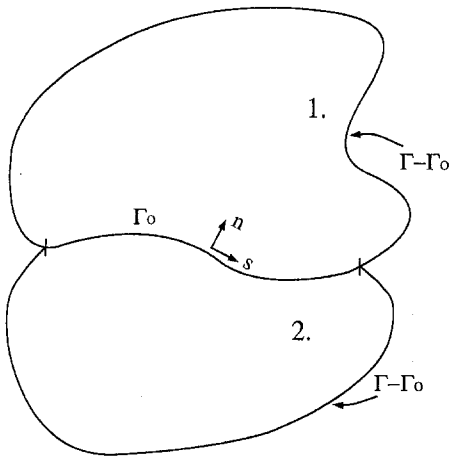


Fig. 1(*a*)

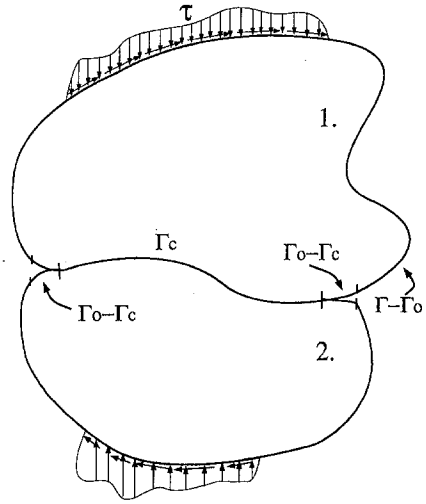


Fig. 1(*b*)

Fig. 1 A receding contact problem. (*a*) With no applied tractions, the contact region is Γ_0. (*b*) With applied tractions, τ, the contact region is Γ_c, and $\Gamma_c \subseteq \Gamma_0$.

where h is a heat transfer coefficient. Finally, in the region $\Gamma_0 - \Gamma_c$ which loses contact as a result of thermoelastic distortion and loading, we assume there is heat transfer between the bodies through a heat transfer coefficient h_0, i.e.,

$$-K_1 \frac{\partial T^1}{\partial n} = -K_2 \frac{\partial T^2}{\partial n} = h_0(T^2 - T^1) \quad \text{in} \quad \Gamma_0 - \Gamma_c. \tag{6}$$

We note that by setting $h = 0$ or $h_0 = 0$ in (5), (6), we recover the thermally insulated boundary condition, while by setting $h = \infty$ or $h_0 = \infty$, we obtain the boundary conditions $T = 0$ and $T^1 = T^2$, respectively.

2.2 Mechanical conditions. We assume that there are no body forces—i.e., that the bodies are loaded only on the boundaries Γ—in which case the stress field in both bodies must satisfy the equilibrium relations

$$\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} = 0; \quad \frac{\partial \sigma_{yx}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} = 0 \tag{7}$$

and the resulting strains must satisfy the compatibility equation

$$\frac{\partial^2 e_{xx}}{\partial y^2} + \frac{\partial^2 e_{yy}}{\partial x^2} = 2 \frac{\partial^2 e_{xy}}{\partial x \partial y}. \tag{8}$$

These quantities are related by the thermoelastic Hooke's law

$$e^i_{xx} = \frac{\sigma^i_{xx}}{2\mu_i} - \frac{(3 - \kappa_i)(\sigma^i_{xx} + \sigma^i_{yy})}{8\mu_i} + \eta_i T^i \tag{9}$$

$$e^i_{xy} = \frac{\sigma^i_{xy}}{2\mu_i} \tag{10}$$

etc., where

$$\kappa_i = (3 - 4\nu_i); \quad \eta_i = \gamma_i(1 + \nu_i) \quad \text{for plane strain,} \tag{11}$$

$$\kappa_i = \frac{(3 - \nu_i)}{(1 + \nu_i)}; \quad \eta_i = \gamma_i \quad \text{for plane stress} \tag{12}$$

and μ_i, ν_i, γ_i are, respectively, the shear modulus, Poisson's ratio, and coefficient of thermal expansion[1] for material i.

We assume traction boundary conditions on the free boundaries, i.e.,

$$\sigma_{nn} = \tau_n; \quad \sigma_{ns} = \tau_s \quad \text{in} \quad \Gamma - \Gamma_0, \tag{13}$$

where n, s are Cartesian coordinates normal and tangential to the local boundary and τ_n, τ_s are prescribed normal and tangential tractions, respectively.

In the contact region, there is no in-plane shear traction and the normal traction is continuous, so

$$\sigma^1_{nn} = \sigma^2_{nn} \equiv -p; \quad \sigma^1_{ns} = \sigma^2_{ns} = 0 \quad \text{in} \quad \Gamma_c. \tag{14}$$

In addition, we have the kinematic contact condition

$$u^1_n = u^2_n \quad \text{in} \quad \Gamma_c \tag{15}$$

and the contact pressure is required to be non-negative; i.e.,

$$p \geq 0 \quad \text{in} \quad \Gamma_c. \tag{16}$$

Finally, the region that loses contact must be traction-free

$$\sigma^1_{nn} = \sigma^2_{nn} = 0; \quad \sigma^1_{ns} = \sigma^2_{ns} = 0 \quad \text{in} \quad \Gamma_0 - \Gamma_c \tag{17}$$

and the displacements must result in a non-negative gap, giving

$$u^1_n - u^2_n \geq 0 \quad \text{in} \quad \Gamma_0 - \Gamma_c. \tag{18}$$

We restrict attention to the case where the two bodies are both simply connected[2], in which case the above conditions completely define the problem of Fig. 1.

3 Dimensionless Formulation

In order to demonstrate the dependence of the solution on material parameters, it is convenient to recast these conditions in terms of dimensionless quantities, normalized with respect to the applied tractions. For this purpose, we introduce a loading parameter $S > 0$ with units of stress and define dimensionless tractions $\hat{\tau}_n$, $\hat{\tau}_s$ through the relations

$$\tau_n = S\hat{\tau}_n; \quad \tau_s = S\hat{\tau}_s. \tag{19}$$

We also introduce a geometric length scale a which will enable us to expose simplifications resulting from geometric self-similarity.

With these preliminaries, we now define dimensionless coordinates, stresses, displacements, etc., through the relations

$$x = a\hat{x}; \quad y = a\hat{y}; \quad n = a\hat{n}; \quad s = a\hat{s}; \tag{20}$$

$$t = \frac{a^2 \hat{t}}{k_1}; \tag{21}$$

[1] Notice that we here use the symbol γ rather than the more familiar α for the coefficient of thermal expansion to avoid confusion with Dundurs bimaterial constant α which appears in Section 3.

[2] If either body is multiply connected, the compatibility condition must be supplemented a set of Cesaro integrals around each hole (Love, 1927).

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\sigma = S\hat{\sigma}; \quad p = S\hat{p}; \tag{22}$$

$$u^i = \frac{aS(\kappa_i + 1)\hat{u}^i}{\mu_i}; \tag{23}$$

$$T^i = \frac{S(\kappa_i + 1)\hat{T}^i}{\mu_i \eta_i}; \tag{24}$$

$$q^i = \frac{SK_i(\kappa_i + 1)\hat{q}^i}{a\mu_i \eta_i}. \tag{25}$$

We also define the alternate strain measure $\hat{e}^i$ through the relation

$$e^i = \frac{S(\kappa_i + 1)\hat{e}^i}{\mu_i}. \tag{26}$$

In this notation, the governing equations (2), (7)–(10) take the simpler forms

$$\nabla^2 \hat{T}^1 = \frac{\partial \hat{T}^1}{\partial \hat{t}}; \quad \nabla^2 \hat{T}^2 = \hat{k}\frac{\partial \hat{T}^2}{\partial \hat{t}}; \tag{27}$$

$$\frac{\partial \hat{\sigma}_{xx}}{\partial \hat{x}} + \frac{\partial \hat{\sigma}_{xy}}{\partial \hat{y}} = 0; \quad \frac{\partial \hat{\sigma}_{yx}}{\partial \hat{x}} + \frac{\partial \hat{\sigma}_{yy}}{\partial \hat{y}} = 0; \tag{28}$$

$$\frac{\partial^2 \hat{e}_{xx}}{\partial \hat{y}^2} + \frac{\partial^2 \hat{e}_{yy}}{\partial \hat{x}^2} = 2\frac{\partial^2 \hat{e}_{xy}}{\partial \hat{x}\partial \hat{y}}; \tag{29}$$

$$\hat{e}^i_{xx} = \frac{\hat{\sigma}^i_{xx}}{2(\kappa_i + 1)} - \frac{(3 - \kappa_i)(\hat{\sigma}^i_{xx} + \hat{\sigma}^i_{yy})}{8(\kappa_i + 1)} + \hat{T}^i; \tag{30}$$

$$\hat{e}^i_{xy} = \frac{\hat{\sigma}^i_{xy}}{2(\kappa_i + 1)}, \tag{31}$$

where

$$\hat{k} = \frac{k_1}{k_2}. \tag{32}$$

A more convenient form of the compatibility Eq. (29) can be obtained by substituting for the strains using (30), (31) and using the equilibrium Eq. (28) to eliminate the shear stress terms. After cancelling a nonzero dimensionless parameter, we obtain

$$\nabla^2(\hat{\sigma}^i_{xx} + \hat{\sigma}^i_{yy} + 8\hat{T}^i) = 0. \tag{33}$$

The corresponding dimensionless forms of the boundary conditions (3)–(6), (13)–(18) are

$$\hat{T}^1 = \hat{\eta}\hat{M}\hat{T}^2; \tag{34}$$

$$-\frac{\partial \hat{T}^1}{\partial \hat{n}} + \frac{\hat{M}\hat{\eta}}{\hat{K}}\frac{\partial \hat{T}^2}{\partial \hat{n}} = \hat{V}\hat{p}; \tag{35}$$

$$\hat{\sigma}^1_{nn} = \hat{\sigma}^2_{nn} \equiv -\hat{p}; \quad \hat{\sigma}^1_{ns} = \hat{\sigma}^2_{ns} = 0; \tag{36}$$

$$\hat{u}^1_n = \hat{M}\hat{u}^2_n; \tag{37}$$

$$\hat{p} \geq 0, \tag{38}$$

in Γ_c,

$$-\frac{\partial \hat{T}^1}{\partial \hat{n}} = \hat{h}\hat{T}^1; \quad -\frac{\partial \hat{T}^2}{\partial \hat{n}} = \hat{K}\hat{h}\hat{T}^2; \tag{39}$$

$$\hat{\sigma}_{nn} = \hat{\tau}_n; \quad \hat{\sigma}_{ns} = \hat{\tau}_s, \tag{40}$$

in $\Gamma - \Gamma_0$ and

$$-\frac{\partial \hat{T}^1}{\partial \hat{n}} = -\frac{\hat{M}\hat{\eta}}{\hat{K}}\frac{\partial \hat{T}^2}{\partial \hat{n}} = \hat{h}_0(\hat{M}\hat{\eta}\hat{T}^2 - \hat{T}^1); \tag{41}$$

$$\hat{\sigma}^1_{nn} = \hat{\sigma}^2_{nn} = 0; \quad \hat{\sigma}^1_{ns} = \hat{\sigma}^2_{ns} = 0; \tag{42}$$

$$\hat{u}^1_n - \hat{M}\hat{u}^2_n \geq 0, \tag{43}$$

in $\Gamma_0 - \Gamma_c$, where

$$\hat{\eta} = \frac{\eta_1}{\eta_2}; \quad \hat{M} = \frac{\mu_1(\kappa_2 + 1)}{\mu_2(\kappa_1 + 1)}; \quad \hat{K} = \frac{K_1}{K_2};$$

$$\hat{h} = \frac{ha}{K_1}; \quad \hat{h}_0 = \frac{h_0 a}{K_1} \tag{44}$$

and

$$\hat{V} = \frac{fV\mu_1\eta_1 a}{K_1(\kappa_1 + 1)}. \tag{45}$$

Notice that $\hat{M}$ can be defined in terms of Dundurs bimaterial parameter α through

$$\hat{M} = \frac{1 - \alpha}{1 + \alpha}. \tag{46}$$

It follows that in the dimensionless formulation, the problem depends on the material properties only through κ_i and the dimensionless parameters defined in Eqs. (32), (44), (45).

4 Linearity With Loading Parameter

The loading parameter S does not appear in the dimensionless formulation of the problem as defined by the above equations. It follows immediately that the extent of the contact area Γ_c, which is determined by the inequalities (38), (43), is independent of S. Furthermore, since the only nonlinearity in the problem is that associated with these inequalities and the possibility of the contact area varying with load, it then follows that the temperature, stress, and displacement fields are all linear with S.

It must be emphasised that this result applies only for the case where S is positive, since in deriving the inequalities (38), (43) from (16), (18) we have cancelled a common factor S. However, this is not a serious restriction since with an appropriate sign convention the case $S < 0$ will correspond to loading that would not support contact even without thermoelastic effects.

Notice that these results apply to transient as well as steady-state problems. Thus, if the bodies are initially at temperature $T = 0$ (for example) and sliding starts at time $t = 0$, we anticipate that the contact area will start at some value $\Gamma_c(0) \leq \Gamma_0$ and will get progressively smaller as thermal distortion occurs. The fact that the dimensionless formulation does not contain S demonstrates that the evolution of the contact area in time, $\Gamma_c(\hat{t})$, is independent of S and that the instantaneous temperature, displacement, and stress fields for different values of S will be linear scalings of each other for a particular value of $\hat{t}$.

The same argument applies if the dimensionless tractions $\hat{\tau}_n$, $\hat{\tau}_s$ and/or the sliding speed $\hat{V}$ are functions of time.

We have developed Eqs. (27)–(43) for the simpler two-dimensional case, but it is clear that a similar line of reasoning could be developed for the three-dimensional problem of Fig. 2, in which a brake pad slides against a rotating annular disk.

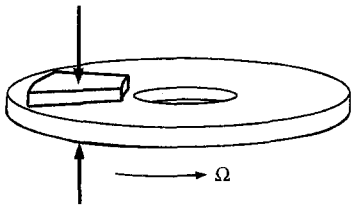


Fig. 2 A three-dimensional receding contact problem for a disk brake. As in the two-dimensional case, the contact area will be independent of the applied load.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

In addition, the argument is readily extended to cases involving multiply connected regions and/or body force loading. To see this, suppose that a solution has been obtained for a particular value of $S(=S_1)$, so that the dimensional temperature, displacement, and stress fields are all known and in particular that they satisfy the unilateral contact inequalities analogous to (16, 18) in $\Gamma_0 - \Gamma_c(t)$ at all times t. We now postulate that the corresponding fields for a different but proportional loading scenario $S = S_2$ will be S_2/S_1 times the original solution. Clearly all the linear governing equations and boundary conditions will be satisfied by this assumption and the new inequality constraints will simply be a positive multiplier of the old ones and hence will also still be satisfied.

We conclude therefore that if a brake pad conforms to the disk at zero applied load, the evolution of the contact area due to combined elastic and thermoelastic effects will be independent of the loading parameter.

This result leads to a remarkable conclusion in the case of mechanical seals. Burton et al. (1973) showed that if a seal between two elastic half-planes operates above the critical speed, the system will be unstable, but will tend to a stable nonuniform steady state in which there are regions of contact and separation, permitting leakage to occur through the separation regions. If the applied load is increased in a misguided attempt to reestablish full contact, the extent of the separation areas will be unchanged, but all the thermal and mechanical fields will be increased, leading in particular to an increased gap and hence increased leakage. The original proof by Burton et al. (1973) was restricted to their particular system, but it is clear from the arguments of this section that the result is of completely general application.

Notice that the three-dimensional argument depends on the unloaded contact problem being conforming in all kinematic configurations. For this to occur, all parts of the contacting surfaces must have the same shape and this restricts the result to cases where the contacting surfaces are plane or (possibly incomplete) surfaces of revolution (e.g., a drum brake). No such restriction applies in the two-dimensional case, since the motion then occurs out-of-plane.

5 The Contact Boundary Conditions

We now return to the two-dimensional formulation of Eqs. (27) – (43) and consider in particular the contact boundary conditions (37), (43). Under certain circumstances Eq. (37) can be replaced by a relationship involving strains only. Dundurs (1989) showed that the change $\Delta\hat{C}$ in the curvature of an interface can be written in terms of strains as

$$\Delta\hat{C} = 2\frac{\partial\hat{e}_{sn}}{\partial s} - \frac{\partial\hat{e}_{ss}}{\partial n} - \hat{C}\hat{e}_{nn}, \tag{47}$$

where the dimensionless curvature $\hat{C}$ is defined such that $\hat{C}\hat{s} = -\partial\hat{n}/\partial\hat{s}$. Markenscoff and Wheeler (1996) showed that the contact condition (37) implies that

$$\Delta\hat{C}^1 - \Delta\hat{C}^2 = \hat{C}'(\hat{u}_s^1 - \hat{u}_s^2) \quad \text{in} \quad \Gamma_c. \tag{48}$$

They were also able to express this condition in terms of strains only in the form

$$\frac{d}{d\hat{s}}\left(\frac{\Delta\hat{C}^1 - \Delta\hat{C}^2}{\hat{C}'}\right) = \hat{e}_{ss}^1 - \hat{e}_{ss}^2 \quad \text{in} \quad \Gamma_c. \tag{49}$$

In the present problem, $\hat{\sigma}_{ns}^i$ and hence $\hat{e}_{ns}^i$ are zero throughout Γ_0, from (36), (42), so (47) simplifies to

$$\Delta\hat{C} = -\frac{\partial\hat{e}_{ss}}{\partial n} - \hat{C}\hat{e}_{nn} \quad \text{in} \quad \Gamma_0. \tag{50}$$

Substituting for the strains from (30), we have

$$\Delta\hat{C}^i = -\frac{1}{2(\kappa^i+1)}\frac{\partial\hat{\sigma}_{ss}^i}{\partial\hat{n}} + \frac{(3-\kappa^i)}{8(\kappa^i+1)}\left(\frac{\partial\hat{\sigma}_{ss}^i}{\partial\hat{n}} + \frac{\partial\hat{\sigma}_{nn}^i}{\partial\hat{n}}\right) - \frac{\partial\hat{T}^i}{\partial\hat{n}} - \hat{C}\left(\frac{\hat{\sigma}_{nn}^i}{2(\kappa^i+1)} - \frac{(3-\kappa^i)}{8(\kappa^i+1)} \times (\hat{\sigma}_{ss}^i + \hat{\sigma}_{nn}^i) + \hat{T}^i\right) \quad \text{in} \quad \Gamma_0. \tag{51}$$

This expression can be simplified using the equilibrium equation

$$\frac{\partial\hat{\sigma}_{nn}^i}{\partial\hat{n}} = -\frac{\partial\hat{\sigma}_{ns}^i}{\partial\hat{s}} - \hat{C}(\hat{\sigma}_{ss}^i - \hat{\sigma}_{nn}^i) \tag{52}$$

(see, for example, Dundurs (1989), Eq. (21)). Substituting (52) into (51) and noting once more that $\hat{\sigma}_{ns}^i = 0$ in Γ_0, we obtain

$$\Delta\hat{C}^i = -\frac{1}{8}\frac{\partial\hat{\sigma}_{ss}^i}{\partial\hat{n}} - \frac{\partial\hat{T}^i}{\partial\hat{n}} + \hat{C}\left(\frac{(2-\kappa^i)\hat{\sigma}_{nn}^i}{8(\kappa^i+1)} - \hat{T}^i\right) \quad \text{in} \quad \Gamma_0. \tag{53}$$

6 Contact Along a Straight Boundary

In general, the terms involving $\hat{C}$ in Eq. (53) will introduce a dependence on κ^i into the problem. However, for the special case where the initial contact boundary is straight, $\hat{C} = 0$ and (53) reduces to

$$\Delta\hat{C}^i = \frac{\partial^2\hat{u}_n^i}{\partial\hat{s}^2} = -\frac{1}{8}\frac{\partial\hat{\sigma}_{ss}^i}{\partial\hat{n}} - \frac{\partial\hat{T}^i}{\partial\hat{n}} \quad \text{in} \quad \Gamma_0. \tag{54}$$

We can therefore integrate along the straight line Γ_0 to obtain

$$\hat{u}_n^1 - \hat{M}\hat{u}_n^2 = -\int\int\left(\frac{1}{8}\left[\frac{\partial\hat{\sigma}_{ss}^1}{\partial\hat{n}} + \hat{M}\frac{\partial\hat{\sigma}_{ss}^2}{\partial\hat{n}}\right] + \left[\frac{\partial\hat{T}^1}{\partial\hat{n}} + \hat{M}\frac{\partial\hat{T}^2}{\partial\hat{n}}\right]\right)dsds \tag{55}$$

in Γ_0. The integral in (55) will contain two arbitrary constants of integration corresponding to rigid-body translation and rotation.

If we use (55) to substitute for $\hat{u}_n^1 - \hat{M}\hat{u}_n^2$ in (37), (43), the entire problem of Fig. 1 can be stated in terms of the two functions $\hat{T}^i$, $\hat{\sigma}^i$, using the boundary conditions (34) – (43) and the governing Eqs. (27), (28), (33). It follows that the solution depends upon the material properties only through the dimensionless parameters $\hat{M}$, $\hat{\eta}$, $\hat{K}$, $\hat{k}$, $\hat{h}$, $\hat{h}_c$, $\hat{V}$. In the steady-state problem, there is no dependence on $\hat{k}$ and if the thermal boundary conditions in $\Gamma - \Gamma_c$ are all of the form $q = 0$ or $T = 0$, the dependence on material properties is further reduced to the set $\hat{M}$, $\hat{\eta}$, $\hat{K}$, $\hat{V}$.

6.1 Similar Materials. If the two materials are similar, we have $\hat{M} = \hat{\eta} = \hat{K} = \hat{k} = 1$ and the problem depends only on $\hat{h}$, $\hat{h}_c$, $\hat{V}$. Remarkably, this reduces to dependence on the single dimensionless group $\hat{V}$ in the case where the thermal boundary conditions in $\Gamma - \Gamma_c$ are all of the form $q = 0$ or $T = 0$. This remains true in the case of transient problems.

6.2 One Material a Rigid Nonconductor. If material 2 is a rigid nonconductor, we have $\hat{M} = 0$, $\hat{\eta} = \hat{K} = \hat{k} = \infty$ and once again the problem depends on the single parameter $\hat{V}$ for homogeneous temperature or heat flux boundary conditions in $\Gamma - \Gamma_c$.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

6.3 Thermoelastic Contact Stability. Dow and Burton (1972) were the first to demonstrate that a problem defined by the frictional heating boundary condition (1) would be unstable if the sliding speed exceeded a certain critical value. They also pioneered a perturbation method for determining the critical speed V_0 in which they examine the conditions under which a small perturbation in the stress and temperature fields can grow exponentially with time. This method has since been used extensively for determining critical speed in more complex systems, both in analytical and numerical implementations (see, for example, Lee and Barber, 1993; Du et al., 1997).

Burton's perturbation method is essentially equivalent to a solution of the transient thermoelastic contact problem of Fig. 1—indeed, the general transient problem can in principle be written down as an eigenfunction expansion using the eigenmodes of Burton's solution. Thus, the above arguments carry over to the stability of any solutions obtained. In particular, we conclude that if two bodies of similar materials make contact at a plane boundary or one body makes contact with a rigid nonconducting plane surface, the stability of a given steady-state or transient solution can be defined in terms of a single dimensionless critical speed

$$\hat{V}_0 = \frac{fV_0\mu\eta a}{K(\kappa + 1)}, \tag{56}$$

for homogeneous temperature or heat flux boundary conditions in $\Gamma - \Gamma_c$.

This result is of significant value in the design of sliding systems against thermoelastic instability. In many cases, the design configuration is too complex to make a full numerical solution of the problem practicable, but using equation (56), we can conclude that the critical speed will be increased (and hence the system made more stable) by any change in material properties or dimensions that reduces the value of $\hat{V}_0$ for given V_0; i.e., that reduces

$$\frac{f\mu\eta a}{K(\kappa + 1)}.$$

7 Conclusions

We have demonstrated that the extent of the contact area Γ_c in the thermoelastic contact problem of Fig. 1 remains invariant under proportional loading if $\Gamma_c \subseteq \Gamma_0$; i.e., if the final contact area is included within that occurring at zero applied load. It then also follows that the temperature, stress, and displacement fields are all linear with the applied tractions.

For the two-dimensional problem with sliding at a common straight boundary, we have established a reduction in the dependence of the solution on the properties of the materials. In particular, we find that for problems involving two similar materials or the contact of a single deformable body against a rigid plane surface, the solution depends only on a single parameter, which is a dimensionless sliding speed $\hat{V}$. The stability of the solution is also determined by this parameter.

Since most practical problems of this class require numerical solution, these results extend the generality of such solutions and permit conclusions to be drawn about the effect of changes in material properties on stability behavior.

Acknowledgments

The authors are pleased to acknowledge support from the Ford Motor Company and from the National Science Foundation under contract number CMS-9619527. They are also grateful to Mr. M. Ciavarella of the Politecnico di Bari for some helpful contributions to the calculations in Section 5.

References

Azarkhin, A., and Barber, J. R., 1986, "Thermoelastic Instability for the Transient Contact Problem of Two Sliding Half-Planes," ASME JOURNAL OF APPLIED MECHANICS, Vol. 53, pp. 565–572.

Barber, J. R., 1969, "Thermoelastic instabilities in the sliding of conforming solids," *Proc. Roy. Soc., Lond.*, Vol. A312, pp. 381–394.

Barber, J. R., 1976, "Some thermoelastic contact problems involving frictional heating," *Q. J. Mech. Appl. Math.*, Vol. 29, pp. 1–13.

Burton, R. A., Kilaparti, S. R., and Nerlikar, V., 1973, "A limiting stationary configuration with partially contacting surfaces," *Wear*, Vol. 24, pp. 199–206.

Dow, T. A., and Burton, R. A., 1972, "Thermoelastic instability of sliding contact in the absence of wear," *Wear*, Vol. 19, pp. 315–328.

Du, Shuqin, Zagrodzki, P., Barber, J. R., and Hulbert, G. M., 1997, "Finite element analysis of frictionally-excited thermoelastic instability," *J. Thermal Stresses*, Vol. 20, pp. 185–201.

Dundurs, J., 1974, "Distortion of a body caused by free thermal expansion," *Mech. Res. Comm.*, Vol. 1, pp. 121–124.

Dundurs, J., 1975, "Properties of elastic bodies in contact," *The Mechanics of the Contact between Deformable bodies*, A. D. de Pater and J. J. Kalker eds., Delft University Press, pp. 54–66.

Dundurs, J., 1989, "Cavities vis-à-vis Rigid Inclusions and Some Related General Results in Plane Elasticity," ASME JOURNAL OF APPLIED MECHANICS, Vol. 56, pp. 786–790.

Dundurs, J., and Stippes, M., 1970, "Role of Elastic Constants in Certain Contact Problems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 37, pp. 965–970.

Keer, L. M., Dundurs, J., and Tsai, K. C., 1972, "Problems Involving a Receding Contact Between a Layer and a Half-Space," ASME JOURNAL OF APPLIED MECHANICS, Vol. 39, pp. 1115–1120.

Lee, Kwangjin, and Barber, J. R., 1993, "Frictionally-Excited Thermoelastic Instability in Automotive Disk Brakes," ASME *Journal of Tribology*, Vol. 115, pp. 607–614.

Love, A. E. H., 1927, *A Treatise on the Mathematical Theory of Elasticity*, 4th Ed., Cambridge University Press, Cambridge, UK, Section 156A.

Markenscoff, X., and Wheeler, L., 1996, "On conditions at an interface between two materials in plane deformation," *J. Elasticity*, Vol. 45, pp. 33–44.

Zagrodzki, P., 1990, "Analysis of thermomechanical phenomena in multidisc clutches and brakes," *Wear*, Vol. 140, pp. 291–308.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Two Interpretations of Rigidity in Rigid-Body Collisions

A. Chatterjee
Department of Engineering Science and Mechanics,
The Pennsylvania State University,
University, PA 16802
e-mail: anindya@crash.esm.psu.edu

A. Ruina
Department of Theoretical and Applied Mechanics,
Cornell University,
Ithaca, NY 14853
e-mail: ruina@cornell.edu

We distinguish between, and discuss the applicability of, two levels of rigidity in rigid-body collision modeling. For rigidity in the strong force-response sense, collisional contact deformations must be highly localized. The bodies then move according to second-order rigid-body mechanics during the collision. Incremental collision laws and most collision models using continuum mechanics for the contact region depend on force-response rigidity. For rigidity in the weaker impulse-response sense, the deformations need not be localized but displacements during the collision need to be small everywhere. Only the time-integrated rigid-body equations, involving before-collision and after-collision velocities, then need apply. Although a force-response rigid body is also impulse-response rigid the converse is not true. Algebraic collision laws depend only on impulse-response rigidity. Elastic vibration models of collisions are also generally consistent with impulse-response rigidity.

1 Introduction

Modeling collisions, the brief strong contact interactions of solid bodies, is of interest both for applications including robotics and general multibody dynamics simulations and to help complete classical mechanics, which does not include a settled approach to constitutive laws for contact. In dynamic models of mechanical systems, an extremely popular and useful idealization of a solid object is as a *rigid body*. The world of ideal rigid bodies has clear and well-defined rules for how objects move under the action of forces and moments, as well as how ideal constraints like perfect rolling, frictionless sliding, and frictionless joints restrict the motions of systems of objects. To complete the model world of ideal rigid bodies, rules are required for frictional contact, dissipative rolling, and (the topic of this paper) collisional contact.

The effect of collisional contact on motion is determined by the momentum balance equations and collisional constitutive relations (also called collision laws). But many questions are still not settled: how general, sensible collision laws might be constructed; how good the basis is for any existing law; and how useful the predictions are of these or any other laws in this world of not-truly-rigid bodies.

One set of important questions concerns the role of rigidity in rigid-body collision mechanics. We aim here to help clarify the applicability of two different notions of rigidity in collisions.

2 Deformations and Displacements During Collisions

A collision of two solid objects is shown schematically in Fig. 1. There is a small region of intense deformation near the initial contact point C. Outside but near this region is a reference point R. A typical point in the body is shown as P. The issue at hand is to what extent the points C, R, and P in one body may be viewed as rigidly linked.

Since the contact forces of colliding, nominally nondeforming, "rigid" bodies are governed by their deformations, some confusion naturally follows from the oxymoronic phrase "rigid-body collisions." One approach to resolving the resulting ambiguities is to consider a sequence of approximating systems having high stiffness and/or damping (see Brogliato, 1996, and references therein). The limiting case of infinite stiffness and/or damping gives a possible realization of a "rigid" object, and a way to develop rules for the behavior of such rigid objects.

A slightly different view is that rigidity itself is an approximate description of real bodies that are suitably stiff. In this view, the only well-defined aspects of rigid-body mechanics are those aspects of deformable body mechanics that do not depend on how limits of infinite stiffness and/or viscosity are approached. Features of the mechanics of deformable bodies that remain dependent on the details of the deformations, even as the deformations vanish with increasing stiffness, need further resolution by the introduction of additional constitutive assumptions beyond just "rigidity." In particular, "rigid-body" collision models are completed by introducing constitutive assumptions for the net collisional outcome, for contact deformations, or for whole-body deformation. Despite the implied deformations in each case, some notion of rigidity can still be retained.

Consider, for example, the Hertzian contact between two frictionless elastic spheres pressed together with remotely applied, equal, and opposite forces (possibly including smooth D'Alembert inertial terms) of magnitude F (for a detailed solution, see Johnson, 1985). Important features of such contact interaction are: (i) the contact region is small compared to the overall dimensions of the contacting bodies, and (ii) the contact stresses σ_{ij}, and the associated deformations, decay rapidly at large distances from the contact region. For spheres and other nonconforming three-dimensional elastic contact problems stresses decay like $1/r^2$, where r is the distance from the contact region. Large portions of the contacting spheres move with relatively smaller deformation. As a result, the net relative displacement $\delta = \Delta\mathbf{u}$ between typical points P, say the centers of the spheres, can be accurately calculated *without* knowledge of exactly where and how the remote forces (i.e., F) are applied on the spheres. For three-dimensional contact that is sufficiently smooth in time, the relative displacement between pairs of points that are not in the contact-deformation regions (say points R and P) can be accurately calculated by treating the spheres as rigid everywhere except in a small localized contact region governed by the Hertz relations.

In general point-contact problems, similar qualitative features are expected to that of the sphere for a useful range of material properties and of objects. The key ideas are that (a) the contact

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Mar. 18, 1996; final revision, July 15, 1998. Associate Technical Editor: S. W. Shaw.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

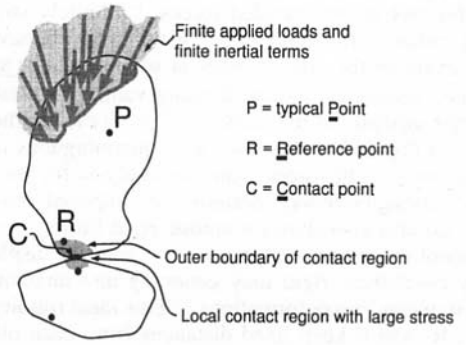


Fig. 1 Schematic diagram of generic three-dimensional collision. At issue is to what extent the points C, R, and P on one body may be regarded as rigidly linked.

stresses are large and decay rapidly with distance from the contact region; (b) the contact region stresses and deformations depend only on the net resultant contact force (and possibly its history—see, e.g., Mindlin and Deresiewicz, 1953); (c) there is an intermediate-sized region inside which the interaction is pseudostatic ($\sigma_{ij,j} \gg \rho\ddot{u}_i$, where ρ is density), and outside which the object has small deformations; and (d) the true contact region and intermediate size region are much smaller than the colliding objects. The natural rigid approximation is that the connection between P and R is rigid and all deformation between R and C is pseudo-static.

Consider next the longitudinal contact of long bars. In this case the deformation in the bulk of the rods (i.e., terms like FL/AE where F is the force, E is Young's modulus, L is length, and A is the cross-sectional area) will usually dominate any contact deformation and the motions between points on the two rods can be well approximated without including any extra contact compliance. That is, the region from C to R may be represented as arbitrarily small and rigid and all deformation is between R and P.

Finally, in the intermediate case of Hertzian contact between cylinders in plane strain or disks in plane stress, the stress field decays slowly enough (like $1/r$) that the deformations are not localized near the contact region. The net relative displacement between the centers of the cylinders, say, cannot be determined without consideration of the global deformations characterized by the relative displacement of R and P (even though the strains between R and P are small compared to unity).

In the case of spheres the deformation between R and P can be neglected. In the case of longitudinal contact of long rods the deformation between R and C can be neglected. And in the case of disks and cylinders both contact and far-field deformations are important for determining motions of general points on the body. Yet in all three cases the relative displacements of typical points can be far less than the distance between them.

Motivated by the observations above, we examine qualitative features of collisions where the effects of localized contact deformations dominate over global deformations, and vice versa.

3 The Contact Point and Contact Region Deformation

We discuss here only single-point contact collisions where the material contact point C is the point of first contact of two bodies that are locally relatively convex, or nonconforming (see Fig. 1). Some of the ideas here are relevant to simultaneous collisions at multiple points[1] although, as explained clearly by Ivanov (1995), "simultaneous" collision behavior is often severely ill-posed. (Ivanov's paper refers to some early works in this area, including D'Alembert's (1743) demonstration that a detailed solution incorporating elastic deformations leads to a prediction different from considering a sequence of pairwise collisions.)

One might, unlike what is shown in Fig. 1, take the reference contact point R to be the position in space that the material point C would have if not for local contact stresses. For example, if the body is effectively rigid but for localized contact deformations, then point R might be chosen as the point in space that is the same distance from points interior to the body as was material point C before the collision.[2]

Since we assume a priori that the region of contact is small compared to the size of the body, impulse momentum relations are only slightly affected by which point inside or near it is used for the application of the collisional impulse. Consider a collision where a net impulse of magnitude P acts at a point C on a body of characteristic length L. The moment of the impulse about the body center of mass G typically has a magnitude comparable to PL. If the impulse were taken to act at a point a small distance δ from C, the correction to the impulsive moment about the center of mass is comparable to $P\delta$, which is negligible if $\delta \ll L$.

In the degenerate case where the impulse **P** is parallel to CG, such as when a uniform sphere is dropped vertically onto a horizontal floor, possibly with spin about a vertical axis, a term of magnitude $P\delta$ may not be small compared to the nominal impulsive moment about G (Brach (1991) attributes this example to Horak). If terms of order $P\delta$ are deemed important then either the collision is not strictly treated as a point collision, or for consistency impulsive moments about the contact point need be considered. Here we ignore such terms because they are small, as described, for single-point collisions of stiff bodies.

So, with little loss of accuracy, the reference point R can be taken as any material point in the body that is slightly removed from the local contact region. One might think of R as representative of a point in the region surrounding the contact region that separates the series connection of contact and bulk deformations.

In the proof stage of this paper we found a discussion by Lord Rayleigh (1906, p. 257) about the location of the reference point (here called R) that is similar to the one here.

In any case, our definition of contact-region deformation is the relative displacement of the reference points on the two bodies (say R_1 and R_2) during the collision (see Fig. 2). If one of the bodies, say body 2, deforms much less than the other,

[1] Attempts to model multiple simultaneous collisions have been made by many authors, e.g., Pfeiffer and Glocker (1996).

[2] The extrapolation of the position of a material point based on neglecting local contact deformations is not always well defined, e.g., for two-dimensional elastic objects with Hertzian contact. We do not know of a rigorous argument that delineates when such a point R is or is not well defined in some suitable asymptotic sense.

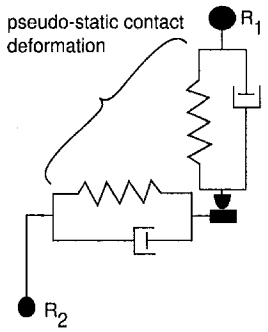


Fig. 2 Schematic diagram showing the deformations in the contact region. Contact deformation region acts as a massless, pseudo-static interaction mechanism between the reference points R_1 and R_2 on the two bodies. The interaction mechanism can be any passive mechanism including, e.g., spring/dashpot/friction elements or Hertz contact.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

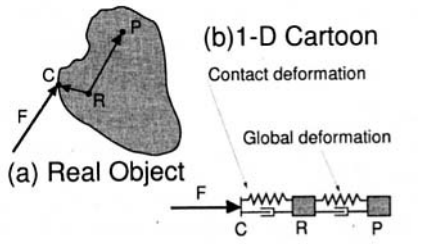


Fig. 3 (*a*) Force *F* acts at point C on an object—R is a reference point, and P is a typical point; (*b*) one-dimensional cartoon: between C and R is contact deformation and between R and P is deformation of the full body

then its initial contact point C_2 may be used for its reference point R_2. For this reason, a collision with a rigid wall serves our explanatory purposes and henceforth we take the contact region deformation as occuring between points R and C on one body.

Following the treatments of "approximating problems" by Brogliato (1996) and references therein, we illustrate the ideas here with the one-dimensional cartoon shown in Fig. 3(b) as well as special cases shown in Figs. 4(a–c). Figure 3(b) shows the cartoon idealization of a solid body. The contact point C and a nearby reference point R are separated by a massless contact deformation mechanism. The point R and point P are separated by the small-deformation mechanics of the bulk of the body.

4 Ideal Rigidity is Well Defined for Smooth Motions

In studying the dynamics of a "real" object as idealized in Fig. 3(a) the simplifying assumption of rigidity is that the change in distance between any pair of points on the body is zero for all time. Under the assumption of rigidity the velocity of *any* material point P relative to the material point C is given by the cross product of the object's angular velocity vector and the position vector from C to P, i.e., $\mathbf{v}_{P/C} = \boldsymbol{\omega} \times \mathbf{r}_{P/C}$. In one-dimensional rigidity, this reduces to $v_{P/C} = 0$.

Consider the cartoon in Fig. 4(a). If a smooth, bounded force $F(t)$ is applied at point C, then the acceleration of the block, i.e., point R (or P), is F/m. The acceleration of point C is *not* F/m, due to the deformation of the spring. However, the limiting case of infinitely stiff spring and dashpot is well defined. The motions of both points, C and R, may be treated as identical if we ignore some vanishingly small-amplitude phenomena. In energy calculations, the work done by the force will be almost totally converted to kinetic energy of the block, with a vanishingly small portion dissipated in the dashpot. In this sense, for smooth motions under bounded forces, "infinitely stiff" is a sufficient criterion for describing "ideal" rigid behavior. We need not examine the relative rates at which k and c approach infinity, i.e., quantities like the limiting value of $c/\sqrt{mk}$.

A similar analysis of the model in Fig. 4(c), and hence for the model in Fig. 3(b), also leads to an unambiguous infinitely stiff limit for smooth forces. Another analysis for the case of smoothly varying boundary motions (i.e., imposed velocities at boundary points) also shows a unique rigid limit.

For smoothly applied forces, or for smooth displacement boundary conditions, rigid may generally and unambiguously be taken to mean "no deformations." This ideal rigidity, where points C, R, and P keep fixed distances from each other, is a well-posed limit of high stiffness.

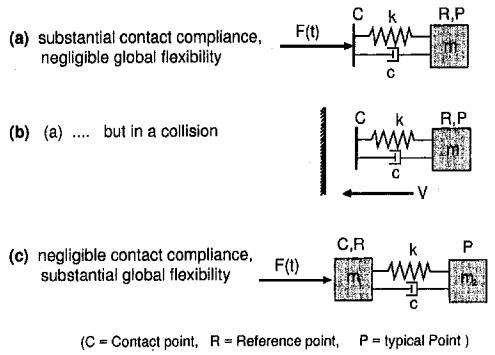


Fig. 4 (*a*) One-dimensional model with negligible global flexibility so that R and P are modeled as on the same rigid body but there is substantial contact compliance; (*b*) same model as "(*a*)" but in a collisional context; (*c*) one-dimensional model with negligible contact compliance and substantial global flexibility

5 Ideal Rigidity is Not Well Defined for Collisions

Now consider the system in Fig. 4(a) colliding with a rigid wall as in Fig. 4(b). In this case, there is a period of compression at the end of which the block is momentarily at rest, followed by a period of "restitution" or rebound during which the block accelerates away from the wall. At the end of compression, all the energy of the block has been absorbed by the spring/dashpot. Some of it subsequently returns to the block, while the rest is dissipated. If we now make the spring and dashpot stiffer, the interaction with the wall occurs over a shorter time. In the limiting case of $k \to \infty$, we have an instantaneous[3] "collision." However, the entire energy of the block still goes into the spring and dashpot at the end of the infinitely short compression phase, and what fraction of it is returned depends on the limiting value of $c/\sqrt{mk}$ (see, e.g., Chatterjee (1997) or Brogliato (1996) for slightly different treatments). Thus, there is no single, unambiguous "ideal rigid" limit for this model and hence for solid bodies in general.

In the case of collisions the limit of infinite stiffness—for which the distances between points C, R, and P stay fixed—does not lead to a unique outcome. That is, various degrees of rebound, energy dissipation, etc., are possible in the infinitely stiff limit. To predict such quantities, we need to make additional hypotheses about the collisional interaction.

6 Basic Assumptions

We assume, as do most authors, several features which seem well approximated by many collisional interactions of interest. All contact forces are in a small contact region; before and some time soon after the collision the velocities of points in the bodies are described with rigid-body dynamics relations; the configuration of the bodies does not change appreciably during the collision (i.e., the displacements of points in the bodies are small compared to the sizes of the bodies); and during collision the collisional forces are much bigger than other forces or $O(\omega^2)$ inertial forces.

These assumptions imply that the net change in linear momentum, angular momentum, and energy of the bodies can be determined by their pre and post-collision rigid-body velocities and angular velocities. Also, the contact interaction is summarized by the net resultant contact impulse, equally and oppositely applied to the contacting bodies at their contact points.

7 What is a Rigid-Body Collision Law?

Given a pair of colliding rigid bodies (or mechanisms composed of rigid bodies connected with frictionless geometric constraints) in known configurations and with known pre-collision velocities, and given the relevant parameters (which could be

[3] If the limiting value of k/c is finite, the interaction time does not actually go to zero. However, in this case deformations are infinitesimal, and only some infinitesimal restitution occurs over finite time.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

coefficients of friction, restitution, material properties, geometric information like local surface shape, etc.) a *collision law* predicts the contact impulse.

Linear and angular momentum balance relations are not part of the collision law but, given the contact impulse, do uniquely determine the post-collision velocities.

If instead of impulses one wants to express the collision law directly in terms of the motions of the bodies, then these motions must be chosen so as to satisfy the basic momentum relations.

8 Definitions of Rigidity in Collisions

8.1 Force-Response Rigidity. We refer to an object as effectively *force-response rigid* in a given collision if (a) a well-defined pseudo-static contact region (R to C) can be identified, and (b) bulk deformations between R and P are negligible at all times including *during* the collision. Force-response rigidity is the strongest sense in which a body may be considered as rigid during a collision. Away from the contact region, such an object behaves like an ideal rigid body moving under the action of a concentrated force acting at the contact point.

Since the collisional contact force is dominant, the dynamic behavior of a force-response rigid object is summarized by the linear relation between the contact force vector **F** and the acceleration $\mathbf{a}_R$ of the near-contact reference point

$$\mathbf{F} = \mathbf{M}\mathbf{a}_R \tag{1}$$

where **M** is a second-order tensor with units of mass, that includes effects of mass, rotational inertia, and kinematic constraints, and that expresses the net anisotropic inertia encountered by a force acting at that point[4] (this inertia tensor or "mass matrix" is well known—see, e.g., Batlle (1993), Smith and Liu (1992), Chatterjee (1997)). Given that real bodies do deform, the approximation being made in force-response rigidity is that the contact region relative displacements (between C and R) are much greater than any deformation-induced relative displacements between points removed from that region (e.g., R and P).

In principle one can test the accuracy of the force-response rigidity assumption. First assume the body is force-response rigid and calculate (using the micromechanics of the contact region, the initial velocities, and the mass tensor **M**) the time history of the net contact force. Then apply this force to a more realistic model of the body as a whole to check that the deformation-induced relative displacements between R and P in this second calculation are much smaller than those between C and R. In our one-dimensional cartoon this test corresponds to first using the model in Fig. 4(*b*) and then using the force history thus found on the model in Fig. 4(*c*) and checking that the R to P relative displacement in the second calculation is relatively much smaller than the C to R relative displacement.

If a body is well described as force-response rigid the net collision impulse can only be a function of material behavior in the contact deformation region, the relative velocities of the contact points before collision, and the mass tensor **M**. That is, the collision impulse cannot depend on any aspects of the shape or mass distribution of the object or mechanism that are not encompassed in **M** (see Chatterjee, 1997). Collision laws based on contact micromechanics mediating the interaction of otherwise rigid bodies (*incremental* collision laws) are based on the implicit assumption of force-response rigidity.

Numerical Example: For illustration, assume that the contact force history from the model in Fig. 4(*b*) is a constant force F over a time interval Δt with a net impulse $P = F\Delta t$. We now examine the behavior of the system in Fig. 4(*c*) for large stiffness and damping. In Fig. 5, the velocity of the mass m_1 is shown as a function of time. It is seen that for increasing k and c (for simplicity, holding $c/\sqrt{km}$ constant), the response of the system approaches that predicted by force-response rigidity. The reference point R on mass 1 follows the second order Newton-Euler dynamics relations based on the entire mass of the system: $F \approx ma_1$ applies (where a_1 is the acceleration of block 1). The common description of this situation is that the settling time for vibrations of the body as a whole are much less than the contact interaction time.

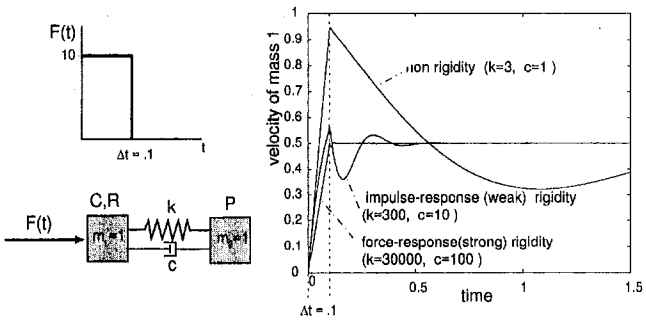


Fig. 5 Simulation of two mass system: $F = 10$, $\Delta t = 0.1$, $P = 1$, $m_1 = m_2 = 1$ and k, c as shown in figure

8.2 Impulse-Response Rigidity. We refer to an object as effectively *impulse-response rigid* in a given collision if (a) displacements are small compared to the dimensions of the object, everywhere and at all times during the collision, and (b) the object has well-defined rigid-body motions before and after the (brief) collision.

For such an object, integration of the deformable body equations of motion show that the net impulse vector transmitted in the collision is related to the net change in the velocity of the near-contact reference point, vector $\mathbf{v}_R$, by an impulse-momentum relation of the form

$$\mathbf{P} = \mathbf{M}\Delta\mathbf{v}_R, \tag{2}$$

where **M** is the same second-order tensor with units of mass, described earlier, that would be used for the bodies were they force-response rigid. In particle mechanics the momentum equation $P = M\Delta v$ is derived from $F = ma$. For the problem at hand, however, the integrated form of the rigid-body momentum equations Eq. (2) is accurate even for cases where the rigid-body differential equation Eq. (1) is inaccurate.

An object is rigid in the weaker impulse-response sense if it moves enough like a rigid body before the collision and soon enough after that its momentum, angular momentum, energy and the velocity of point R can be calculated accurately using rigid-body relations with the same configuration (position and orientation) after the collision as before. Impulse-response rigidity applies to a wider range of collisions than force-response rigidity.

In the cartoon model there is a range of k and c for which the transients decay on a time scale comparable to Δt. In this range of stiffness, the system is rigid in the weak (impulse-response) sense—on time scales that are long compared to the time scale of the impulse, *before* and *after* but *not during* the application of the impulse, the system moves essentially like a rigid body (see the "weak-rigidity" curve in Fig. 5). The net difference between the pre and post-collision rigid-body-like motions is accurately described by rigid-body impulse-momentum relations. In other words, the collisional behavior of the object is well described by $P = m\Delta v_1$ (where v_1 is the velocity of block 1) even though the equation $F = ma_1$ is inaccurate. After application of the impulse P and after transients die out, the velocity of each block is equal to $P/(m_1 + m_2)$, so the *net*

[4] The nonlinear ω^2 terms vanish in proportion to other inertial terms in the limit as the interaction time goes to zero and the interaction force goes to infinity.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

change in velocity *can* be found by approximating the system as a single rigid object.

If a body is only impulse-response rigid there is no reason to expect that collision laws based solely on the local deformation mechanism, the incoming velocity difference, and the mass tensor **M** should be accurate. Aspects of the body not encompassed in **M** are likely to be important determinants of the collisional impulse. But it is still true that the net jumps in velocities of all points are determined by the collisional impulse and that the mass tensor **M** helps in the expression of reasonable bounds on possible collisional impulses (Chatterjee and Ruina, 1998).

8.3 Nonrigidity. Finally we consider a case where rigidity does not even apply in the weak impulse-response sense.

If the spring and dashpot in Fig. 4(*c*) are soft, so that the time scale of the decaying transients (or the settling time) is long compared to the interaction time Δt, the system is not yet ready for rigid-body approximations until a time long after the contact period. We might be tempted to consider the system to be rigid in an *average* sense, in that it has well-defined *mean* rigid-body motions, before and after the application of the impulse, although the vibrations persist for a long time. However, for general two and three-dimensional bodies that are, say, vibrating for a long time, it may not be possible to define a "mean" equivalent rigid-body motion.

For example, consider two equal point masses joined by a spring of nonzero free length, moving in a plane. Let initial conditions be such that the center of mass is stationary, while the angular momentum of the system about its center of mass is H (a conserved quantity). Each point mass executes central force motion. Let the radial position of each point mass oscillate about its mean value with a small amplitude A. In that case, the *average* angular velocity of the system can be shown to be of the form $C_0 + C_1A^2$, for suitable constants C_0 and C_1. The small correction to the mean angular velocity produces large changes in orientation over long enough times. Thus, if two bodies with identical mass distribution are given identical impulses, and if the vibrations in them die out over widely different times, then eventually their orientations may be quite different. If the amplitude of vibrations is large (say, comparable to the size of the body), then the orientations of the vibrating and nonvibrating object may differ appreciably after a short time.

The previous discussion shows some of the angular momentum-related difficulties in two dimensions and three dimensions which do not appear in one dimension, if vibrations have large amplitudes and/or last long. Note that the position and velocity of the center of mass do not suffer such ambiguities: unlike angular momentum, linear momentum is the derivative of a function of the system's coordinates.

There is another way in which persistent vibrations can violate a rigid-body modeling approach (J. Jenkins, private communication). If we want only a smaller (finite or $O(1)$) time prediction, then small persistent vibrations will not lead to large errors. However, if the body has another collision before the vibrations are dead, then some vibrational kinetic energy may get transferred back to rigid-body modes. Thus, a model for the *second* collision will need to account for the vibrations of the object. In a general rigid-body dynamics simulation environment, vibrations need to die out on a time scale that is short compared to the typical time between collisions, i.e., the time scale of overall motions.

Thus, due to considerations of angular momentum for long-time predictions, and energy for short-time predictions where several collisions might occur sequentially, impulse-response rigidity is the weakest sense in which an object may be considered as rigid in general two-dimensional and three-dimensional collisions.

9 Algebraic, Incremental, and Full-Body Deformation Rigid-Body Collision Laws

There are three general approaches to finding the net resultant collisional impulse.

In *incremental* collision laws, rigid-body force acceleration and angular momentum balance equations are used to describe the dynamics *during* the collision. That is, incremental collision laws use force-response rigidity. The collision is modeled using a pseudo-static local-deformation calculation interacting with a mass matrix **M** (or a rigid-body dynamics calculation equivalent to using **M**).

Incremental laws can roughly be divided into two classes. In the first the contact region is described by a simplistic combination of simple elements such as springs and dashpots. In the second class the contact region is modeled using continuum mechanics. This class includes, for example, models based on Hertz contact with or without distributed friction.

Incremental rigid-body collision laws, such as Routh's law (Routh, 1897), Hertz-contact-based treatments (e.g., Maw et al., 1981; Jaeger, 1994), as well as more ad hoc spring-dashpot type approaches (e.g., Goyal et al., 1994) *all* depend on force-response rigidity, whether such an assumption is justified or not. Even when force-response rigidity may be an inaccurate description, these laws might still be used to calculate a plausible collisional impulse. That is, force-response based predictions using standard passive contact elements will respect basic constraints like non-negative energy dissipation, noninterpenetration and (if applicable) the Coulomb friction inequality at the impulse level. However, if the incremental model is not a good approximation of the actual contact mechanics, and/or if the body is not well approximated as force-response rigid, incremental laws should not be expected to make accurate predictions.

In *algebraic* laws, the impulse-momentum equations (Eq. (2) or equivalent rigid-body calculations) are used to summarize the mechanics of the bodies. The *net* interaction is predicted without paying attention to the detailed dynamic interaction during the collision. Algebraic laws are generalizations of the familiar one-dimensional Newton or Poisson laws based on a coefficient of restitution e. Algebraic rigid-body collision laws (such as Whittaker's (1944) or Kane and Levinson's (1985) collision law, Smith's (1991) law, the laws proposed in Brach's text (1991), or the new law proposed in Chatterjee and Ruina (1998)), assume impulse-response rigidity. Note that algebraic collision laws do not deny the existence of deformations during the collisions. They just do not pay attention to their details. As is commonly acknowledged (see, e.g., Smith 1991), any given algebraic rigid-body collision law should be looked upon as an approximate treatment.[5]

In *full-body-deformation* collision treatments the deformations of the full bodies are modeled during the collision. If it is assumed that at times before and soon after the collision the overall motions of the bodies have settled to something well described by rigid-body kinematics then these calculations may still be viewed as a means to finding the collisional impulse between two nominally rigid bodies. The well-known one-dimensional wave propagation model of longitudinally colliding elastic bars falls in this class. See also the model used by Stoianovici and Hurmuzlu (1996).

The algebraic and incremental approaches to calculating collisional impulses are sometimes called, respectively, "hard" and "soft" (Goyal et al., 1994; Walton, 1992). Algebraic laws are called "hard" because they pay no attention to the details of even the contact-region deformation. Incremental laws are called "soft" because they follow the details of some contact compliance model. The words "soft" and "hard" are somewhat inverted in meaning, however. "Soft" collision laws actually depend on the stronger force-response rigidity (the whole body, but for a small contact region, behaves like a rigid body

[5] This is not to say that special-purpose algebraic collision laws for special objects in particular configurations cannot be accurate (see, e.g., a special law for collisions between slender rods and massive objects presented in Hurmuzlu, 1998).

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

throughout the collision) while the "hard" collision laws depend only on the weaker impulse-response rigidity (rigid motions only before and soon after the collision).

The following examples may further clarify some of these collisional rigidity issues.

9.1 Routh's Law. Routh's law has received attention from many authors (e.g., Mayer, 1902; Plyavniyeks, 1970; Keller, 1986; Wang and Mason, 1992; Bhatt and Koechling, 1995; Batlle, 1996). In addition to force-response rigidity, this law assumes that the contact interaction is well described by a (nonzero) finite compliance in the normal direction along with *zero* compliance in the tangential direction. Errors in predictions from Routh's model can therefore arise from tangential compliance effects or other contact model defects, even for objects accurately modeled as force-response rigid.

9.2 Robots and Linkages. One might assume that colliding robots/linkages are rigid with ideal geometric constraints (see e.g., Bhatt and Koechling, 1993; Marghitu and Hurmuzlu, 1995; Mills and Nguyen, 1992). Under such an assumption, incremental rigid-body collision laws might be used. However, real machines are frequently made of long slender components which are not expected to be force-response rigid. Moreover, small imperfections like finite (nonzero) bearing clearances complicate matters further. To be effectively force-response rigid in a collision, a mechanism must have suitably rigid components *and* the bearing clearances and compliance effects have to be negligible compared to the relative displacements in the contact region. Real mechanisms in colliding contact will typically *not* be well approximated as force-response rigid. However, with low joint friction, impulse-response rigidity may be quite accurate (for these objects the tensor $\mathbf{M}$ is still second order, or 3×3, but depends on all the rigid bodies in the chain).

9.3 Examples of Nonlocal Deformations in Rigid-Body Collisions.

Example 1 (Elastic Vibrations): Global deformations associated with elastic vibration are expected to be important for bodies when the net contact interaction time is not large compared to the lowest natural vibration frequencies. Short contact times can come from conforming contact in shock loading or when a body is hit with a much smaller or stiffer body. On the other hand, when an object is much larger in one dimension than in others, as for rods, relatively large-time vibrational modes are expected. Accurate models might still be obtained by considering the small localized interaction in the contact region to be pseudostatic (i.e., using mechanisms like nonlinear springs, dashpots, and friction elements, all without inertia), as demonstrated for the case of slender rods colliding with a massive anvil by Stoianovici and Hurmuzlu (1996). They found that the coefficient of restitution depends strongly on the orientation of the rods because of the importance of bending vibrations and consequent multiple micro impacts for each macroscopic collision. Such behavior was not foreseen in several papers on general collisions where slender rods are used as "example" problems (Brach, 1989; Stronge, 1990; Wang and Mason, 1992). The treatments in these papers are not wrong for "rigid" rods in that they do not violate laws or principles of mechanics. The treatments are just inaccurate for these slender steel rods. One might think (Hurmuzlu, 1996) that such collisions should therefore not be treated as rigid-body collisions. However, several essential features of rigid-body mechanics are retained (Chatterjee and Ruina, 1997). Rather the specific models considered (Brach, 1989; Stronge, 1990; Wang and Mason, 1992) are implicitly and inappropriately based on force-response rigidity. However, models based on impulse-response rigidity (Hurmuzlu, 1998) or on full deformations (Stoianovici and Hurmuzlu, 1996) can still be usefully applied to determine the relation between the before-collision and after-collision rigid motions.

Example 2 (Disks): In Chatterjee (1997) and Calsamiglia et al. (1998), results are given of collision experiments with thin disks against a massive steel plate. Practically all rigid-body collision models, for the special case of sliding collisions in two dimensions, predict that the tangential component of impulse transmitted is equal to μ (coefficient of friction) times the normal component. This includes, for example, the models in Routh (1897), Whittaker (1944), or more recent models as in Smith (1991).[6] This idea is generally accepted as correct, particularly for spheres and disks, and has been verified for some cases (Maw et al., 1981). However, the new data (Chatterjee, 1997; Calsamiglia et al., 1998) indicate that for thin enough disks, *even for* collisions where the tangential motion is not reversed or stopped (i.e., in the "sliding range"), the ratio of tangential to normal impulses is not constant—it can vary by roughly a factor of two. Only in the limiting case of grazing incidence, for the disks studied, does the impulse ratio become approximately equal to the friction coefficient measured in separate experiments. The discrepancy between the standard frictional prediction and the experimental result may stem in part from the importance of the whole-body deformation during collision (see discussion in Chatterjee (1997) or Calsamiglia et al. (1998)). We remark that for the disks, normal restitution was roughly constant for all incidence angles, as usually expected, and in contrast to the slender rods mentioned earlier.

Again, the thin disk experiments do not negate the concept of a rigid-body collision law, but point out shortcomings in particular laws and the need for care in the use of the term "rigidity" when only impulse-response rigidity may apply.

Example 3 (Pseudo-Rigid Bodies): A special theoretical approach to collisions uses pseudo-rigid bodies (e.g., Cohen and Mac Sithigh, 1991), where the strain in a colliding object is uniform (i.e., the strain is a function of time but not position). Pseudo-rigid bodies are a special case of using a finite number of approximate modes to represent the global deformations in the objects. This approach does not allow the possibilities of dominant localized deformations (as in, e.g., collisions of spheres or other roundish, solid objects); of dominant stress waves (as in, e.g., the longitudinal collisions of long rods); and of slow bending vibrations (as in, e.g., the transverse collisions of slender rods, plates, etc.). In other words, the choice of mode shapes used in the pseudo-rigid-body approach may not be appropriate for many collisions; however, as in the case of incremental models, this approach provides a means to calculate a plausible collision impulse that will satisfy basic restrictions, whether or not the object being modeled is well approximated as pseudo-rigid. A pseudo-rigid body is not force-response rigid but might be impulse-response rigid if the vibrations are assumed to damp out soon after collision.

10 Summary

The prediction of a collision model for nominally rigid bodies depends on the momentum balance equations and constitutive laws. Depending on the degree to which rigid-body kinematics is followed by the real body, different types of constitutive laws are appropriate. The degree to which rigid-body kinematics is followed can be summarized by the extent to which the Newton-Euler equations for a rigid body describe the motion of a near-contact reference point in response to the collision contact force. If the Newton-Euler equations, exactly equivalent for collisions to Eq. (1), apply through the collision, then the body is *force-response* rigid. If rigid-body dynamics equations only accurately describe the net changes in velocity, as expressed by Eq.

[6] But not including Brach's approach (1991) where the ratio of impulses is not predicted but assumed to be an input parameter known a priori.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

(2), then the bodies are *impulse-response* rigid. Force-response bodies are impulse-response rigid, but impulse-response bodies may not be force-response rigid. Collision laws that use a micromechanical contact model interacting with rigid bodies assume force-response rigidity, whether the assumption is justified or not. Algebraic laws, or collisions calculated using bulk deformation, depend only on impulse-response rigidity.

If incremental collision laws are applied when force-response rigidity is not accurate, or if they use an inaccurate micromechanical model, then they can predict plausible collisional outcomes but should not be expected to be more or less accurate than algebraic collision laws.

Acknowledgments

The authors thank Jim Jenkins and Bernard Brogliato for informative discussions, and Don Conway for pointing out Lord Rayleigh's paper.

References

Batlle, J. A., 1993, "On Newton's and Poisson's Rules of Percussive Dynamics," ASME JOURNAL OF APPLIED MECHANICS, 60:376–381, Vol. 60, pp. 376–381.

Batlle, J. A., 1996, "Rough Balanced Collisions," ASME JOURNAL OF APPLIED MECHANICS, Vol. 63, pp. 168–172.

Bhatt, V., and Koechling, J., 1995, "Three-Dimensional Frictional Rigid-Body Impact," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 893–898.

Bhatt, V., and Koechling, J. C., 1993, "Incorporating frictional impacts in dynamic simulations of planar multi link chains," *Proceedings of the IASTED International Conference on Modelling and Simulation,* M. H. Hamza, ed., Pittsburgh, PA, pages 454–457.

Brach, R. M., 1989, "Rigid-body collisions," ASME JOURNAL OF APPLIED MECHANICS, Vol. 56, pp. 133–138.

Brach, R. M., 1991, *Mechanical Impact Dynamics: Rigid Body Collisions,* John Wiley and Sons, New York.

Brogliato, B., 1996, *Nonsmooth Impact Mechanics: Models, Dynamics and Control.* Springer, London, UK.

Calsamiglia, J., Kennedy, S., Chatterjee, A., Ruina, A., and Jenkins, J., 1998, "Anomalous Frictional Behavior in Collisions of Thin Disks," ASME JOURNAL OF APPLIED MECHANICS, in press.

Chatterjee, A., 1997, "Rigid Body Collisions: Some General Considerations, New Collision Laws, and Some Experimental Data. Ph.D. thesis, Cornell University, Ithaca, NY.

Chatterjee A., and Ruina, A., 1997, discussion of "A Critical Study of the Applicability of Rigid-Body Collision Theory," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 247–248.

Chatterjee, A. and Ruina, A., 1998, "A New Algebraic Rigid Body Collision Law Based on Impulse Space Considerations," ASME JOURNAL OF APPLIED MECHANICS, Vol. 65, pp. 939–951.

Cohen, H., and Mac Sithigh, G. P., 1991, "Impulsive Motions of Elastic Pseudo-Rigid Bodies," ASME JOURNAL OF APPLIED MECHANICS, Vol. 58, pp. 1042–1048.

D'Alembert, J., 1743, *Traite de Dynamique,* David, Paris.

Goyal, S. Pinson, E. N., and Sinden, F. W., 1994, "Simulation of dynamics of interacting rigid bodies including friction I: General problem and contact model," *Engineering with Computers,* Vol. 10, pp. 162–174.

Hurmuzlu, Y., 1998, "An Energy Based Coefficient of Restitution for Planar Impacts of Slender Bars With Massive External Surfaces," ASME JOURNAL OF APPLIED MECHANICS, Vol. 65, pp. 952–962.

Ivanov, A. P., 1995, "On multiple impact," *Journal of Applied Mathematics and Mechanics,* Vol. 59, No. 6, pp. 887–902.

Jaeger, J., 1994, "Oblique impact of similar bodies with circular contact," *Acta Mechanica,* Vol. 107, pp. 101–115.

Johnson, K. L., 1985, *Contact Mechanics,* Cambridge University Press, Cambridge, UK.

Kane, T. R., and Levinson, D. A., 1985, *Dynamics: Theory and Applications,* McGraw-Hill, New York.

Keller, J. B., 1986, "Impact With Friction," ASME JOURNAL OF APPLIED MECHANICS, Vol. 53, pp. 1–4.

Marghitu, D. B., and Hurmuzlu, Y., 1995, "Three-Dimensional Rigid-Body Collisions With Multiple Contact Points," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 725–732.

Maw, N., Barber, J. R., and Fawcett, J. N., 1981, "The role of elastic tangential compliance in oblique impact," *Journal of Lubrication Technology,* Vol. 103, pp. 74–80.

Mayer, A., 1902, "Über den zusammenstoßzweier körper unter berücksichtigung der gleitenden reibung." *Berichte über die Verhandlungen der königlich sächsischen Gesellschaft der Wissenschaften zu Leipzig,* Vol. 54, pp. 208–243.

Mills, J. K., and Nguyen, C. V., 1992, "Robotic Manipulator Collisions: Modeling and Simulation," ASME *Journal of Dynamic Systems, Measurement, and Control,* Vol. 114, pp. 650–659.

Mindlin, R. D., and Deresiewicz, H., 1953, "Elastic spheres in contact under varying oblique forces," ASME JOURNAL OF APPLIED MECHANICS, Vol. 20, pp. 327–344.

Pfeiffer F., and Glocker, C., 1996, *Dynamics of rigid body systems with unilateral constraints,* John Wiley and Sons, New York.

Plyavniyeks, V. Yu., 1970, "Three-dimensional collision of two bodies," *Vopr. Dinamiki i Prochnosti,* Vol. 20, pp. 75–88.

Routh, E. J., 1897, *Dynamics of a System of Rigid Bodies,* 6th Ed., Macmillan and Co., London.

Lord Rayleigh, 1906, "On the production of vibrations by forces of relatively long duration, with application to the theory of collisions," *Philosophical Magazine,* Series 6, Vol. II, pp. 283–291.

Smith, C. E., 1991, "Predicting Rebounds Using Rigid-Body Dynamics," ASME JOURNAL OF APPLIED MECHANICS, Vol. 58, pp. 754–758.

Smith, C. E., and Liu, Pao-Pao, 1992, "Coefficients of restitution," ASME JOURNAL OF APPLIED MECHANICS, Vol. 59, pp. 963–969.

Stoianovici, D., and Hurmuzlu, Y., 1996, "A Critical Study of the Applicability of Rigid-Body Collision Theory," ASME JOURNAL OF APPLIED MECHANICS, Vol. 63, pp. 307–316.

Stronge, W. J., 1990, "Rigid body collisions with friction," *Proceedings of the Royal Society of London,* Vol. A431, pp. 169–181.

Walton, O., 1992, *Particulate Two-Phase Flow,* M. C. Roco, ed., Butterworth-Heinemann, Boston, MA, Chapter 25.

Wang, Yu, and Mason, M. T., 1992, Two-Dimensional Rigid-Body Collisions With Friction," ASME JOURNAL OF APPLIED MECHANICS, Vol. 59, 635–642.

Whittaker, E. T., 1944, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies,* 4th Ed., Dover, New York.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Free and Forced Vibration of an Axially Moving String With an Arbitrary Velocity Profile

W. D. Zhu
University of North Dakota,
Grand Forks, ND
Assoc. Mem. ASME

B. Z. Guo
Beijing Institute of Technology,
Beijing, China

The exact response of an axially moving string with an arbitrary velocity profile is determined under general initial conditions and external excitation. In terms of the curvilinear characteristic coordinates, the time-dependent governing equation reduces to one with constant coefficients. The domain of interest in the characteristic coordinate plane is bounded by two monotonic curves with the same shape corresponding to the spatial boundaries of the string. Solutions of the transformed equation are derived for the free and forced responses. The amplitude of the free response is shown to be bounded under arbitrary variation of the transport speed in the subcritical regime. The unbounded displacement of the string predicted by Floquet theory for the spatially discretized models does not occur. The method is demonstrated on two illustrative examples with piecewise constant and sinusoidal accelerations.

1 Introduction

A wide class of flexible translating machine elements, including magnetic tapes, transmission belts, textile threads, band saws, and tramway cables, is commonly referred to as axially moving materials (Wickert and Mote, 1988). During operation, systems of this class often experience intermittent periods of axial acceleration and deceleration. For instance, in data storage devices, the high-speed magnetic tape that travels at a speed of 5 m/s can experience an axial acceleration of 50 m/s^2 (Wickert, 1995). In automotive engine accessory drives, the belt that drives the engine-mounted accessories undergoes periodic speed fluctuations by virtue of discrete engine firing cycles (Hwang et al., 1994). In the design and operation of these translating elements, an understanding of the effect of longitudinal acceleration on the transverse vibration is desirable.

Although research interest in the dynamic analysis of axially moving materials has increased in recent years, most studies were restricted to the case of constant transport velocity. Analysis of axially accelerating continua remains scarce. Basic characteristics of translating media with a constant speed include speed-dependent natural frequencies, complex eigenfunctions, and existence of a critical transport speed at which divergence instability occurs. For the case of variable transport velocity, the time-dependent acceleration components in the equation of motion render the solution a transient problem. Eigensolutions do not exist, and the response is not characterized by the natural modes of vibration. The development of a general understanding of the response and stability of an axially accelerating medium is heretofore hindered by the lack of an analytical technique to handle the time-dependent distributed system.

The equation of motion of an axially accelerating string was first derived by Miranker (1960). Replacing the time-dependent coefficients in the governing equation by time-averaged values shows that under harmonic end excitation, transport mass acceleration generates positive damping and deceleration negative damping (Mote, 1975). The dynamic behavior of the axially accelerating string under periodically varying transport speeds was recently examined by Pakdemirli and Batan (1993) and Pakdemirli et al. (1994). Parametric resonances characterized by unbounded amplitudes of the response were predicted based on the truncated models using the Floquet theory. When the transport speed varies slowly on the time scale of the natural periods of oscillation, a perturbation analysis following the asymptotic method of Krylov, Bogoliubov, and Mitropolsky was presented by Wickert (1995). All approaches resort to numerical and approximate techniques. General treatment for an arbitrarily prescribed velocity profile has not yet become available.

In this study, an exact analysis is developed for the response of an axially moving string with an arbitrary velocity profile under general excitation and initial conditions. Through transformation using the characteristic coordinates, the time-dependent equation of motion reduces to a standard second-order partial differential equation with constant coefficients. The resulting domain in the coordinate plane is bounded by two monotonic curves of the same shape. The solution technique for the transformed equation is presented. The analysis is exemplified through two sample problems with piecewise constant and sinusoidal acceleration profiles.

The boundedness of the displacement of the string, in the absence of exogenous forces, is established for all variable, subcritical transport speeds. In particular, when the transport velocity varies periodically, the governing equation possesses time-periodic coefficients. While unbounded displacements of the string were predicted based on the discretized models using the Floquet theory (Pakdemirli and Batan, 1993; Pakdemirli et al., 1994), they are absent in the distributed model examined herein.

2 Equation of Motion

Consider a uniform string of mass per unit length ρ and tension P, traveling at a variable speed $V(T)$, between two eyelets separated by distance L. The transverse displacement of a material particle that is instantaneously located at an axial position $0 < X < L$ is described by $U(X, T)$. The string is subjected to the distributed external force $F(X, T)$. In terms of the dimensionless parameters

$$x = X/L \quad u = U/L \quad t = T(P/\rho L^2)^{1/2}$$
$$f = FL/P \quad v = V(\rho/P)^{1/2}, \tag{1}$$

the linear transverse vibration of the string is described by the

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, May 28, 1996; final revision, Mar. 25, 1998. Associate Technical Editor: N. C. Perkins.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

governing equation with time-varying coefficients (Miranker, 1960)

$$u_{tt}(x, t) + 2v(t)u_{xt}(x, t) - [1 - v^2(t)]u_{xx}(x, t) + \dot{v}(t)u_x(x, t) = f(x, t), \quad 0 < x < 1, \quad (2a)$$

where the subscripts denote partial differentiation and $f(x, t)$ is the dimensionless external force. Without loss of generality, we assume homogeneous boundary conditions

$$u(0, t) = u(1, t) = 0 \quad (2b)$$

and arbitrary excitations at the boundaries are incorporated in $f(x, t)$ in (2a) through standard transformations (Meirovitch, 1966). The initial conditions are specified by

$$u(x, 0) = u_0(x) \quad u_t(x, 0) = u_1(x) \quad (2c)$$

where $u_0(x)$ and $u_1(x)$ are the initial displacement and velocity of the string. Relative to an inertial reference frame, each particle on the string experiences local u_{tt}, Coriolis $2vu_{xt}$, centripetal v^2u_{xx}, and tangential $\dot{v}u_x$ accelerations. The governing Eqs. (2a, b) can be cast in the structured operator form: $Mu_{tt} + Gu_t + (K_e + K_c + K_t)u = 0$ for an unsteady gyroscopic system (Wickert, 1995), where M is the inertia operator, G is the gyroscopic operator, and $K_e + K_c + K_t$ is the net stiffness operator. They are defined by

$$M = I \quad G = 2v\frac{\partial}{\partial x} \quad K_e = -\frac{\partial^2}{\partial x^2} \quad K_c = v^2\frac{\partial^2}{\partial x^2} \quad K_t = \dot{v}\frac{\partial}{\partial x} \quad (3)$$

where I is the identity operator. The stiffness operators in (3) include K_e that derives from tension, and K_c and K_t that derive from centripetal and tangential accelerations. With respect to the standard inner product over the spatial domain, M, K_e, and K_c are symmetric, while G and K_t are skew-symmetric. In what follows, we assume subcritical transport speed with $|v(t)| \le r < 1$, where r is a constant. The net stiffness operator is positive definite.

3 Method of Solution

The response of the axially moving string with an arbitrary velocity profile $v(t)$ is determined for general initial conditions and external excitation. Similar to the case of constant transport velocity, the solution of (2a)–(2c) is characterized by waves that propagate downstream and upstream at time-varying speeds $1 + v(t)$ and $1 - v(t)$, respectively, as measured by a stationary observer. Introduce the transformation

$$\begin{cases} \xi = \phi(x, t) = x - \int_0^t v(\tau)d\tau - t \\ \eta = \psi(x, t) = x - \int_0^t v(\tau)d\tau + t \end{cases} \quad (4)$$

where (ξ, η) are the characteristic coordinates and are curvilinear when $\dot{v}(t) \neq 0$. The correspondence between the two coordinate systems, (ξ, η) and (x, t), is one to one because the Jacobian

$$J(x, t) = \phi_x\psi_t - \phi_t\psi_x = 2 \quad (5)$$

does not vanish. Let $z(\xi, \eta)$ be the new dependent variable, that is,

$$z(\xi, \eta) = z(\phi(x, t), \psi(x, t)) = u(x, t). \quad (6)$$

By use of the chain rule and (4), (2a) is transformed to

$$Az_{\xi\xi} + 2Bz_{\xi\eta} + Cz_{\eta\eta} + Dz_\xi + Ez_\eta = F(\xi, \eta) \quad (7a)$$

where

$$A = \phi_t^2 + 2v(t)\phi_t\phi_x - [1 - v^2(t)]\phi_x^2 = 0$$

$$C = \psi_t^2 + 2v(t)\psi_t\psi_x - [1 - v^2(t)]\psi_x^2 = 0$$

$$B = \phi_t\psi_t + v(t)[\phi_t\psi_x + \psi_t\phi_x] - [1 - v^2(t)]\phi_x\psi_x = -2$$

$$D = \phi_{tt} + 2v(t)\phi_{xt} - [1 - v^2(t)]\phi_{xx} + \dot{v}(t)\phi_x = 0$$

$$E = \psi_{tt} + 2v(t)\psi_{xt} - [1 - v^2(t)]\psi_{xx} + \dot{v}(t)\psi_x = 0$$

$$Q(\xi, \eta) = f\left(\frac{\xi + \eta}{2} + \int_0^{(\eta - \xi)/2} v(\tau)d\tau, \frac{\eta - \xi}{2}\right). \quad (7b)$$

Hence the transformed Eq. (7a) becomes

$$z_{\xi\eta}(\xi, \eta) = -\frac{1}{4}Q(\xi, \eta). \quad (8)$$

The boundary and initial conditions for the transformed Eq. (8) are obtained next. When $x = 0$, from the first equation in (4), we have

$$\xi = -\int_0^t v(\tau)d\tau - t. \quad (9)$$

Because $\xi'(t) = -1 - v(t) < 0$, the function ξ is a strictly monotonically decreasing function of t. In addition, $-\infty < \xi \le 0$ when $t \ge 0$. Hence the inverse function of (9), denoted by $t = w(\xi)$, exists, and it is a strictly monotonically decreasing function of $\xi \in (-\infty, 0]$. By the second equation in (4), we have

$$\eta = -\int_0^t v(\tau)d\tau + t = \xi + 2t = \xi + 2w(\xi). \quad (10)$$

The line $x = 0$ in the upper half (x, t) plane, as shown in Fig. 1, corresponds to the curve L_1: $\eta = \xi + 2w(\xi)$ in the (ξ, η) plane, where $-\infty < \xi \le 0$ and $\eta \ge 0$. By the rule of differentiation for an inverse function,

$$w'(\xi) = -\frac{1}{\xi'(t)} = -\frac{1}{1 + v(t)}. \quad (11)$$

By (10) and (11), we have

$$\eta'(\xi) = 1 + 2w'(\xi) = -\frac{1 - v(t)}{1 + v(t)} < 0. \quad (12)$$

Hence L_1 is a strictly monotonically decreasing curve in the (ξ, η) plane. The boundary condition $u(0, t) = 0$ corresponds to

$$z(\xi, \eta) = 0 \quad \text{for all} \quad (\xi, \eta) \in L_1 \quad (13)$$

for the transformed equation. When $|v(t)| \le r < 1$, by (12), we have $-(1 + r)/(1 - r) \le \eta'(\xi) \le -(1 - r)/(1 + r)$. Hence L_1 is not asymptotic to any vertical line $\xi = \text{const}$ in the left half of the coordinate plane. Therefore $-\infty < \xi \le 0$. Similarly, when $x = 1$, from the first equation in (4), we obtain $\xi = 1 - \int_0^t v(\tau)d\tau - t$, whose inverse function is $t = w(\xi - 1)$, where $-\infty < \xi \le 1$ and $t \ge 0$. The line $x = 1$ in the upper half (x, t) plane corresponds to the curve L_2:

$$\eta = 1 - \int_0^t v(\tau)d\tau + t = \xi + 2t = \xi + 2w(\xi - 1) \quad (14)$$

in the (ξ, η) plane, where $-\infty < \xi \le 1$ and $\eta \ge 1$. It is easily seen that L_2 is a strictly monotonically decreasing curve in the (ξ, η) plane with the same shape as L_1, as depicted in Fig. 1. The boundary condition $u(1, t) = 0$ becomes

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

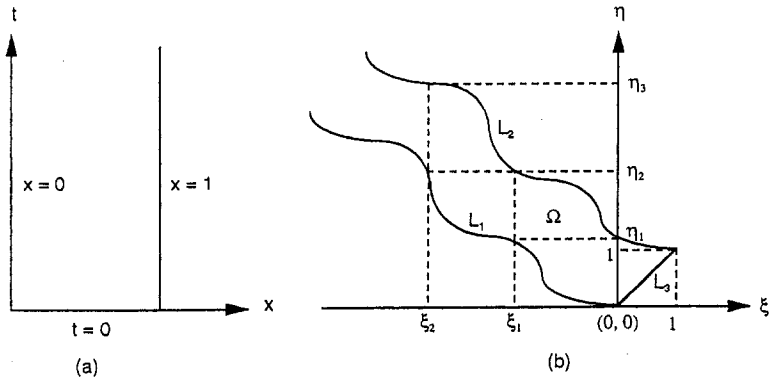


Fig. 1 Domains of the original (*a*) and transformed (*b*) systems; $x = 0$ corresponds to L_1, $x = 1$ to L_2, and $t = 0$ to L_3. Locations of ξ_n ($n = 1, 2$) and η_n ($n = 1, 2, 3$) are labeled.

$$z(\xi, \eta) = 0 \quad \text{for all} \quad (\xi, \eta) \in L_2. \tag{15}$$

When $t = 0$, from (4), we have $\xi = \eta = x$. The line segment $\{(x, t) | t = 0, x \in [0, 1]\}$ in the (x, t) plane corresponds to the line L_3: $\eta = \xi$ in the (ξ, η) plane, where $0 \le \xi \le 1$. The initial condition $u(x, 0) = u_0(x)$ becomes

$$z(\xi, \eta) = u_0(\xi) \quad \text{for all} \quad (\xi, \eta) \in L_3. \tag{16}$$

We differentiate (16) with respect to ξ by the chain rule and note that $\eta = \xi$ on L_3 to obtain

$$z_\xi(\xi, \xi) + z_\eta(\xi, \xi) = u_0'(\xi), \quad 0 \le \xi \le 1. \tag{17}$$

Next, we differentiate (6) with respect to t by the chain rule, and evaluate the resulting expression at $t = 0$. Because $t = 0$ corresponds to $\xi = \eta = x$, in this way we obtain

$$\begin{aligned} u_t(x, 0) &= -[1 + v(0)] z_\xi(\xi, \xi) + [1 - v(0)] z_\eta(\xi, \xi) \\ &= u_1(\xi). \end{aligned} \tag{18}$$

Multiply (17) by $1 + v(0)$ and add the expression to (18) to obtain

$$z_\eta(\xi, \xi) = \frac{1}{2} u_1(\xi) + \frac{1 + v(0)}{2} u_0'(\xi) \equiv z_1(\xi) \quad \text{for all} \quad (\xi, \eta) \in L_3. \tag{19}$$

Let Ω be the region enclosed by L_i ($i = 1, 2, 3$) in the (ξ, η) plane. It corresponds to the domain $(0, 1) \times \mathbf{R}_+$ in the (x, t) plane. The solution of (8) with the boundary and initial conditions (13), (15), (16), and (19) is determined next.

3.1 Free Vibration. Consider first the free response corresponding to $f(x, t) = 0$. Equation (8) becomes

$$z_{\xi\eta}(\xi, \eta) = 0. \tag{20}$$

The general solution of (20) is given by

$$z(\xi, \eta) = h(\xi) + g(\eta), \quad (\xi, \eta) \in \Omega, \tag{21}$$

where h and g are arbitrary functions of class C^2. Substituting (21) into (15) yields

$$h(\xi) = -g[\xi + 2w(\xi - 1)]. \tag{22}$$

Substituting (22) into (21) yields

$$z(\xi, \eta) = -g[\xi + 2w(\xi - 1)] + g(\eta), \quad (\xi, \eta) \in \Omega. \tag{23}$$

Substituting (23) into (13) yields

$$g[\xi + 2w(\xi - 1)] = g[\xi + 2w(\xi)], \quad -\infty < \xi \le 0. \tag{24}$$

By the theorem of the mean for integration (Protter and Morrey, 1977) we have by use of (11)

$$\begin{aligned} &[\xi + 2w(\xi - 1)] - [\xi + 2w(\xi)] \\ &\quad = 2[w(\xi - 1) - w(\xi)] = -2w'(\theta) = \frac{2}{1 + v}, \end{aligned} \tag{25}$$

where $\xi - 1 < \theta < \xi$. Hence for constant v, the function g is periodic with period $2/(1 + v)$, as given by (24) and (25).

Application of the initial conditions (19) and (16) to (23) yields

$$z_\eta(\xi, \xi) = g'(\xi) = z_1(\xi), \quad 0 \le \xi \le 1 \tag{26}$$

$$z(\xi, \xi) = -g[\xi + 2w(\xi - 1)] + g(\xi) = u_0(\xi), \quad 0 \le \xi \le 1. \tag{27}$$

Integration of (26), followed by substitution of the resulting expression in (27), yields

$$g(\xi) = \int_0^\xi z_1(\zeta) d\zeta, \quad 0 \le \xi \le 1 \tag{28}$$

and

$$g[\xi + 2w(\xi - 1)] = -u_0(\xi) + \int_0^\xi z_1(\zeta) d\zeta, \quad 0 \le \xi \le 1. \tag{29}$$

Note that the constant of integration in (28) and (29) is set to zero because it is canceled once one substitutes these expressions in (23). The principal observation here is that the solution $u(x, t) = z(\xi, \eta)$ can be readily determined from (23) for all $(x, t) \in (0, 1) \times \mathbf{R}_+$, once one knows $g(\eta)$ for all $\eta \ge 0$.

Let $(\xi_0, \eta_0) = (0, 0)$ be the origin in the (ξ, η) plane. One can define the coordinates of the grid points η_n and ξ_n shown in Fig. 1 for all $n \in \mathbf{N}$, through recursive use of the following equations defining the boundary curves of L_1 and L_2:

$$\eta_n = \xi_{n-1} + 2w(\xi_{n-1} - 1) \tag{30a}$$

$$\xi_n = p(\eta_n). \tag{30b}$$

Here $\xi = p(\eta)$ denotes the inverse function of $\eta = \xi + 2w(\xi)$, and it exists because L_1 is monotonic. For instance, we obtain $\eta_1 = 2w(-1)$ from (30*a*), and $\xi_1 = p(\eta_1)$ from (30*b*). Subsequent applications of (30*a*, *b*) yield $\eta_2 = \xi_1 + 2w(\xi_1 - 1)$ and $\xi_2 = p(\eta_2)$, respectively. Repeating the same procedure, one can determine (ξ_n, η_n) for all $n \ge 3$. Note that because L_1 and L_2 are not asymptotic to any vertical lines in the (ξ, η) plane, (ξ_n, η_n) exist for all $n \ge 1$.

The method to determine $g(\eta)$ for all $\eta \ge 0$ is outlined below. We determine $g(\eta)$ for all $\eta \in [0, 1]$ by use of (28). Because $\eta = \xi + 2w(\xi - 1) \in [1, \eta_1]$ when $0 \le \xi \le 1$, we can determine $g(\eta)$ for all $\eta \in [1, \eta_1]$ by use of (29), where

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$\xi = p(\eta - 1) + 1$. For $\eta \in [\eta_n, \eta_{n+1}]$, where $n \geq 1$, $g(\eta)$ is determined by a recursive process by use of (24), as illustrated in what follows. For instance, when $\eta \in [\eta_1, \eta_2]$, there exists a unique $\xi \in [\xi_1, \xi_0]$ on L_2, such that $\xi = p(\eta - 1) + 1$. By (24), we have $g[\eta = \xi + 2w(\xi - 1)] = g[\xi + 2w(\xi)]$, where the latter is known because $\xi + 2w(\xi) \in [\eta_0, \eta_1]$. Likewise, for any $\eta \in [\eta_2, \eta_3]$, there corresponds a unique $\xi = p(\eta - 1) + 1 \in [\xi_2, \xi_1]$. By (24), $g[\eta = \xi + 2w(\xi - 1)] = g[\xi + 2w(\xi)]$, where the latter has been determined earlier because $\xi + 2w(\xi) \in [\eta_1, \eta_2]$. This process can be continued to find the values of $g(\eta)$ for all $\eta \geq \eta_3$.

From the above procedure we can easily see that, for all $\eta \geq \eta_1$, $g(\eta)$ falls into its range corresponding to the domain $[\eta_0, \eta_1]$. As a result, an upper bound for the amplitude of the response is derived in the following theorem.

Theorem. When $f(x, t) = 0$ and $|v(t)| \leq r < 1$, where r is any constant the solution to (2a–c) is bounded by

$$|u(x, t)| \leq 2 \max \left\{ \max_{0 \leq \xi \leq 1} \left| \int_0^\xi \left[\frac{1}{2} u_1(\zeta) + \frac{1 + v(0)}{2} u_0'(\zeta) \right] d\zeta \right|, \max_{0 \leq \xi \leq 1} \left| u_0(\xi) - \int_0^\xi \left[\frac{1}{2} u_1(\zeta) + \frac{1 + v(0)}{2} u_0'(\zeta) \right] d\zeta \right| \right\} \quad (31)$$

for all $(x, t) \in (0, 1) \times \mathbf{R}_+$.

Proof: Recall that for all $\eta \geq 0$, the values of $g(\eta)$ lie in its range corresponding to the domain $[0, \eta_1]$. Hence by (23), (28), and (29) we have for all $(x, t) \in (0, 1) \times \mathbf{R}_+$:

$$\begin{aligned} |u(x, t)| = |z(\xi, \eta)| &\leq |g[\xi + 2w(\xi - 1)]| + |g(\eta)| \\ &\leq 2 \max_{0 \leq \eta \leq 2w(-1)} |g(\eta)| \\ &= 2 \max \left\{ \max_{0 \leq \xi \leq 1} \left| \int_0^\xi z_1(\zeta) d\zeta \right|, \max_{0 \leq \xi \leq 1} \left| u_0(\xi) - \int_0^\xi z_1(\zeta) d\zeta \right| \right\} \end{aligned} \quad (32)$$

where ξ and η are expressed in terms of x and t by (4). Substituting the expression for z_1 in (19) in (32) yields (31). (QED)

It is interesting to note that, for arbitrarily large acceleration $|\dot{v}(t)|$, the amplitude of the free response remains bounded whenever $|v(t)| \leq r < 1$. In particular, the solution is bounded, when $v(t)$ is any periodic function within the subcritical regime. For the case of periodic, piecewise constant (Pakdemirli and Batan, 1993) and sinusoidal (Pakdemirli et al., 1994) accelerations, parametric resonances characterized by unbounded displacements of the string were predicted at sufficiently high frequency based on the truncated models. They are not manifested in the present analysis for the distributed model.

3.2 Forced Vibration. The solution technique developed in the previous section can be easily extended to the case when $f(x, t) \neq 0$. The general solution to (8) is

$$z(\xi, \eta) = h(\xi) + g(\eta) - \frac{1}{4} \int_0^\xi \int_0^\eta Q(s, \tau) d\tau ds, \quad (\xi, \eta) \in \Omega \quad (33)$$

where h and g are arbitrary functions of class C^2. Applying the boundary condition (15) to (33) yields

$$z(\xi, \eta) = -g[\xi + 2w(\xi - 1)] + g(\eta) + \frac{1}{4} \int_0^\xi \int_\eta^{\xi + 2w(\xi - 1)} Q(s, \tau) d\tau ds, \quad (\xi, \eta) \in \Omega. \quad (34)$$

Applying the boundary condition (13) to (34) yields

$$g[\xi + 2w(\xi - 1)] = g[\xi + 2w(\xi)] + \frac{1}{4} \int_0^\xi \int_{\xi + 2w(\xi)}^{\xi + 2w(\xi - 1)} Q(s, \tau) d\tau ds, \quad -\infty \leq \xi \leq 0. \quad (35)$$

Applying the initial conditions (19) and (16) to (34) yields, respectively,

$$g'(\xi) = \frac{1}{4} \int_0^\xi Q(s, \xi) ds + z_1(\xi), \quad 0 \leq \xi \leq 1 \quad (36)$$

$$g[\xi + 2w(\xi - 1)] = g(\xi) + \frac{1}{4} \int_0^\xi \int_\xi^{\xi + 2w(\xi - 1)} Q(s, \tau) d\tau ds - u_0(\xi), \quad 0 \leq \xi \leq 1. \quad (37)$$

Similar to the case of free response, one can determine $g(\eta)$ for all $\eta \geq 0$. The grid points (ξ_n, η_n) for $n \in \mathbf{Z}^+$ are those defined in Section 3.1. Integration of (36) yields

$$g(\xi) = \int_0^\xi z_1(\zeta) d\zeta + \frac{1}{4} \int_0^\xi \int_0^\tau Q(s, \tau) ds d\tau, \quad 0 \leq \xi \leq 1 \quad (38)$$

where, for the same reason as in (28), we have set the constant of integration equal to zero in (38). By (38) one can determine $g(\eta)$ for all $\eta \in [0, 1]$. When $1 \leq \eta \leq 2w(-1)$, there exists unique $\xi \in [0, 1]$, such that $\eta = \xi + 2w(\xi - 1)$. Substituting (38) in (37) yields

$$g(\eta) = -u_0(\xi) + \int_0^\xi z_1(\zeta) d\zeta + \frac{1}{4} \int_0^\xi \int_0^\tau Q(s, \tau) ds d\tau + \frac{1}{4} \int_0^\xi \int_\xi^\eta Q(s, \tau) d\tau ds, \quad 1 \leq \eta \leq 2w(-1) \quad (39)$$

where $\xi = p(\eta - 1) + 1$. By (39) one can determine $g(\eta)$ for all $\eta \in [1, \eta_1]$. As observed in (38) and (39), both the initial conditions, $u_0(x)$ and $u_1(x)$, and the external force, $f(x, t)$, contribute to the response. Following the same recursive procedure as that outlined in Section 3.1, we can successively determine $g(\eta)$ for all $\eta \geq \eta_1$ by use of (35). Once $g(\eta)$ is known for all $\eta \geq 0$, the solution $u(x, t) = z(\xi, \eta)$ can be readily determined from (34). Due to the presence of the forcing integral in (35), $g(\eta)$ for all $\eta \geq 2w(-1)$ does not necessarily fall into its range corresponding to the domain $[0, 2w(-1)]$, as is the case for the free response. To obtain the forced response, evaluation of the forcing integrals in (34), (35), (38), and (39) becomes necessary.

4 Examples and Discussions

To determine the transient response at any position $x \in (0, 1)$ along the span during the interval $t \in [0, t_f]$, where t_f is an arbitrary final time, the domain Ω in the (ξ, η) plane is restricted within $-\xi_{\min} \leq \xi \leq 1$ and $0 \leq \eta \leq \eta_{\max}$, where

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\xi_{\min} = -\int_0^{t_f} v(\tau)d\tau - t_f, \quad \eta_{\max} = \xi_{\min} + 2w(\xi_{\min} - 1). \quad (40)$$

Consequently, it suffices to determine $g(\eta)$ for all $\eta \in [0, \eta_{\max}]$. In what follows the solution technique developed in Section 3 is demonstrated on two sample problems.

4.1 Piecewise Constant Acceleration. The free and forced responses of a constantly accelerating string are obtained in closed forms. The transport speed is prescribed by $v(t) = at$ for $0 \le t \le c$ and $v(t) = ac$ for $t \ge c$, where $c > 0$ and $|a| < 1/c$. Inclusion of an interval during which the string is under constant deceleration can be handled in a similar manner. The inverse function of (9) is

$$t = w(\xi) = \begin{cases} (-1 + \sqrt{1 - 2a\xi})/a, & -c - ac^2/2 \le \xi \le 0 \\ -\xi/(1 + ac) + ac^2/2(1 + ac), & \xi \le -c - ac^2/2. \end{cases} \quad (41)$$

The boundaries at $x = 0, 1$ correspond to the curves L_1 and L_2 in the (ξ, η) plane:

$$L_1: \quad \eta = \xi + 2w(\xi) = \begin{cases} \xi + 2(\sqrt{1 - 2a\xi} - 1)/a, & -c - ac^2/2 \le \xi \le 0 \\ (ac - 1)\xi/(ac + 1) + ac^2/(ac + 1), & \xi \le -c - ac^2/2 \end{cases} \quad (42)$$

$$L_2: \quad \eta = \xi + 2w(\xi - 1) = \begin{cases} \xi + 2(\sqrt{1 + 2a - 2a\xi} - 1)/a, & 1 - c - ac^2/2 \le \xi \le 1 \\ (ac - 1)\xi/(ac + 1) + (ac^2 + 2)/(ac + 1), & \xi \le 1 - c - ac^2/2. \end{cases} \quad (43)$$

The inverse functions of (42) and (43) are given, respectively, by

$$\xi = p(\eta) = \begin{cases} \eta - 2(1 - \sqrt{1 - 2a\eta})/a, & 0 \le \eta \le c - ac^2/2 \\ [(ac + 1)\eta - ac^2]/(ac - 1), & \eta \ge c - ac^2/2 \end{cases} \quad (44)$$

$$\xi = p(\eta - 1) + 1 = \begin{cases} \eta - 2(1 - \sqrt{1 - 2a\eta + 2a})/a, & 1 \le \eta \le 1 + c - ac^2/2 \\ [(ac + 1)\eta - 2 - ac^2)]/(ac - 1), & \eta \ge 1 + c - ac^2/2. \end{cases} \quad (45)$$

When $t_f \ge c$, the limiting value of η is obtained from (40) and (43):

$$\eta_{\max} = \frac{1}{2}ac^2 - (ac - 1)t_f + \frac{2}{ac + 1}. \quad (46)$$

Given $(\xi_0, \eta_0) = (0, 0)$, η_n $(n = 1, 2, \ldots, N)$ and ξ_n $(n = 1, 2, \ldots, N - 1)$, where N is the minimal integer satisfying $\eta_N \ge \eta_{\max}$, are determined using (30a, b), (43), and (44).

4.1.1 Free Response. Given the initial conditions $u_0(x) = b \sin \pi x$ and $u_1(x) = 0$, the exact response of the midpoint of the string for $a = 0.03$, $b = 0.05$, $c = 10$, and $t_f = 30$ is shown in Fig. 2 by a solid line. In this case, $\eta_{\max} = 15.82$ and $N = 9$, $g(\eta) = (b \sin \pi\eta)/2$ when $\eta \in [0, 1]$, and $g(\eta) = -(b \sin \pi\xi)/2$, where $\xi = p(\eta - 1) + 1$, when $\eta \in [1, \eta_1]$. The function $g(\eta)$ for all $\eta \in [\eta_n, \eta_{n+1}]$ $(n = 1, 2, \ldots, 8)$ can be successively determined by use of (24). As predicted by (31), the amplitude of the response is bounded by $|u(x, t)| \le b = 0.05$. When $a = 0$ with other parameters unchanged, the response shown by a dashed line reduces to that of the stationary string. Given the initial conditions for the stationary string, only the first mode of dimensionless frequency π is excited. During axial acceleration, the frequency of oscillation of the accelerating string gradually decreases, as compared to that of the stationary string.

4.1.2 Forced Response. Under the sinusoidal excitation, $u(1, t) = e \sin \alpha t$, at the right boundary, and zero initial conditions, the transient response of the string during the period of acceleration $[0, c]$ is determined. By (46), $\eta_{\max} = c - ac^2/2 + 2/(1 + ac)$. Through transformation $\tilde{u}(x, t) = u(x, t) - ex \sin \alpha t$, the function $\tilde{u}(x, t)$ satisfies (2a, b) with $f(x, t) = e\alpha^2 x \sin \alpha t - 2ae\alpha t \cos \alpha t$. The initial conditions for $\tilde{u}$ are $\tilde{u}(x, 0) = 0$ and $\tilde{u}_t(x, 0) = -e\alpha x$. Hence by (19), $z_1(x) = -e\alpha x/2$. By the last equation in (7b), we have

$$Q(\xi, \eta) = \frac{1}{2}\left[\xi + \eta + \frac{1}{4}a(\eta - \xi)^2\right] e\alpha^2 \sin \frac{\alpha}{2}(\eta - \xi) - ae\alpha(\eta - \xi) \cos \frac{\alpha}{2}(\eta - \xi). \quad (47)$$

Evaluating the single and double integrals in (38) and setting $\eta = \xi$ yields

$$g(\eta) = -\frac{1}{4}e\alpha\eta^2 + \frac{e}{4}\left(\frac{2a}{\alpha}\eta + \alpha\eta^2\right) - e\left(\frac{a}{\alpha^2} + \frac{1}{2}\eta + \frac{a}{8}\eta^2\right) \sin \frac{\alpha}{2}\eta \quad (48)$$

where $0 \le \eta \le 1$. Likewise, by (39), we have for $1 \le \eta \le 2w(-1)$

$$g(\eta) = -\frac{1}{4}e\alpha\xi^2 + \frac{ae}{2\alpha}\xi + \frac{e\alpha}{4}\xi^2 - e\left(\frac{1}{2}\eta + \frac{1}{8}a\eta^2 + \frac{a}{\alpha^2}\right) \sin \frac{\alpha}{2}\eta + e\left[\frac{a}{\alpha^2} + \frac{1}{2}(\xi + \eta) + \frac{a}{8}(\eta - \xi)^2\right] \sin \frac{\alpha}{2}(\eta - \xi), \quad (49)$$

where $\xi = p(\eta - 1) + 1$. For $\eta \in [\eta_n, \eta_{n+1}]$, where $n = 1, 2, \ldots, N - 1$, there exists unique $\xi = p(\eta - 1) + 1$ in $[\xi_n, \xi_{n-1}]$. Let $\tilde{\eta} = \xi + 2w(\xi)$, we have $\tilde{\eta} \in [\eta_{n-1}, \eta_n]$. Then (35) becomes

$$g(\eta) = g(\tilde{\eta}) + \tfrac{1}{4}\phi(\xi, \eta) - \tfrac{1}{4}\phi(\xi, \tilde{\eta}) \quad (50a)$$

where

$$\phi(\xi, \eta) = \int_0^{\xi}\int_0^{\eta} Q(s, \tau)d\tau ds = e\left[\frac{4a}{\alpha^2} + 2(\xi + \eta) + \frac{a}{2}(\eta - \xi)^2\right] \sin \frac{\alpha}{2}(\eta - \xi)$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

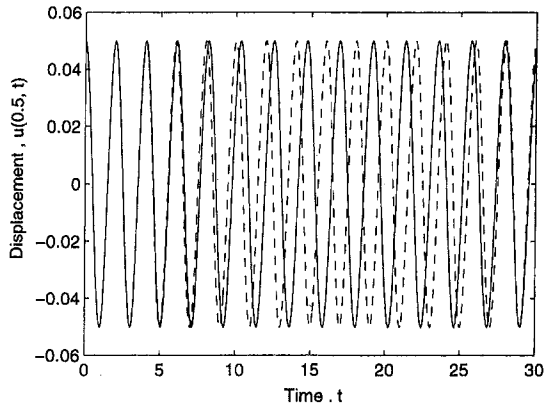


Fig. 2 Free response of the midpoint of the string to an initial displacement $u_0(x) = 0.05 \sin \pi x$; $t_f = 30$. Solid line, accelerating string, $a = 0.03$ and $c = 10$; dashed line, stationary string, $a = 0$.

$$+ e\left(\frac{4a}{\alpha^2} + 2\xi + \frac{a}{2}\xi^2\right) \sin \frac{\alpha}{2}\xi$$

$$- e\left(\frac{4a}{\alpha^2} + 2\eta + \frac{a}{2}\eta^2\right) \sin \frac{\alpha}{2}\eta. \quad (50b)$$

The function $g(\eta)$ for $\eta \in [\eta_0, \eta_1]$ is determined from (48) and (49). The function $g(\eta)$ for $\eta \in [\eta_n, \eta_{n+1}]$ $(n = 1, 2, \ldots, N - 1)$ is subsequently determined through n iterations by use of (50a, b). The solution $u(x, t)$ for all $(x, t) \in (0, 1) \times [0, c]$ is obtained by use of (34),

$$u(x, t) = z(\xi, \eta) + ex \sin \alpha t = -g(\hat{\eta}) + g(\eta) + \tfrac{1}{4}\phi(\xi, \hat{\eta}) - \tfrac{1}{4}\phi(\xi, \eta) + ex \sin \alpha t, \quad (51)$$

where ξ and η are expressed in terms of x and t by (4) and $\hat{\eta} = \xi + 2w(\xi - 1)$.

When $a = 0.05$, $c = 10$, $e = 0.01$, and $\alpha = 10$, the exact response of the midpoint of the string is shown in Fig. 3 by a solid line. By (46), $\eta_{\max} = 8.83$ and $N = 6$. Because the initial disturbance from the right end travels upstream at speed $1 - at$, the response of the midpoint of the string remains zero until the disturbance reaches it at $t = 0.5064$. When $a = 0$ with other parameters unchanged, the response shown by a dashed line reduces to that of the stationary string and is in full agreement with that predicted by the standard Green's function method.

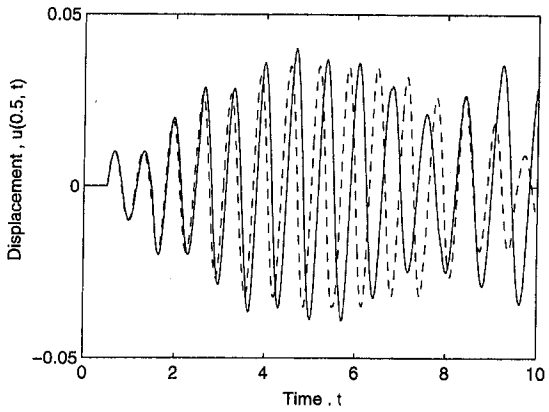


Fig. 3 Forced response of the midpoint of the string to harmonic end excitation $u(1, t) = 0.01 \sin 10t$; $t_f = 10$. Solid line, accelerating string, $a = 0.05$ and $c = 10$; dashed line, stationary string, $a = 0$.

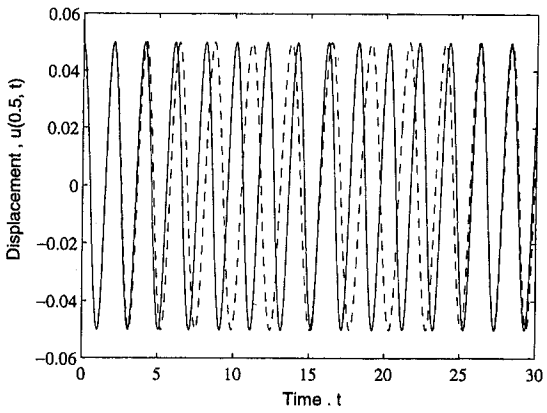


Fig. 4 Free response of the midpoint of the string under sinusoidally varying transport speed $v = 0.1 + d \sin \omega t$ and the same initial conditions as those in Fig. 2; $t_f = 30$. Solid line, $d = 0.03$ and $\omega = 10$; dashdot line, $d = 0.03$ and $\omega = 150$; dashed line, $d = 0.4$ and $\omega = 0.1$.

The response of the midpoint of the string vanishes, as expected, until the disturbance propagates there at $t = 0.5$. Similar to the case of free response, the frequency of oscillation of the accelerating string decreases during the period of axial acceleration.

4.2 Sinusoidal Acceleration. For the case of sinusoidally varying transport speed $v(t) = v_0 + d \sin \omega t$, where v_0 is the mean velocity, and d and ω are the amplitude and frequency of the alternating velocity, respectively, (9) and (10) become

$$\xi = -(1 + v_0)t + \frac{d}{\omega}(\cos \omega t - 1) \quad (52)$$

$$\xi + \frac{1 + v_0}{2}(\eta - \xi) - \frac{d}{\omega}\cos \frac{\omega}{2}(\eta - \xi) + \frac{d}{\omega} = 0. \quad (53)$$

Unlike the case of piecewise constant acceleration discussed in the previous section, closed-form expressions for the inverse functions $w(\xi)$ and $p(\eta)$ do not exist. These inverse functions are determined here through numerical solutions of (52) and (53). Given $\xi \in (-\infty, 0]$, the corresponding solution of (52), $t = w(\xi) \in [0, \infty)$, is unique, because (52) is monotonic. Likewise, for any $\eta \geq 0$, the solution of (53), $\xi = p(\eta) \in (-\infty, 0]$, is also unique. By (40),

$$\xi_{\min} = -(1 + v_0)t_f + \frac{d}{\omega}(\cos \omega t_f - 1), \quad (54)$$

and $\eta_{\max}$ can be determined once w is defined. Given the same initial conditions as those in Section 4.1.1, $z_1(x) = [(1 + v_0)/2]b\pi \cos \pi x$. Consequently by (28) and (29), $g(\eta) = [(1 + v_0)/2]b \sin \pi\eta$ when $\eta \in [0, 1]$, and $g(\eta) = -[(1 - v_0)/2]b \sin \pi\xi$, where $\xi = p(\eta - 1) + 1$, when $\eta \in [1, \eta_1]$. It is interesting to note that the values of g are independent of d and ω in the present case.

The free response of the midpoint of the string for $v_0 = 0.1$, $d = 0.03$, $\omega = 10$, $b = 0.05$, and $t_f = 30$ is shown in Fig. 4 by a solid line. As predicted by (31), the amplitude of the response is bounded by $(1 + v_0)b = 0.055$. When $d = 0$, the response, which is not shown here, reduces to that of the uniformly traveling string as predicted by the Green's function method (Wickert and Mote, 1990; Zhu and Mote, 1995). With $\omega = 150$ and other parameters unchanged, the response shown by a dashdot line is indistinguishable from that of a solid line within the scale of the figure. This occurs because g is independent of ω and the resulting changes in (4) are negligible. When $d = 0.4$ and $\omega = 0.1$, significant changes in (4) result and the response is shown by a dashed line. Increasing d/ω reduces the frequency of oscillation of the response.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Though the displacement of the string is bounded under arbitrary variation of the transport speed, the energy of vibration of the string may become unbounded for certain values of v_0, d, and ω. The dynamic behavior of the velocity and the energy of vibration of the string under a varying transport speed remains to be further investigated.

5 Conclusions

1 A general method of solution is developed for the response of an axially accelerating string under arbitrary initial conditions, external forces, and boundary excitation. It determines the exact response of the string with an arbitrary velocity profile, and is applicable to a stationary string as a special case.

2 The free and forced responses are determined once the values of $g(\eta)$ are known for all $\eta \geq 0$. The monotonicity of L_1 and L_2 is crucial in the determination of $g(\eta)$ for all $\eta \geq 2w(-1)$. All the values of g fall into its range corresponding to the domain $[0, 2w(-1)]$ in the case of free response. This is not the case for the forced response in general.

3 In the absence of external excitation, the displacement of the freely vibrating string remains bounded under arbitrary variation of the transport speed in the subcritical regime. The uniform amplitude bound of the response depends only on the initial conditions. Contrary to the prediction based on the truncated models, the unbounded displacement of the string does not exist in the distributed model. Spatial discretization may yield an inaccurate prediction for the response of the distributed system.

Acknowledgments

This work was supported by an Earmarked Grant of the Research Grants Council and a Post-doctoral Fellowship of the Chinese University of Hong Kong.

References

Hwang, S. J., Perkins, N. C., Ulsoy, A. G., and Meckstroth, R. J., 1994, "Rotational Response and Slip Prediction of Serpentine Belt Drive Systems," ASME *Journal of Vibration and Acoustics,* Vol. 116, pp. 71–78.

Meirovitch, L., 1966, *Analytical Methods in Vibrations,* MacMillan, New York, pp. 300–308.

Miranker, W. L., 1960, "The Wave Motion in a Medium in Motion," *IBM Journal of Research and Development,* Vol. 4, pp. 36–42.

Mote, C. D., Jr, 1975, "Stability of Systems Transporting Accelerating Axially moving Materials," ASME *Journal of Dynamic Systems, Measurement, and Control,* Vol. 97, pp. 96–98.

Pakdemirli, M., and Batan, H., 1993, "Dynamic Stability of a Constantly Accelerating Strip," *Journal of Sound and Vibration,* Vol. 168, No. 2, pp. 371–378.

Pakdemirli, M., Ulsoy, A. G., and Ceranoglu, A., 1994, "Transverse Vibration of an Axially Accelerating String," *Journal of Sound and Vibration,* Vol. 169, No. 2, pp. 179–196.

Protter, M. H., and Morrey, C. B., Jr., 1977, *A First Course in Real Analysis,* Springer-Verlag, New York.

Wickert, J. A., and Mote, C. D., Jr., 1988, "Current Research on the Vibration and Stability of Axially Moving Materials," *The Shock and Vibration Digest,* Vol. 20, No. 5, pp. 3–13.

Wickert, J. A., and Mote, C. D., Jr., 1990, "Classical Vibration Analysis of Axially Moving Continua," ASME JOURNAL OF APPLIED MECHANICS, Vol. 57, pp. 738–744.

Wickert, J. A., 1995, "Transient Vibration of Gyroscopic Systems with Unsteady Superposed Motion," *Proceedings of the 1995 Design Engineering Technical Conferences,* Vol. 3, Part B, DE-Vol. 84-2, ASME, New York, pp. 1435–1442.

Zhu, W. D., and Mote, C. D., Jr., 1995, "Propagation of Boundary Disturbances in an Axially Moving Strip in Contact With Rigid and Flexible Constraints," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 873–879.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Flexibility-Based Finite Element Analysis for Structures With Random Material Properties

Yongjian Ren

I. Elishakoff

Department of Mechanical Engineering,
Florida Atlanta University,
Boca Raton., FL 33431-0991

In this paper, we propose a flexibility-based finite element formulation for beams with stochastic bending stiffness. We formulate the element-level finite element equilibrium equation in terms of the element flexibility, so that the stochasticity appears on the right side of the equilibrium equation as the random flexibility. As a result, the unknown internal forces, which are dependent on the stochastic bending stiffness for statically indeterminate beams, appear explicitly in the global finite element equilibrium equations. The internal forces associated with the mean stiffness are used to approximate the unknown internal forces in the process of computation. With the new formulation, the mean and covariance of the displacement for stochastic beams can be directly calculated in terms of mean and covariance function of the flexibility, which can be evaluated from one-dimensional probability density function and the correlation function of the stiffness by the Monte Carlo simulation. The numerical example is given to illustrate the advantages of the new formulation over the conventional perturbation solution.

1 Introduction

The finite element method for stochastic problems (FEMSP) has gained significant interest among researchers in the past decade. The literature includes three monographs (Nakagiri and Hisada, 1985; Ghanem and Spanos, 1991; Kleiber and Hien, 1993), several reviews, and numerous papers. To the best of our knowledge, the existing finite element method for stochastic problems falls mainly into the category of perturbation and/or series expansion methods. It is basically a combination of (a) the deterministic finite element method, (b) perturbation technique, and (c) probabilistic analysis. The perturbed components of the response are obtained through the perturbed components of the uncertain parameters. The perturbation-based FEMSP is associated with the disadvantages pertinent to the perturbation method. For example, (a) only low-order (usually first-order) perturbation technique is practically implementable, since the recursive algorithm to obtain high-order perturbations is extremely time-consuming; (b) the accuracy of the low-order perturbation method is acceptable only for the problems whose uncertain parameters have only small deviations, or, in the probabilistic terminology, the small coefficients of variation; (c) for dynamic response analysis, the perturbation-based method results in a divergent solution when the time variable increases (Shinozuka and Yamazaki, 1988). In fact, an efficient FEMSP, which can efficiently be applied to dynamic analysis of stochastic structures, is still beyond the scope of the present day state of the art of stochastic mechanics.

The present authors realize that adopting the perturbation technique in the FEMSP is due to the difficulty of the probabilistic coupling of the response (which is stochastic due to the presence of uncertain parameters) and the stochastic input parameters. If in some manner the problem of coupling between the response and the input parameters could be resolved, one could then, at least conceptually, obtain the moments of the response in terms of the uncertain parameters. Based on this line of thought, the present authors have proposed several new versions of the finite element method for stochastic beams. There include the moment-related finite element method for the statically determinate stochastic beams (Ren, Elishakoff, and Shinozuka, 1996); new formulation of the stiffness matrix, that generalizes Fuchs' idea (Fuchs, 1991) and related finite element scheme for the statically determinate and indeterminate stochastic beams (Elishakoff, Ren, and Shinozuka, 1996). These new methods are applicable for *any value of the coefficient of variation* of the stochastic stiffness, in contrast to the conventional analysis, which is applicable to the can of the small variation.

However, the moment-related finite element method is not applicable to the statically indeterminate stochastic beams. The generalized Fuchs' approach is applicable for statically indeterminate structures, but it requires the knowledge of the high-order probability density functions, which are usually not available in realistic situations; it also necessitates a tremendous computational effort if the structures have several redundent degrees-of-freedom, when an extremely fine mesh is adopted. It is difficult to generalize Fuchs' approach, which is applicable for beam structures, to general two or three-dimensional structures.

In this study, we propose a flexibility-based finite element method for stochastic problems. The first step in the new formulation is to formulate the element-level finite element equilibrium equation in terms of the reciprocal of the stiffness (the flexibility) in the right side of the equation. The equilibrium equation thus becomes uncoupled from the unknown displacement. The mean and covariance of the displacement are then approximately obtained in terms of the mean and covariance function of the flexibility.

2 Formulation

Consider a beam with length L and spatially varying stochastic bending stiffness $D(x)$. Following the conventional procedure of the finite element method, we discretize the beam into N elements. The finite element equilibrium equation for arbitrary jth element can be obtained from the principle of virtual work, as follows:

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

Discussion on this 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.

Manuscript received by the ASME Applied Mechanics Division, Jan. 28, 1996; final revision, May 30, 1997. Associate Technical Editor: W. K. Liu.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$D_j\mathbf{k}_j\mathbf{u}_j = \mathbf{P}_j^i + \mathbf{P}_j^0 \quad (j = 1, 2, \ldots, N) \tag{1}$$

where D_j = bending stiffness of jth element, which is a random variable representing the spatially random stiffness field $D(x)$ on the jth element. In calculation, D_j can be simply represented by the value of the stiffness at the center of the element, or can be represented by the local average of the stiffness over the element; $\mathbf{u}_j$ = unknown nodal displacement vector, which is also random due to randomness of the stiffness, $\mathbf{P}_j^i$ = nodal internal force vector, $\mathbf{P}_j^0$ = equivalent nodal force vector caused by external loads, $\mathbf{k}_j$ = modified element stiffness matrix, which equals to the element stiffness matrix divided by the bending stiffness D_j. Suppose that the two nodes of jth element are denoted by x_j and x_{j+1}, respectively. Then

$$\mathbf{u}_j = [w_j, w_j', w_{j+1}, w_{j+1}']^T = [\delta_j^T, \delta_{j+1}^T]^T$$

$$\mathbf{P}_j = [Q_{j1}, M_{j1}, Q_{j2}, M_{j2}]^T = [\mathbf{p}_{j1}^T, \mathbf{p}_{j2}^T]^T \tag{2}$$

$$\mathbf{k}_j = \frac{1}{a^3}\begin{bmatrix} 12 & 6a & -12 & 6a \\ 6a & 4a^2 & -6a & 2a^2 \\ -12 & -6a & 12 & -6a \\ 6a & 2a^2 & -6a & 4a^2 \end{bmatrix} = \begin{bmatrix} \mathbf{k}_{11}^j & \mathbf{k}_{12}^j \\ \mathbf{k}_{21}^j & \mathbf{k}_{22}^j \end{bmatrix} \tag{3}$$

where $a = x_{j+1} - x_j$ = length of the element, $\mathbf{p}_{j1}$ and $\mathbf{p}_{j2}$ are nodal forces on left and right nodes, respectively, of the jth element, $\mathbf{k}_{rs}$'s ($r, s = 1, 2$) are two by two submatrices of the modified element stiffness matrix $\mathbf{k}_j$. Dividing Eq. (1) by the bending stiffness D_j yields

$$\mathbf{k}_j\mathbf{u}_j = f_j(\mathbf{P}_j^i + \mathbf{P}_j^0) \tag{4}$$

where $f_j = 1/D_j$ = bending flexibility of jth element. Note that $\mathbf{k}_j$ is a deterministic matrix and assume that the beam is subjected to deterministic external loads. After taking mathematical expectation on both sides of Eq. (4), we arrive at

$$\mathbf{k}_j\bar{\mathbf{u}}_j = \bar{f}_j\mathbf{P}_j^0 + E[f_j\mathbf{P}_j^i]. \tag{5}$$

Equation (5) can be treated as two sets of equations, based on Eqs. (2), (3), as

$$\mathbf{k}_{11}^j\bar{\delta}_j + \mathbf{k}_{12}^j\bar{\delta}_{j+1} = \bar{f}_j\mathbf{p}_{j1}^0 + E[f_j\mathbf{p}_{j1}^i]$$

$$\mathbf{k}_{21}^j\bar{\delta}_j + \mathbf{k}_{22}^j\bar{\delta}_{j+1} = \bar{f}_j\mathbf{p}_{j2}^0 + E[f_j\mathbf{p}_{j2}^i]$$

$$(j = 1, 2, \ldots, N) \tag{6}$$

Similarly, for the $(j + 1)$th element, we have

$$\mathbf{k}_{11}^{j+1}\bar{\delta}_{j+1} + \mathbf{k}_{12}^{j+1}\bar{\delta}_{j+2} = \bar{f}_{j+1}\mathbf{p}_{j+1,1}^0 + E[f_{j+1}\mathbf{p}_{j+1,1}^i]$$

$$\mathbf{k}_{21}^{j+1}\bar{\delta}_{j+1} + \mathbf{k}_{22}^{j+1}\bar{\delta}_{j+2} = \bar{f}_{j+1}\mathbf{p}_{j+1,2}^0 + E[f_{j+1}\mathbf{p}_{j+1,2}^i]$$

$$(j = 1, 2, \ldots, N - 1). \tag{7}$$

Assembling the second set equation of Eq. (6) and the first set equation of Eq. (7), we obtain

$$\mathbf{k}_{21}^j\bar{\delta}_j + (\mathbf{k}_{22}^j + \mathbf{k}_{11}^{j+1})\bar{\delta}_{j+1} + \mathbf{k}_{12}^{j+1}\bar{\delta}_{j+2}$$

$$= \bar{f}_j\mathbf{p}_{j2}^0 + \bar{f}_{j+1}\mathbf{p}_{j+1,1}^0 + E[f_j\mathbf{p}_{j2}^i + f_{j+1}\mathbf{p}_{j+1,1}^i]. \tag{8}$$

Note that the external equivalent nodal force vectors are independent from the stochastic stiffness and can be directly calculated from the applied loads. On the other hand, the internal forces between two adjacent elements, namely, the internal force acting on node 2 of the jth element and the internal force acting on node 1 of the $(j + 1)$th element, have the same magnitude and opposite signs due to the principle of action and reaction

$$\mathbf{p}_{j2}^i = -\mathbf{p}_{j+1,1}^i = \mathbf{p}_{j+1}^i. \tag{9}$$

Substitution of Eq. (9) into Eq. (8) yields

$$\mathbf{k}_{21}^j\bar{\delta}_j + (\mathbf{k}_{22}^j + \mathbf{k}_{11}^{j+1})\bar{\delta}_{j+1} + \mathbf{k}_{12}^{j+1}\bar{\delta}_{j+2}$$

$$= \bar{f}_j\mathbf{p}_{j2}^0 + \bar{f}_{j+1}\mathbf{p}_{j+1,1}^0 + E[(f_j - f_{j+1})\mathbf{p}_{j+1}^i]$$

$$(j = 1, 2, \ldots, N - 1). \tag{10}$$

For statically determinate beams, the internal forces are evaluated through the applied loads, irrespective of the bending stiffness of the beam. Therefore, $\mathbf{p}_{j+1}^i$ can be taken out of the expectation operator in Eq. (10) if the external forces are deterministic. If the stochastic bending stiffness of the beam is weakly homogeneous, the term with expectation operator in Eq. (10) vanishes, since $E[f_j] = E[f_{j+1}] = \bar{f}$. Equation (10) then reduces to

$$\mathbf{k}_{21}^j\bar{\delta}_j + (\mathbf{k}_{22}^j + \mathbf{k}_{11}^{j+1})\bar{\delta}_{j+1} + \mathbf{k}_{12}^{j+1}\bar{\delta}_{j+2} = \bar{f}_j\mathbf{p}_{j2}^0 + \bar{f}_{j+1}\mathbf{p}_{j+1,1}^0$$

$$(j = 1, 2, \ldots, N - 1). \tag{11}$$

For statically indeterminate beams, the third term of the right side of Eq. (10) is negligible compared to first two terms, if a fine finite element mesh is adopted. Indeed, the third term should approach zero when the elements are extremely small, since $E[(f_j - f_{j+1})]$ approaches zero. Hence, Eq. (11) can be applied for both statically determinate and indeterminate beams, representing approximation for statically indeterminate beams.

By examining Eq. (11), one can find that the internal force $\mathbf{p}_{j+1}(j = 1, 2, \ldots, N - 1)$ on $(j + 1)$th junction nodes no longer appears after assembling jth and $(j + 1)$th elements. However, at the two ends of the beam, there is no assembling of elements. At the left end of the beam we obtain, from Eq. (6),

$$\mathbf{k}_{11}^1\bar{\delta}_1 + \mathbf{k}_{12}^1\bar{\delta}_2 = \bar{f}_1\mathbf{p}_{11}^0 + E[f_1\mathbf{p}_{11}^i], \tag{12}$$

and at the right end of the beam

$$\mathbf{k}_{21}^N\bar{\delta}_N + \mathbf{k}_{22}^N\bar{\delta}_{N+1} = \bar{f}_N\mathbf{p}_{N2}^0 + E[f_N\mathbf{p}_{N2}^i]. \tag{13}$$

The internal forces in Eqs. (12), (13) are reaction forces of the supports and will be determined in the same way as in the conventional finite element analysis. Therefore, by assembling all elements together, the global finite element equilibrium equation will have the following form:

$$\mathbf{K}\bar{\mathbf{U}} = \mathbf{F}, \tag{14}$$

where $\mathbf{K}$ = modified global stiffness matrix, $\mathbf{U}$ = vector of global unknown nodal displacements, $\mathbf{F}$ = global vector of modified equivalent nodal forces consisting of $\bar{f}_1\mathbf{p}_{11}^0$, $\bar{f}_j\mathbf{p}_{j2}^0 + \bar{f}_{j+1}\mathbf{p}_{j+1,1}^0$ $(j = 1, 2, \ldots, N - 1)$ and $\bar{f}_N\mathbf{p}_{N2}^0$ as its two-element subvectors. In case that the stochastic stiffness field is weakly homogenous, $\bar{f}_j = \bar{f}_{j+1} = \bar{f}$. Equation (14) becomes

$$\mathbf{K}\bar{\mathbf{U}} = \bar{f}\mathbf{P} \tag{15}$$

where $\mathbf{P}$ = global vector of equivalent nodal forces. The mean displacement is then

$$\bar{\mathbf{U}} = \mathbf{K}^{-1}\mathbf{F} = \bar{f}\mathbf{K}^{-1}\mathbf{P}. \tag{16}$$

To obtain the covariance matrix of the displacement, we subtract Eq. (5) from Eq. (4) and get

$$\mathbf{k}_j\Delta\mathbf{u}_j = \Delta f_j\mathbf{P}_j^0 + f_j\mathbf{P}_j^i - E[f_j\mathbf{P}_j^i] \tag{17}$$

where $\Delta(\cdot) = (\cdot) - (\bar{\cdot})$ is deviation of $(\cdot)$ from its mean value. In perfect analogy to Eqs. (6), (7), we write

$$\mathbf{k}_{11}^j\Delta\delta_j + \mathbf{k}_{12}^j\Delta\delta_{j+1} = \Delta f_j\mathbf{p}_{j1}^0 + f_j\mathbf{p}_{j1}^i - E[f_j\mathbf{p}_{j1}^i]$$

$$\mathbf{k}_{21}^j\Delta\delta_j + \mathbf{k}_{22}^j\Delta\delta_{j+1} = \Delta f_j\mathbf{p}_{j2}^0 + f_j\mathbf{p}_{j2}^i - E[f_j\mathbf{p}_{j2}^i]$$

$$(j = 1, 2, \ldots, N) \tag{18}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

and

$$\mathbf{k}_{11}^{j+1}\Delta\delta_{j+1} + \mathbf{k}_{12}^{j+1}\Delta\delta_{j+2} = \Delta f_{j+1}\mathbf{p}_{j+1,1}^{0} + f_{j+1}\mathbf{p}_{j+1,1}^{i} - E[f_{j+1}\mathbf{p}_{j+1,1}^{i}]$$

$$\mathbf{k}_{21}^{j+1}\Delta\delta_{j+1} + \mathbf{k}_{22}^{j+1}\Delta\delta_{j+2} = \Delta f_{j+1}\mathbf{p}_{j+1,2}^{0} + f_{j+1}\mathbf{p}_{j+1,2}^{i} - E[f_{j+1}\mathbf{p}_{j+1,2}^{i}] \quad (j = 1, 2, \ldots, N-1). \quad (19)$$

The assembly of the second equation of Eq. (18) and the first equation of Eq. (19) results in

$$\mathbf{k}_{21}^{j}\Delta\delta_{j} + (\mathbf{k}_{22}^{j} + \mathbf{k}_{11}^{j+1})\Delta\delta_{j+1} + \mathbf{k}_{12}^{j+1}\Delta\delta_{j+2} = \Delta F_{j+1} \quad (j = 1, 2, \ldots, N-1) \quad (20)$$

where

$$\Delta\mathbf{F}_{j+1} = \Delta f_j\mathbf{p}_{j2}^{0} + \Delta f_{j+1}\mathbf{p}_{j+1,1}^{0} + f_j\mathbf{p}_{j2}^{i} + f_{j+1}\mathbf{p}_{j+1,1}^{i} - E[f_j\mathbf{p}_{j2}^{i} + f_{j+1}\mathbf{p}_{j+1,1}^{i}]. \quad (21)$$

Due to Eq. (9), Eq. (21) reduces to

$$\Delta\mathbf{F}_{j+1} = \Delta f_j\mathbf{p}_{j2}^{0} + \Delta f_{j+1}\mathbf{p}_{j+1,1}^{0} + (f_j - f_{j+1})\mathbf{p}_{j+1}^{i} - E[(f_j - f_{j+1})\mathbf{p}_{j+1}^{i}]. \quad (22)$$

In analogy with Eqs. (12), (13), at the left end of the beam we get

$$\mathbf{k}_{11}^{1}\Delta\delta_1 + \mathbf{k}_{12}^{1}\Delta\delta_2 = \Delta f_1\mathbf{p}_{11}^{0} + f_1\mathbf{p}_{11}^{i} - E[f_1\mathbf{p}_{11}^{i}]. \quad (23)$$

At the right end of the beam we have

$$\mathbf{k}_{21}^{N}\Delta\delta_N + \mathbf{k}_{22}^{N}\Delta\delta_{N+1} = \Delta f_N\mathbf{p}_{N2}^{0} + f_N\mathbf{p}_{N2}^{i} - E[f_N\mathbf{p}_{N2}^{i}]. \quad (24)$$

The global finite element equilibrium equation for the displacement deviation can then be written, after assembling all the elements, as

$$\mathbf{K}\Delta\mathbf{U} = \Delta\mathbf{F}. \quad (25)$$

The covariance matrix of the displacement is

$$\mathrm{Cov}\,(\mathbf{U}, \mathbf{U}^T) = \mathbf{K}^{-1}\,\mathrm{Cov}\,(\Delta\mathbf{F}, \Delta\mathbf{F}^T)\mathbf{K}^{-T} \quad (26)$$

where the jkth element of the covariance matrix Cov ($\Delta\mathbf{F}$, $\Delta\mathbf{F}^T$) is obtained from Eqs. (22)–(24)

$$\begin{aligned} E[\Delta\mathbf{F}_j, \Delta\mathbf{F}_k^T] &= \mathrm{Cov}\,(f_j, f_k)\mathbf{p}_{j2}^{0}(\mathbf{p}_{k2}^{0})^T \\ &+ \mathrm{Cov}\,(f_j, f_{k+1})\mathbf{p}_{j2}^{0}(\mathbf{p}_{k+1,1}^{0})^T + \mathrm{Cov}\,(f_{j+1}, f_k)\mathbf{p}_{j+1,1}^{0}(\mathbf{p}_{k2}^{0})^T \\ &+ \mathrm{Cov}\,(f_{j+1}, f_{k+1})\mathbf{p}_{j+1,1}^{0}(\mathbf{p}_{k+1,1}^{0})^T \\ &+ \mathbf{p}_{j2}^{0}\,\mathrm{Cov}\,[f_j, (f_k - f_{k+1})(\mathbf{p}_{k+1}^{i})^T] \\ &+ \mathbf{p}_{j+1,1}^{0}\,\mathrm{Cov}\,[f_{j+1}, (f_k - f_{k+1})(\mathbf{p}_{k+1}^{i})^T] \\ &+ \mathrm{Cov}\,[(f_j - f_{j+1})\mathbf{p}_{j+1}^{i}, f_k](\mathbf{p}_{k2}^{0})^T \\ &+ \mathrm{Cov}\,[(f_j - f_{j+1})\mathbf{p}_{j+1}^{i}, f_{k+1}](\mathbf{p}_{k+1,1}^{0})^T \\ &+ \mathrm{Cov}\,[(f_j - f_{j+1})\mathbf{p}_{j+1}^{i}, (f_k - f_{k+1})(\mathbf{p}_{k+1}^{i})^T] \\ &- E[(f_j - f_{j+1})\mathbf{p}_{j+1}^{i}]E[(f_k - f_{k+1})(\mathbf{p}_{k+1}^{i})^T]. \end{aligned} \quad (27)$$

For statically determinate structures, the internal forces are independent from the stochastic bending stiffness or flexibility. In this case, Eq. (27) reduces to

$$\begin{aligned} E[\Delta\mathbf{F}_j, \Delta\mathbf{F}_k^T] &= \mathrm{Cov}\,(f_j, f_k)\mathbf{p}_{j2}^{0}(\mathbf{p}_{k2}^{0})^T \\ &+ \mathrm{Cov}\,(f_j, f_{k+1})\mathbf{p}_{j2}^{0}(\mathbf{p}_{k+1,1}^{0})^T + \mathrm{Cov}\,(f_{j+1}, f_k)\mathbf{p}_{j+1,1}^{0}(\mathbf{p}_{k2}^{0})^T \\ &+ \mathrm{Cov}\,(f_{j+1}, f_{k+1})\mathbf{p}_{j+1,1}^{0}(\mathbf{p}_{k+1,1}^{0})^T \\ &+ \mathbf{p}_{j2}^{0}\,\mathrm{Cov}\,[f_j, (f_k - f_{k+1})](\mathbf{p}_{k+1}^{i})^T \\ &+ \mathbf{p}_{j+1,1}^{0}\,\mathrm{Cov}\,[f_{j+1}, (f_k - f_{k+1})](\mathbf{p}_{k+1}^{i})^T \\ &+ \mathbf{p}_{j+1}^{i}\,\mathrm{Cov}\,[(f_j - f_{j+1}), f_k](\mathbf{p}_{k2}^{0})^T \\ &+ \mathbf{p}_{j+1}^{i}\,\mathrm{Cov}\,[(f_j - f_{j+1}), f_{k+1}](\mathbf{p}_{k+1,1}^{0})^T \\ &+ \mathbf{p}_{j+1}^{i}\,\mathrm{Cov}\,[(f_j - f_{j+1}), (f_k - f_{k+1})](\mathbf{p}_{k+1}^{i})^T. \end{aligned} \quad (28)$$

Moreover, when the size of elements tends to zero, the terms with internal forces become progressively smaller, since the terms like Cov $[f_j, (f_k - f_{k+1})]$ tend to zero. Therefore, as the result of approximation, we can simply neglect the terms with internal forces in Eq. (28). In these circumstances

$$\begin{aligned} E[\Delta\mathbf{F}_j, \Delta\mathbf{F}_k^T] &= \mathrm{Cov}\,(f_j, f_k)\mathbf{p}_{j2}^{0}(\mathbf{p}_{k2}^{0})^T \\ &+ \mathrm{Cov}\,(f_j, f_{k+1})\mathbf{p}_{j2}^{0}(\mathbf{p}_{k+1,1}^{0})^T + \mathrm{Cov}\,(f_{j+1}, f_k)\mathbf{p}_{j+1,1}^{0}(\mathbf{p}_{k2}^{0})^T \\ &+ \mathrm{Cov}\,(f_{j+1}, f_{k+1})\mathbf{p}_{j+1,1}^{0}(\mathbf{p}_{k+1,1}^{0})^T. \end{aligned} \quad (29)$$

The above equation can also be applied approximately for statically indeterminate structures, based on the reasoning pertinent to the mean displacement.

3 Evaluation of Mean and Covariance of Flexibility

In section 2 we have derived the mean and the covariance matrix of the displacement vector in terms of those of the element-level flexibilities. The two-dimensional probability density function of the stiffness coefficient is required in order to calculate the covariance function of the flexibility function. However, it is a common case in engineering that the statistical information available on a stochastic parameter consists of a one-dimensional probability density function of the parameter and its correlation function. The required two-dimensional probability density function is seldom available. Under these circumstences, we can adopt the Monte Carlo simulation method to evaluate the mean and covariance function of the flexibility, based on the available one-dimensional probability density function and the correlation function of the stiffness.

Assume that $D(x) = [D(x_1), D(x_2), \ldots, D(x_2), \ldots, D(x_f), \ldots, D(x_N)]^T$ is a vector consisting of the element stiffness over finite elements $j = 1, 2, \ldots, N$. Following the simulation technique for non-Gaussian random fields (Yamazaki and Shinozuka, 1988; Ren, Elishakoff, and Shinozuka, 1995), we can generate a set of samples, $\mathbf{D}_i(\mathbf{x}) = [D_i(x_1), D_i(x_2), D_i(x_3), \ldots, D_i(x_N)]^T$ $(i = 1, 2, \ldots, M)$, of $\mathbf{D}(\mathbf{x})$. Assume that the random stiffness $D(x)$ is a weakly homogeneous non-Gaussian field with zero-mean and autocorrelation function $R_{NN}^{(T)}(\xi) = E[D(x_1)D(x_2)]$, where $\xi = x_1 - x_2$. Let B be a random variable with zero-mean and variance $\sigma_B^2 = R_{NN}^{(T)}(0)$ and given non-Gaussian distribution function $F_B(b)$, which is also the (one-dimensional) distribution function of non-Gaussian random field $D(x)$. Assume that $G(x)$ is a homogeneous Gaussian random field having the same first and second moments as the non-Gaussian field $D(x)$, namely the zero-mean and autocorrelation function $R_{GG}(\xi) = R_{NN}^{(T)}(\xi)$. Denote by H a Gaussian random variable with zero-mean and variance $\sigma_H^2 = R_{GG}(0)$ and Gaussian distribution function $F_H(h)$. We first generate a set of samples of the proposed Gaussian field $G(x)$, namely, $\mathbf{G}_i(\mathbf{x}) = [G_i(x_1), G_i(x_2), G_i(x_3), \ldots, G_i(x_N)]^T$ $(i = 1, 2, \ldots, M)$. The corresponding set of non-Gaussian fields $D(x)$, $\mathbf{D}_i(\mathbf{x}) = [D_i(x_1), D_i(x_2), D_i(x_3), \ldots, D_i(x_N)]^T$ $(i = 1, 2, \ldots, M)$ are obtained in terms of the following mapping:

$$\mathbf{D}_i(\mathbf{x}_r) = \mathbf{F}_B^{-1}\{\mathbf{F}_H[\mathbf{G}_i(\mathbf{x}_r)]\} \quad (i = 1, 2, \ldots, M;\ r = 1, 2, \ldots, N). \quad (30)$$

$D_i(x_r)$ $(i = 1, 2, \ldots, M;\ r = 1, 2, \ldots, N)$ consist of a set of approximate samples of the non-Gaussian random field $D(x)$. Because of the nonlinearity of the mapping, N realizations $D_i(x_r)$ of $D(x)$ may not match, in general cases, the targeted autocorrelation function $R_{NN}^{(T)}(\xi)$. When the targeted autocorrelationship is not achieved, an iterative procedure should be applied to improve $D_i(x_r)$ as follows. Suppose that $R_{GG}^{(S)}(\xi)$ is the

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

estimated autocorrelation function of the Gaussian random field $G(x)$ calculated through the above set of samples. Let $R_{NN}^{(S)}(\xi)$ be the estimated autocorrelation function of the non-Gaussian random field $D(x)$ calculated from corresponding simulations by inversely mapping Gaussian simulations into non-Gaussian simulations. The error between the estimated autocorrelation function $R_{NN}^{(S)}(\xi)$ and the targeted autocorrelation function $R_{NN}^{(T)}(\xi)$ generally stems from two sources, namely from (a) the process of simulation and (b) the process of mapping. The simulation error is caused by the simulation of the Gaussian random field itself. The error should be reduced by the simulation technique, for instance, by increasing the size of the sample. The mapping error is caused by nonlinearity of the mapping. The mapping error can be reduced by adjusting the assumed correlation function of the Gaussian random field $G(x)$, as explained below.

The total error of a simulation procedure is the difference between the targeted correlation function $R_{NN}^{(T)}(\xi)$ and the simulated correlation function $R_{NN}^{(S)}(\xi)$ of non-Gaussian fields. It reads

$$e(\xi) = R_{NN}^{(T)}(\xi) - R_{NN}^{(S)}(\xi). \tag{31}$$

The error of simulation of Gaussian fields, denoted by $e^S(\xi)$, can be written as

$$e^S(\xi) = R_{GG}(\xi) - R_{GG}^{(S)}(\xi) \tag{32}$$

where $R_{GG}(\xi)$ is the correlation function of the Gaussian field. In the first iteration, $R_{GG}(\xi)$ is assumed to be identical to the targeted correlation function of the non-Gaussian field, and in later iterations it is obtained through the iteration equation. $R_{GG}^{(S)}(\xi)$ is the simulated correlation function of the Gaussian field. The second part of the error, denoted by $e^M(\xi)$, is then

$$e^M(\xi) = e(\xi) - e^S(\xi) = R_{NN}^{(T)}(\xi) - R_{NN}^{(S)}(\xi) - e^S(\xi) \tag{33}$$

the superscript "M" indicates that $e^M(\xi)$ is the mapping error. It is natural to require that $|e^M(\xi)|$ should be sufficiently small. The following mapping accuracy criterion is adopted herein: If the estimated autocorrelation function satisfies the requirement $|e^M(\xi)| \le \epsilon$ (where ϵ is an allowable error), one stops the iteration. If, however, the criterion is violated, i.e., $|e^M(\xi)| > \epsilon$, one should update the assumed autocorrelation function of Gaussian random field as follows:

$$R_{GG}^1(\xi) = R_{GG}(\xi)\left[1 + k\frac{R_{NN}^{(T)}(\xi) - R_{NN}^{(S)}(\xi)}{R_{NN}^{(T)}(0)}\right] \tag{34}$$

where k is a factor analogous to the relaxation factor in the iteration methods of the solution of linear equations. It affects the convergence of the iteration. In most cases, $k = 1$. In some cases, k should be taken to be less than unity in order to avoid divergence of the iteration. The algorithm then proceeds to Eq. (30) to start the next iteration until the condition $|e^M(\xi)| \le \epsilon$ is satisfied. The iteration Eq. (34) is a natural generalization from the equation utilized by Yamazaki and Shinozuka (1988) in the context of an unconditional simulation

$$S_{GG}^1(\kappa) = \frac{S_{GG}(\kappa)}{S_{NN}^{(S)}(\kappa)} S_{NN}^{(T)}(\kappa) \tag{35}$$

where $S(\kappa)$ is spectral density of random fields.

After generating the set of samples of the random stiffness field $D(x)$, the mean value of the element stiffness $D(x_j)$ is then calculated as follows:

$$E[D(x_j)] = \frac{1}{M-1}\sum_{i=1}^{M} D_i(x_j). \tag{36}$$

The correlation between $D(x_j)$ and $D(x_k)$ is calculated as

$$C[D(x_j), D(x_k)] = \frac{1}{M-1}\sum_{i=1}^{M} D_i(x_j)D_i(x_k). \tag{37}$$

Similarly, the mean and the correlation function of the flexibility (the reciprocal of the stiffness) are computed as follows:

$$E[f(x_j)] = E\left[\frac{1}{D(x_j)}\right] = \frac{1}{M-1}\sum_{i=1}^{M}\frac{1}{D_i(x_j)} \tag{38}$$

$$C[f(x_j), f(x_k)] = C\left[\frac{1}{D(x_j)}, \frac{1}{D(x_k)}\right] = \frac{1}{M-1}\sum_{i=1}^{M}\frac{1}{D_i(x_j)}\frac{1}{D_i(x_k)}. \tag{39}$$

4 Example

Consider a statically indeterminate beam that is clamped at both ends and subjected to the uniform pressure $q = 1$. The length of the beam is assumed to be unity. The bending stiffness of the beam is assumed to be a spatially stochastic field. Let $D(x) = D_0[1 + k\alpha(x)]$ and assume that the normalized field $\alpha(x)$ possesses a uniform one-dimensional probability density function in an interval $[-a, a]$ and exponential correlation function, namely,

$$\rho(x, y) = \exp\left(-\frac{|x-y|}{d}\right), \quad |x-y| \le L \tag{40}$$

where d is the scale of fluctuation of the random field and is fixed at 0.5 in computation. The two-dimensional probability density function of the field $\alpha(x)$ is assumed to be unavailable.

We use three different approaches to calculate the mean and the variance-covariance of the displacements, namely, (i) the element-level flexibility-based finite element method for stochastic structures, developed in Section 2 where the mean and covariance of the flexibility are evaluated by the Monte Carlo simulation technique; (ii) the conventional deterministic finite element method for simulated samples of the stiffness; and (iii) the first-order perturbation finite element method for stochastic structures. Within the first approach, we generate a set of samples of the random field $\alpha(x)$, based on the simulation technique for non-Gaussian random fields discussed in Section 3. The generated samples are used to calculate the mean and variance-covariance of element-level flexibilities, which will then be utilized in the element-level flexibility-based finite element method for stochastic structures to obtain the mean and variance-covariance of the displacement. There is only a single finite element analysis needed in this formulation.

In the second approach, for each generated sample of the random field $\alpha(x)$, the deterministic finite element method is

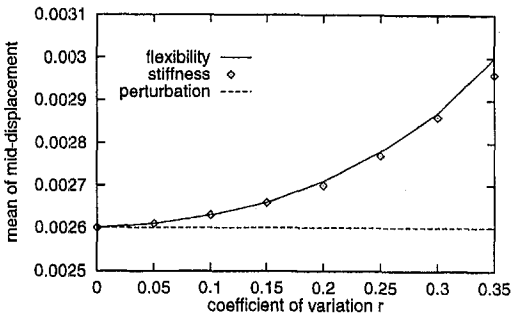


Fig. 1 The mean of the displacement at midsection of the beam versus the coefficient of variation of the stiffness

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

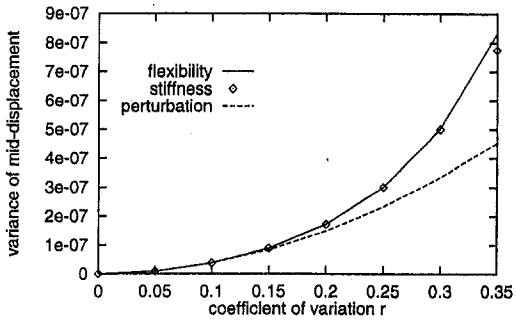


Fig. 2 Variance of the displacement at midsection of the beam versus coefficient of variation of the stiffness

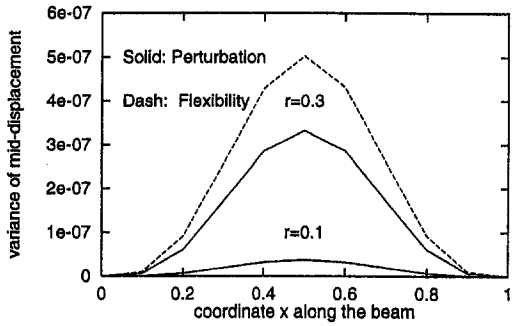


Fig. 4 Variance of the displacement along the beam for coefficient of variation $r = 0.1$ and $r = 0.3$

used to obtain each solution of the displacement. The mean and variance-covariance of the displacement are then obtained by averaging all sets of the solutions, corresponding to the entire set of the samples. The necessary number of finite element analyses conducted equals to the number of the generated samples.

The third approach applies the conventional first-order perturbation-based finite element method for stochastic structures to get the first approximation of the mean value and the variance-covariance matrix of the displacement.

Figures 1 and 2 portray the mean and variance of the displacement at the midsection of the beam, respectively, obtained by the three approaches. One can see that the solution obtained by the first approach and the solution obtained by the second approach agree well, for any value of the coefficient of variation. On the other hand, the perturbation-based solution is acceptable only for small values of the coefficient of variation of the stochastic material property, as expected. Indeed, the perturbation solution turns out to be close to the solution of the flexibility-based method only for small values of the coefficient of variation of the stiffness, $r < = 0.1$. For larger values of the coefficient of variation of the stiffness, the difference between the perturbation solution and the flexibility-based solution increases. For example, the difference constitutes about 18 percent if the coefficient of variation equals 0.3. It can also be concluded from Figs. 1 and 2 that the element-level flexibility based finite element method for stochastic structures is proved to be very efficient and accurate, even though the inner forces corresponding to the mean Young's modulus are utilized instead the real inner forces. The incorporation of the simulation technique can reduce the requirement on data information of the stochastic parameters by the nonperturbative finite element method for stochastic structures, which usually necessitate the knowledge of the two-dimensional probability density function of the stochastic parameters. Also, compared to the conventional simulation-based deterministic finite element method, the approach of incorporating the simulation technique and the nonperturbative finite element method for stochastic structures requires much less computational time than the direct simulation finite element method. For example, in this calculation where 6000 samples and ten elements were adopted and the computing was conducted on a VAX 6000-320 machine, the first approach required approximately one tenth of the time consumed in the direct simulation in conjunction with the deterministic finite element method approach. The structure and the random bending stiffness are discretized with the identical ten-element mesh. The stiffness is modeled as a random variable in each element.

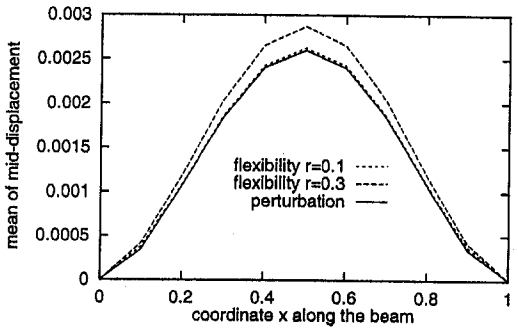


Fig. 3 Mean of the displacement along the beam for coefficient of variation $r = 0.1$ and $r = .3$

Figure 3 depicts the mean displacement of the clamped beam for the values of the coefficient variation of the stochastic stiffness $r = 0.1$, and $r = 0.3$, respectively. The first-order perturbation solution, which is the solution corresponding to the mean stiffness, is independent of the coefficient of variation. One finds that for the case $r = 0.1$, the perturbation solution is close to the flexibility-based solution; however, in the case that $r = 0.3$, the difference between two solutions for mean displacement constitutes about 11 percent. Figure 4 contrasts the results, obtained by the first-order perturbation method and the present flexibility-based method, for the coefficient of variation of the stiffness equals 0.1 and 0.3, respectively. In the case of the small value ($r = 0.1$) of the coefficient of variation of the stiffness, three solutions are quite close to each other. For the case $r = 0.3$, however, the first-order perturbation solution underestimates the displacements' probabilistic characteristics which can be accurately obtained by the flexibility-based method. Such a result can certainly be expected. However, this work appears to be the first one which directly evaluates the errors associated with the perturbation method for the statically indeterminate problems for which the exact solution is unavailable. Moreover, the present work allows one to obtain accurate solutions for arbitrary values of the coefficient of variation.

4 Conclusion

This paper proposes a flexibility-based finite element method for stochastic beams, where the bending stiffness coefficient constitutes a random field. The main idea is to formulate the element-level equilibrium equation in terms of the flexibility coefficients; the displacement is then expressed in terms of the bending flexibility coefficient of the beam. Since the unknown internal forces are involved, an approximate scheme has been proposed to deal with the internal forces. The results of the suggested approximation coincide with the exact solution when the beam is statically determinate. The extention of the proposed approach to two-dimensional structures is under investigation and will appear elsewhere.

Acknowledgment

Support provided by the National Science Foundation through the SGEER grant is appreciated (Program Director: Dr. K. P. Chong). Any opinions, findings, and conclusions or

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

recommendations expressed by this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Elishakoff, I., and Ren, Y. J., 1998, "The bird's eye view on finite element method for stochastic structures," *Computer Methods in Applied Mechanics and Engineering,* to appear.

Elishakoff, I., Ren, Y. J., and Shinozuka, M., 1996a, "Variational principles developed for and applied to analysis of stochastic beams," *Journal of Engineering Mechanics,* pp. 559–565.

Elishakoff, I., Ren, Y. J., and Shinozuka, M., 1996b, "Some thoughts and attendant new results in finite element method for stochastic structures," *Chaos, Solitons, & Fractals,* Vol. 7, pp. 597–609.

Elishakoff, I., Ren, Y. J., Shinozuka, M., 1997, "New formulation of FEM for deterministic and stochastic beams through a generalization of Fuch's approach," *Computer Methods in Applied Mechanics and Engineering,* Vol. 144, 235–243.

Fuchs, M. B., 1991, "Unimodal Beam Elements," *International Journal of Solids and Structures,* Vol. 27, pp. 533–545.

Ghanem, R. G., and Spanos, P. D., 1991, *Stochastic Finite Elements: A Spectral Approach,* Springer-Verlag, New York.

Kleiber, M., and Hien, T. D., 1993, *Stochastic Finite Element Method,* John Wiley and Sons, New York.

Nakagiri, S., and Hisada, T., 1985, *An Introduction to Stochastic Finite Element Method: Analysis of Uncertain Structures,* BaiFuKan, Tokyo (in Japanese).

Ren, Y. J., Elishakoff, I., and Shinozuka, M., 1995, "Conditional simulation of non-Gaussian random fields for earthquake monitoring systems," *Interdisciplinary Journal of Nonlinear Science: Chaos, Solitons & Fractals,* Vol. 5, pp. 91–101.

Ren, Y. J., Elishakoff, I., and Shinozuka, M., 1996, "Finite Element Method for Stochastic Beams Based on Variational Principles," ASME JOURNAL OF APPLIED MECHANICS, Vol. 63, pp. 758–766.

Shinozuka, M., and Yamazaki, F., 1988, "Stochastic Finite Element Analysis: An Introduction," *Stochastic Dynamics,* S. T. Ariaratnam et al., eds., Elsevier Applied Science, London, pp. 241–291.

Yanazaki, F., and Shinozuka, M., 1988, "Digital Generation of non-Gaussian stochastic fields," *Journal of Engineering Mechanics,* Vol. 114, pp. 1183–1197.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Application of Homotopy Analysis Method in Nonlinear Oscillations

Shi-Jun Liao[1]

A. T. Chwang
Fellow ASME

Department of Mechanical Engineering,
The University of Hong Kong,
Pokfulam Road,
Hong Kong

In this paper, we apply a new analytical technique for nonlinear problems, namely the Homotopy Analysis Method (Liao 1992a), to give two-period formulas for oscillations of conservative single-degree-of-freedom systems with odd nonlinearity. These two formulas are uniformly valid for any possible amplitudes of oscillation. Four examples are given to illustrate the validity of the two formulas. This paper also demonstrates the general validity and the great potential of the Homotopy Analysis Method.

1 Introduction

Nonlinear problems are more difficult to solve than linear ones. So far, the most widely applied nonlinear analytical technique is the perturbation method (Nayfeh 1981, 1985; Nayfeh and Mook 1979). The perturbation method is valid in principle only for problems containing small (or large) parameters. Its basic idea is to transform, by means of small parameters, a nonlinear problem into an infinite number of linear subproblems, or a complicated linear problem into an infinite number of simpler ones. Therefore, the small parameter plays a very important role in the perturbation method. It determines not only the accuracy of the perturbation approximations but also the validity of the perturbation method itself. Therefore, it is the small parameter that greatly restricts the applications of the perturbation method. However, in both science and engineering, there exist many nonlinear problems which do not contain any small parameters, especially those with strong nonlinearity. Thus, it is necessary to develop and improve some nonlinear analytical techniques which are independent of small parameters.

There exist some alternative analytical approaches, such as the harmonic balance method (Denman et al. 1964, 1965; Delamotte 1993), the Krylov-Bogoliubov-Mitropolsky (KBM) method (Nayfeh 1981, 1985), and the weighted linearization method (Agrwal and Denman 1985), which sometimes may give analytical approximations even for large parameters. For nonlinear oscillators governed by

$$u'' + F(u) = 0, \tag{1}$$

the abovementioned techniques can produce first approximations of the period of nonlinear oscillations, which are valid even for rather large amplitudes. However, it is usually rather difficult to apply them to produce higher order approximations.

Liao (1992a) proposed a new analytical method called the Homotopy Analysis Method (HAM). The HAM is in principle based on homotopy (Wang and Kao 1991) which is an important part of topology (Duld, 1972; Nash and Sen, 1983; Papy, 1970) so that it has a solid mathematical base. The basic ideas of the HAM are simple. First of all, one introduces an imbedding parameter to construct a homotopy and then analyzes it by the Taylor formula. Subsequently, by means of a linear property of homotopy described by Liao (1992a), one can transform a nonlinear problem into an infinite number of linear subproblems, whether the nonlinear problem contains small parameters or not. Therefore, unlike the perturbation method, the HAM is independent of small parameters and can overcome the restrictions of the perturbation methods. Liao (1992b, c, 1995) and Liao and Chwang (1996) successfully applied the HAM to solve some simple nonlinear problems. These successful applications of the HAM demonstrate the validity and the great potential of the HAM, although the HAM is still in development and further improvements.

For example, let us consider a periodic oscillator governed by

$$(1 + 4q^2u^2)\frac{d^2u}{dt^2} + \kappa u + 4q^2\left(\frac{du}{dt}\right)^2 u = 0 \quad (\kappa > 0) \tag{2}$$

with two initial conditions

$$u(0) = A, \quad u'(0) = 0. \tag{3}$$

Liao (1995) applied the HAM to give two approximate formulas of period T,

$$\frac{T}{T_0} = \sqrt{1 + 2q^2A^2} \tag{4}$$

and

$$\frac{T}{T_0} = \sqrt{5 + 9q^2A^2 - \sqrt{16 + 56q^2A^2 + 53q^4A^4}}, \tag{5}$$

where $T_0 = 2\pi/\sqrt{\kappa}$. As pointed out by Liao (1995), the maximum relative errors of the above two formulas with respect to the exact period

$$\frac{T}{T_0} = \frac{2}{\pi}\sqrt{1 + 4q^2A^2}E(k) \tag{6}$$

are, respectively, ten percent and three percent in the *whole* interval $(qA)^2 \in [0, \infty)$, where $k = 2qA/\sqrt{1 + 4q^2A^2}$ and $E(k)$ is the complete elliptic integral of the second kind. Although Eq. (2) is only a special case, the basic ideas of the HAM are

[1] Present address: Department of Naval Architecture and Ocean Engineering, Shanghai Jiao Tong University, Shanghai 200030, China.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Mar. 13, 1997; final revision, Apr. 1, 1998. Associate Technical Editor: S. W. Shaw.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

general so that it can be applied to solve more nonlinear oscillation problems as shown in this paper.

Let us further consider the free oscillation (undamped and unforced) of a conservative nonlinear system governed by

$$u''(t) + F(u) = 0 \tag{7}$$

with initial conditions

$$u(0) = \beta, \quad u'(0) = 0, \tag{8}$$

where $u(t)$ is a dimensionless variable and $F(u)$ is an odd function of u so that there does not exist the so-called drift or steady-streaming. Suppose that $u = 0$ is a stable equilibrium point, i.e., a center of the oscillation. If $F(u)$ is a nonlinear function, both period T and frequency $\omega = 2\pi/T$ of the corresponding oscillation are dependent upon the amplitude of oscillation β. Usually, period T and frequency ω are functions of β.

In this paper, we first apply the HAM to derive two general formulas for period T of a conservative single-degree-of-freedom system with odd nonlinearity. Then, four examples are given to illustrate that these two period formulas are uniformly valid for *any* possible amplitudes of oscillation.

2 Basic Ideas of the HAM and Main Derivation

Following the approach of Liao (1992a), we construct a one-parameter family of equations

$$\frac{\partial^2 U(t;p)}{\partial t^2} + F[U(t;p)] = (1-p)\left\{\frac{d^2 u_0(t)}{dt^2} + \left[\frac{\Omega^2(p)}{\omega_0^2}\right] F[u_0(t)]\right\}$$

$$t \in [0, \infty), \quad p \in [0, 1] \tag{9}$$

with initial conditions

$$U(t;p) = \beta, \quad \frac{\partial U(t;p)}{\partial t} = 0, \quad \text{for} \quad t = 0, \quad p \in [0, 1], \tag{10}$$

where $p \in [0, 1]$ is an imbedding parameter, $u_0(t)$, which satisfies Eq. (10), is an initial approximation of $u(t)$, and ω_0 denotes the frequency of $u_0(t)$. Note that $U(t; p)$ is now a function of both t and p so that its frequency, denoted by $\Omega(p)$ in Eq. (9), is naturally dependent upon the value of p and moreover, $\Omega(0) = \omega_0$.

At $p = 0$, we have obviously $U(t; 0) = u_0(t)$. At $p = 1$, Eqs. (9) and (10) are exactly the same as Eqs. (7) and (8), respectively, so that $U(t; 1) = u(t)$ which implies $\Omega(1) = \omega$, where ω is the frequency of $u(t)$. Note that $u(t)$ and its frequency $\omega = 2\pi/T$ are exactly what we want to know. As the imbedding parameter p varies from zero to one, $U(t; p)$ varies continuously from $u_0(t)$ to $u(t)$ and $\Omega(p)$ varies from ω_0 to ω. This kind of continuous variations is called deformation in topology. The continuous deformations of $U(t; p)$ and $\Omega(p)$ are completely governed by Eqs. (9) and (10). Therefore, we call them the zeroth-order deformation equations, which represent the changing or becoming process of $U(t; p)$ and $\Omega(p)$. Here, let us emphasize that, in modern science, becoming is considered more important than being, as pointed out by Prigogine and Stengers (1984).

Let

$$\Lambda(p) = \frac{1}{\Omega(p)}, \quad p \in [0, 1] \tag{11}$$

$\lambda_0 = 1/\omega_0$ and $\lambda = 1/\omega$. Then, we have

$$\lambda_0 = \frac{1}{\omega_0} = \frac{1}{\Omega(0)} = \Lambda(0) \tag{12}$$

and

$$\lambda = \frac{1}{\omega} = \frac{1}{\Omega(1)} = \Lambda(1). \tag{13}$$

Introducing a transformation $\tau = \Omega(p)t$ in Eq. (9) and then dividing both sides by $\Omega^2(p)$, we have

$$\frac{\partial^2 U(\tau;p)}{\partial \tau^2} + \Lambda^2(p) F[U(\tau;p)] = (1-p)\left\{\frac{d^2 u_0(\tau)}{d\tau^2} + \lambda_0^2 F[u_0(\tau)]\right\}$$

$$\tau \in [0, \infty), \quad p \in [0, 1] \tag{14}$$

with two initial conditions

$$U(\tau;p) = \beta, \quad \frac{\partial U(\tau;p)}{\partial \tau} = 0, \text{ for } \tau = 0, \ p \in [0, 1], \tag{15}$$

which are also called the zeroth-order deformation equations.

Here, we emphasize that neither small nor large parameters are necessary in constructing the zeroth-order deformation Eqs. (9) and (10), or (14) and (15). In fact, whether or not Eq. (7) contains small or large parameters is not important at all for the validity of the HAM, because the only assumption made on Eq. (7) is that $F(u)$ should be an odd function of u. We also emphasize that, the initial approximation $u_0(t)$ of $u(t)$, together with its corresponding frequency ω_0, can be selected with large freedom that provides us with a great potential to construct better or even the best zeroth-order deformation equations. This kind of freedom is rather important. It would produce much better approximations.

Suppose that $U(\tau; p)$ and $\Lambda(p)$ have derivatives with respect to the imbedding variable p evaluated at $p = 0$:

$$u_0^{[k]}(\tau;p) = \left.\frac{\partial^k U(\tau;p)}{\partial p^k}\right|_{p=0} \quad (k \geq 1), \tag{16}$$

$$\lambda_0^{[k]}(p) = \left.\frac{\partial^k \Lambda(p)}{\partial p^k}\right|_{p=0} \quad (k \geq 1), \tag{17}$$

which are called the kth-order deformation derivatives. Then, by Taylor's formula, we have

$$U(\tau;p) = u_0(\tau) + \sum_{k=1}^{\infty} \frac{u_0^{[k]}(\tau)}{k!} p^k, \tag{18}$$

$$\Lambda(p) = \lambda_0 + \sum_{k=1}^{\infty} \frac{\lambda_0^{[k]}}{k!} p^k. \tag{19}$$

Setting $p = 1$ and noting that $U(\tau; 1) = u(\tau)$ and $\Lambda(1) = \lambda$, we obtain

$$u(\tau) = u_0(\tau) + \sum_{k=1}^{\infty} \frac{u_0^{[k]}(\tau)}{k!} \tag{20}$$

and

$$\lambda = \lambda_0 + \sum_{k=1}^{\infty} \frac{\lambda_0^{[k]}}{k!}, \tag{21}$$

provided that the radii of convergence of series (18) and (19) are not less than 1. Note that (20) gives a relation between the initial approximation $u_0(\tau)$ and solution $u(\tau)$; meanwhile, (21) provides a link between the initial approximation $\lambda_0 = 1/\omega_0$ and the reciprocal of the frequency $\lambda = 1/\omega$. Now, the key to the problem becomes how to solve these kth-order deformation derivatives $u_0^{[k]}(\tau)$ and $\lambda_0^{[k]}(k \geq 1)$. For this purpose, we must first of all give equations governing $u_0^{[k]}(\tau)$ and $\lambda_0^{[k]}(k \geq 1)$.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Differentiating Eqs. (14) and (15) with respect to p and then setting $p = 0$, we have

$$\frac{d^2u_0^{[1]}(\tau)}{d\tau^2} + \lambda_0^2\left(\left.\frac{dF}{du}\right|_{u=u_0(\tau)}\right)u_0^{[1]}(\tau) = -\left\{\frac{d^2u_0(\tau)}{d\tau^2} + \sigma F[u_0(\tau)]\right\} \quad (22)$$

with two initial conditions

$$u_0^{[1]}(\tau) = 0, \quad \frac{du_0^{[1]}(\tau)}{d\tau} = 0, \quad \text{for} \quad \tau = 0, \quad (23)$$

where

$$\sigma = \lambda_0^2 + 2\lambda_0\lambda_0^{[1]}. \quad (24)$$

We call Eqs. (22) and (23) the first-order deformation equations. In the same way, we can obtain all of the kth-order deformation equations governing $u_0^{[k]}(\tau)$ and $\lambda_0^{[k]}(k \geq 2)$, which are similar in form to Eqs. (22) and (23) except the inhomogeneous terms. However, in this paper, we would like to show that even the first-order HAM approximation

$$\lambda \approx \lambda_0 + \lambda_0^{[1]}$$

can give satisfactory period formulas uniformly valid for any possible amplitudes of oscillation. Let us emphasize that the first-order deformation Eqs. (22) and (23) are *linear* with respect to the first-order deformation derivative $u_0^{[1]}(\tau)$. In fact, every kth-order deformation equation is *linear* with respect to the corresponding kth-order deformation derivative ($k \geq 1$), as proved by Liao (1992a) in a rather general case. It means that every term $u_0^{[k]}(\tau)(k \geq 1)$ in (20) is governed by a *linear* equation. Therefore, by (20), the original *nonlinear* problem governed by Eqs. (7) and (8) can be transformed to an infinite number of *linear* subproblems about the kth-order deformation derivatives $u_0^{[k]}(\tau)(k \geq 1)$. In the Introduction of this paper, we point out that the perturbation method uses the small parameter assumption to transform a nonlinear problem into an infinite number of linear subproblems. By means of the HAM, we also accomplish this kind of transformation but *without* using the small parameter assumption. Thus, the HAM is in principle different from the perturbation method.

For conservative oscillations with an odd restoring force $F(u)$, there exists a periodic motion around the equilibrium point $u = 0$ with frequency ω and amplitude β. Therefore, a reasonable and the simplest initial approximation of $u(\tau)$ is

$$u_0(\tau) = \beta \cos(\tau). \quad (25)$$

Moreover, $F'(u) = dF/du$ is clearly an even function of u. Thus, $F(\beta \cos \tau)$ and $F'(\beta \cos \tau)$ can be expressed by

$$F(\beta \cos \tau) = \sum_{k=0}^{\infty} A_{2k+1} \cos[(2k+1)\tau] \quad (26)$$

and

$$\left.\frac{dF}{du}\right|_{u=\beta\cos\tau} = \frac{B_0}{2} + \sum_{k=1}^{\infty} B_{2k} \cos(2k\tau), \quad (27)$$

respectively, where the coefficients A_{2k+1} and B_{2k} ($k = 0, 1, 2, 3, \ldots$) are determined by

$$A_{2k+1} = \frac{1}{\pi}\int_0^{2\pi} F(\beta \cos\tau)\cos[(2k+1)\tau]d\tau \quad (k = 0, 1, 2, 3, \ldots), \quad (28)$$

$$B_{2k} = \frac{1}{\pi}\int_0^{2\pi}\left(\left.\frac{dF}{du}\right|_{u=\beta\cos\tau}\right)\cos(2k\tau)d\tau \quad (k = 0, 1, 2, 3, \ldots). \quad (29)$$

Clearly, the solution $u_0^{[1]}(\tau)$ of Eq. (22) can be expressed by

$$u_0^{[1]}(\tau) = \sum_{k=0}^{\infty} C_{2k+1} \cos[(2k+1)\tau]. \quad (30)$$

Substituting (26), (27), and (30) into (22) and then equating the coefficients, we obtain a set of linear algebraic equations for unknown coefficients C_{2k+1} ($k = 0, 1, 2, 3, \ldots$),

$$\mathbf{Ec} = \frac{2}{\lambda_0^2}\left(\beta\begin{bmatrix}1\\0\\0\\\vdots\end{bmatrix} - \sigma\begin{bmatrix}A_1\\A_3\\A_5\\\vdots\end{bmatrix}\right) = \frac{2}{\lambda_0^2}(\beta\mathbf{b} - \sigma\mathbf{a}), \quad (31)$$

where $\mathbf{a} = [A_1\ A_3\ A_5\ \cdots]^T$, $\mathbf{b} = [1\ 0\ 0\ 0\ \cdots]^T$, $\mathbf{c} = [C_1\ C_3\ C_5\ \cdots]^T$ are vectors and $\mathbf{E}$ is a matrix with elements e_{ij} defined by

$$\begin{aligned}
e_{11} &= B_0 + B_2 - 2\lambda_0^{-2},\\
e_{12} &= B_2 + B_4,\\
e_{13} &= B_4 + B_6,\\
e_{21} &= B_2 + B_4,\\
e_{22} &= B_0 + B_6 - 18\lambda_0^{-2},\\
e_{23} &= B_2 + B_8,\\
e_{31} &= B_4 + B_6,\\
e_{32} &= B_2 + B_8,\\
e_{33} &= B_0 + B_{10} - 50\lambda_0^{-2},\\
&\cdots
\end{aligned}$$

and so on. According to Cramer's theory, the solution of Eq. (31) can be expressed by

$$C_{2k-1} = \frac{2}{\lambda_0^2 \det(\mathbf{E})}[\beta \det(\mathbf{S}_k) - \sigma \det(\mathbf{R}_k)] \quad (k = 1, 2, 3, \ldots), \quad (32)$$

where both $\mathbf{S}_k = \{s_{ij}^k\}$ and $\mathbf{R}_k = \{r_{ij}^k\}$ are matrices with elements

$$s_{ij}^k = \begin{cases} e_{ij} & \text{when} \quad j \neq k,\\ b_i & \text{when} \quad j = k, \end{cases} \quad (33)$$

and

$$r_{ij}^k = \begin{cases} e_{ij} & \text{when} \quad j \neq k,\\ a_i & \text{when} \quad j = k. \end{cases} \quad (34)$$

Hence, according to (30), we have

$$u_0^{[1]}(\tau) = \frac{2}{\lambda_0^2 \det(\mathbf{E})}\sum_{k=1}^{\infty}[\beta \det(\mathbf{S}_k) - \sigma \det(\mathbf{R}_k)]\cos[(2k-1)\tau]. \quad (35)$$

From initial conditions (23), we obtain

$$\sigma = \left[\frac{\sum_{k=1}^{\infty}\det(\mathbf{S}_k)}{\sum_{k=1}^{\infty}\det(\mathbf{R}_k)}\right]\beta. \quad (36)$$

According to (24) and (36), σ is a function of both β and λ_0. If only n terms are used in (30), we have

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\sigma(\beta, \lambda_0) \approx \left[\frac{W_n(\beta, \lambda_0)}{H_n(\beta, \lambda_0)}\right]\beta, \tag{37}$$

where

$$W_n(\beta, \lambda_0) = \sum_{k=1}^{n} \det(\mathbf{S}_k), \tag{38}$$

$$H_n(\beta, \lambda_0) = \sum_{k=1}^{n} \det(\mathbf{R}_k). \tag{39}$$

Then, by Eqs. (21) and (24), we obtain the first-order approximation of λ as

$$\lambda_1 = \lambda_0 + \lambda_0^{[1]} = \frac{1}{2}\left[\lambda_0 + \frac{\sigma(\beta, \lambda_0)}{\lambda_0}\right]. \tag{40}$$

Note that in all of the above derivations, we do *not* set a definite value to λ_0. As we have emphasized, the HAM provides us with large freedom to select the value of $\lambda_0 = 1/\omega_0$. Obviously, if we replace λ_0 by λ_1 in (40), we will obtain a new approximation λ_2 which should be even better than λ_1. Therefore, expression (40) provides, in fact, such an iterative process

$$\lambda_{k+1} = \frac{1}{2}\left[\lambda_k + \frac{\sigma(\beta, \lambda_k)}{\lambda_k}\right]. \tag{41}$$

In the limit as k tends to infinity and $\lambda = \lim_{k\to+\infty} \lambda_k$, we have

$$\lambda^2 = \sigma(\beta, \lambda). \tag{42}$$

Substituting (37) into (42) gives

$$\lambda^2 = \left[\frac{W_n(\beta, \lambda)}{H_n(\beta, \lambda)}\right]\beta, \tag{43}$$

which is an algebraic equation about λ^2. If only one term in Eq. (30) is used ($n = 1$), we have

$$W_1(\beta, \lambda) = b_1 = 1, \tag{44}$$

$$H_1(\beta, \lambda) = a_1 = A_1, \tag{45}$$

so that

$$\lambda = \sqrt{\frac{\beta}{A_1}}. \tag{46}$$

Therefore, we obtain the first approximate period formula

$$T_1 = 2\pi\sqrt{\frac{\beta}{A_1}}. \tag{47}$$

If two terms are used in (30), we have

$$W_2(\beta, \lambda) = \begin{vmatrix} 1 & e_{12} \\ 0 & e_{22} \end{vmatrix} + \begin{vmatrix} e_{11} & 1 \\ e_{21} & 0 \end{vmatrix}$$

$$= B_0 - B_2 - B_4 + B_6 - 18/\lambda^2 \tag{48}$$

and

$$H_2(\beta, \lambda) = \begin{vmatrix} A_1 & e_{12} \\ A_3 & e_{22} \end{vmatrix} + \begin{vmatrix} e_{11} & A_1 \\ e_{21} & A_3 \end{vmatrix}$$

$$= A_1(B_0 - B_2 - B_4 + B_6 - 18/\lambda^2) + A_3(B_0 - B_4 - 2/\lambda^2). \tag{49}$$

Substituting (48) and (49) into (43) we obtain

$$\lambda^2 = \frac{\beta(B_0 - B_2 - B_4 + B_6 - 18/\lambda^2)}{[A_1(B_0 - B_2 - B_4 + B_6 - 18/\lambda^2) + A_3(B_0 - B_4 - 2/\lambda^2)]}, \tag{50}$$

which gives the second approximate period formula

$$T_2 = \frac{2\pi}{\sqrt{f(\beta) \pm \sqrt{f^2(\beta) - h(\beta)}}}, \tag{51}$$

where

$$f(\beta) = \frac{\beta(B_0 - B_2 - B_4 + B_6) + 18A_1 + 2A_3}{36\beta}, \tag{52}$$

$$h(\beta) = \frac{A_1(B_0 - B_2 - B_4 + B_6) + A_3(B_0 - B_4)}{18\beta}. \tag{53}$$

The sign before $\sqrt{f^2(\beta) - h(\beta)}$ should be determined by the condition $\lim_{\beta\to 0}(T_1/T_2) = 1$.

3 Some Applications

3.1 Example 1: Duffing Equation. Let us first consider the well-known Duffing equation

$$u'' + u + \epsilon u^3 = 0 \tag{54}$$

with initial conditions $u(0) = \beta$ and $u'(0) = 0$, where ϵ is a known parameter and $\beta > 0$ is the amplitude of oscillation. Note that both ϵ and β are not necessarily small. When $\epsilon < 0$, $u = 0$ is a center and $u = \pm(-\epsilon)^{-1/2}$ are saddle points so that there exist periodic oscillations around $u = 0$ only for $\sqrt{|\epsilon|}\beta < 1$. However, when $\epsilon > 0$, there always exist oscillations around the stable equilibrium point $u = 0$ for *any* possible values of $\sqrt{\epsilon}\beta$ so that the corresponding amplitude of oscillation, β, may be very large ($0 \le \beta < +\infty$). The exact period T of the Duffing Eq. (54) is

$$T = \frac{4}{\sqrt{1 + \epsilon\beta^2}} K(\mu), \tag{55}$$

where $\mu = \sqrt{\frac{1}{2}\epsilon\beta^2/(1 + \epsilon\beta^2)}$ and $K(\mu)$ is the complete elliptic integral of the first kind. Its first-order perturbation approximation is

$$T_{\text{pert}} = \frac{2\pi}{\left(1 + \frac{3}{8}\epsilon\beta^2\right)}. \tag{56}$$

Since the restoring force is $F(u) = u + \epsilon u^3$, we have

$$F(\beta\cos\tau) = \beta(1 + \tfrac{3}{4}\epsilon\beta^2)\cos(\tau) + \tfrac{1}{4}\epsilon\beta^3\cos(3\tau), \tag{57}$$

$$\left.\frac{dF}{du}\right|_{u=\beta\cos(\tau)} = \left(1 + \frac{3}{2}\epsilon\beta^2\right) + \frac{3}{2}\epsilon\beta^2\cos(2\tau). \tag{58}$$

According to (28) and (29), we have

$$A_1 = \beta(1 + \tfrac{3}{4}\epsilon\beta^2), \tag{59}$$

$$A_3 = \tfrac{1}{4}\epsilon\beta^3, \tag{60}$$

$$B_0 = 2 + 3\epsilon\beta^2, \tag{61}$$

$$B_2 = \tfrac{3}{2}\epsilon\beta^2, \tag{62}$$

$$A_{2k+1} = B_{2k} = 0 \quad (k \ge 2). \tag{63}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

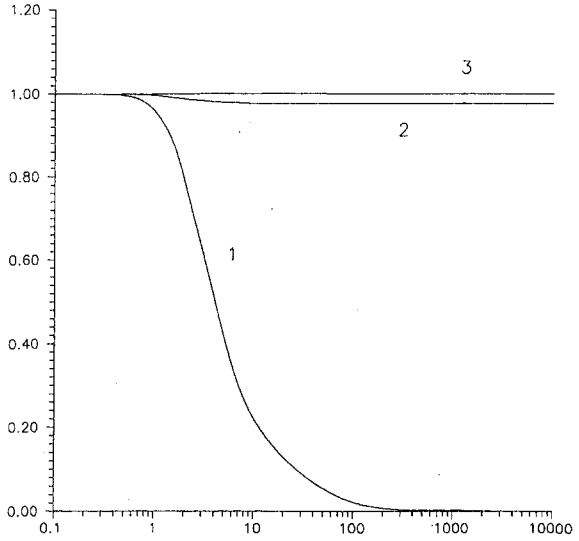


Fig. 1 Comparison of the approximate periods with the exact one (example 1: $\epsilon > 0$); Horizontal axis: $\sqrt{|\epsilon|}\beta$; Vertical axis: ratio of approximate periods to the exact; Curve 1: T_{pert}/T_{exact}; Curve 2: T_1/T_{exact}; Curve 3: T_2/T_{exact}

Substituting (59) into (47) gives

$$T_1 = \frac{2\pi}{\sqrt{1 + \frac{3}{4}\epsilon\beta^2}}. \tag{64}$$

Moreover, according to (51), we have

$$T_2 = \frac{24\pi}{\sqrt{80 + 62\epsilon\beta^2 + \sqrt{4096 + 5888\epsilon\beta^2 + 1684\epsilon^2\beta^4}}}. \tag{65}$$

Here, we select the plus sign before the second square root to ensure $\lim_{\epsilon\beta^2\to 0}(T_1/T_2) = 1$.

Comparisons of the approximate periods (56), (64), and (65) with the exact period (55) are shown in Fig. 1 for $\epsilon > 0$ and Fig. 2 for $\epsilon < 0$, respectively. Clearly, formula (64) and especially (65) are more accurate than (56) and can give good approximations of period T for both small and large values of $\sqrt{|\epsilon|}\beta$. Note that, for $\epsilon > 0$, (64) and (65) give maximum relative errors less than 2.2 percent and 0.2 percent, respectively. In fact,

$$\lim_{\sqrt{|\epsilon|}\beta\to\infty}\frac{T_1}{T} = \left(\frac{\pi}{\sqrt{3}}\right)\frac{1}{K(1/\sqrt{2})} \approx 0.9783 \tag{66}$$

and

$$\lim_{\sqrt{|\epsilon|}\beta\to\infty}\frac{T_2}{T} = \left(\frac{6\pi}{\sqrt{62 + \sqrt{1684}}}\right)\frac{1}{K(1/\sqrt{2})} \approx 1.0016, \tag{67}$$

where K denotes the complete elliptic integral of the first kind. Note that the accuracy of (64) and (65) are *not* strongly dependent upon the values of $\sqrt{|\epsilon|}\beta$ because they are uniformly valid for *any* possible values of $\sqrt{|\epsilon|}\beta$. Therefore, both (64) and (65) are valid for strongly nonlinear oscillations with large amplitudes. Finally, let us point out that the first approximations given by other alternative techniques such as the harmonic balance method may give the same result as (64). However, when we apply the harmonic balance method to obtain a second-order approximation, we have to solve a cubic algebraic equation. Therefore, Eq. (65) is relatively simple and easy to use, because it involves only square roots.

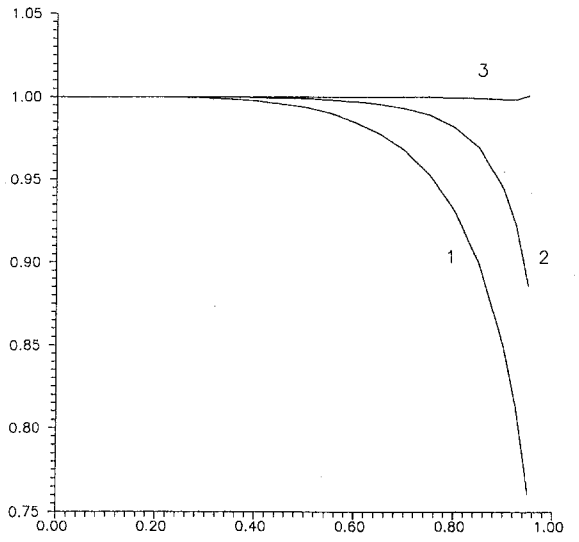


Fig. 2 Comparison of the approximate periods with the exact one (example 1: $\epsilon < 0$); Horizontal axis: $\sqrt{|\epsilon|}\beta$; Vertical axis: ratio of approximate periods to the exact; Curve 1: T_{pert}/T_{exact}; Curve 2: T_1/T_{exact}; Curve 3: T_2/T_{exact}

3.2 Example 2: Simple Pendulum Attached to a Rotating Rigid Frame. As the second example we consider a simple pendulum attached to a rotating rigid frame, governed by

$$\theta'' + \omega_0^2(1 - \kappa\cos\theta)\sin\theta = 0 \tag{68}$$

with two initial conditions $\theta(0) = \beta$ and $\theta'(0) = 0$, where $0 < \kappa < 1$ is a constant and $\beta(0 \le \beta < \pi)$ denotes the amplitude of oscillation. The exact period T of Eq. (68) is

$$T = \frac{4}{\omega_0}\int_0^{\pi/2}\frac{d\phi}{\sqrt{1 - \mu^2\sin^2\phi}\sqrt{1 - \kappa + \kappa\mu^2(1 + \sin^2\phi)}}, \tag{69}$$

where $\mu = \sin(\beta/2)$. Its first-order perturbation approximation is

$$T_{pert} = \left(\frac{2\pi}{\omega_0}\right)\frac{16\sqrt{1 - \kappa}}{[16(1 - \kappa) - (1 - 4\kappa)\beta^2]}. \tag{70}$$

Now, the restoring force is $F(\theta) = \omega_0^2(1 - \kappa\cos\theta)\sin\theta$. Therefore, we have

$$F(\beta\cos\tau) = \omega_0^2\sum_{k=0}^{\infty}(-1)^k\{2J_{2k+1}(\beta) - \kappa J_{2k+1}(2\beta)\}\cos[(2k + 1)\tau], \tag{71}$$

$$F'(\beta\cos\tau) = \omega_0^2[J_0(\beta) - \kappa J_0(2\beta)] + 2\omega_0^2\sum_{k=1}^{\infty}(-1)^k[J_{2k}(\beta) - \kappa J_{2k}(2\beta)]\cos(2k\tau), \tag{72}$$

where $J_k(x)$ is the Bessel function of the first kind. Then, by (26), (27), (47), and (51) we obtain

$$T_1 = \frac{2\pi}{\omega_0}\sqrt{\frac{\beta}{2J_1(\beta) - \kappa J_1(2\beta)}}, \tag{73}$$

$$T_2 = \frac{2\pi}{\omega_0}\frac{1}{\sqrt{Q(\beta) + \sqrt{Q^2(\beta) - P(\beta)}}}, \tag{74}$$

where

$$Q(\beta) = \frac{1}{18\beta}\{\beta[J_0(\beta) + J_2(\beta) - J_4(\beta) - J_6(\beta)] - \kappa\beta[J_0(2\beta) + J_2(2\beta) - J_4(2\beta) - J_6(2\beta)]$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

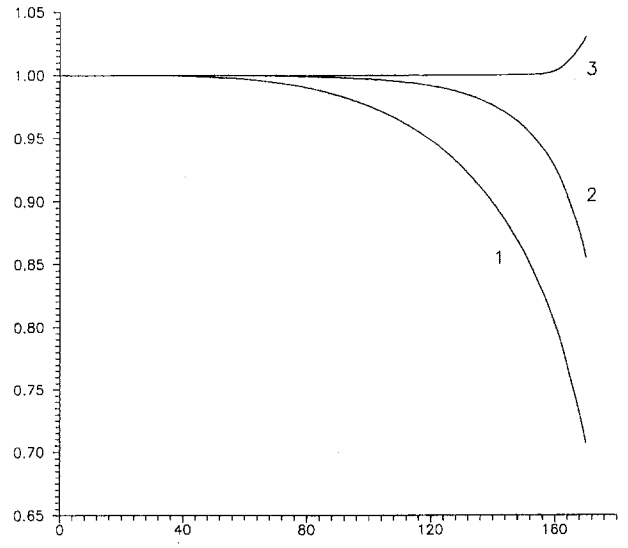


Fig. 3 Comparison of the approximate periods with the exact one (example 2: $\kappa = 0.1$); Horizontal axis: β (degree); Vertical axis: ratio of approximate periods to the exact; Curve 1: T_{pert}/T_{exact}; Curve 2: T_1/T_{exact}; Curve 3: T_2/T_{exact}

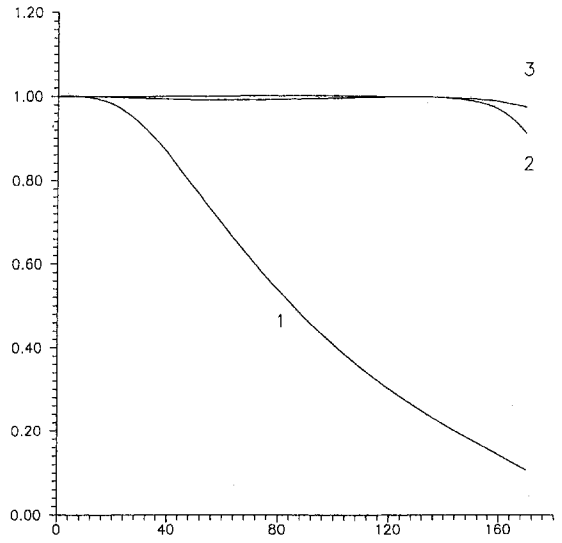


Fig. 5 Comparison of the approximate periods with the exact one (example 2: $\kappa = 0.9$); Horizontal axis: β (degree); Vertical axis: ratio of approximate periods to the exact; Curve 1: T_{pert}/T_{exact}; Curve 2: T_1/T_{exact}; Curve 3: T_2/T_{exact}

$$+ [18J_1(\beta) - 2J_3(\beta)] - \kappa[9J_1(2\beta) - J_3(2\beta)]\}, \quad (75)$$

$$P(\beta) = \frac{1}{9\beta}\{[2J_1(\beta) - \kappa J_1(2\beta)][J_0(\beta) + J_2(\beta) - J_4(\beta) - J_6(\beta)] - \kappa[2J_1(\beta) - \kappa J_1(2\beta)][J_0(2\beta) + J_2(2\beta) - J_4(2\beta) - J_6(2\beta)] - [2J_3(\beta) - \kappa J_3(2\beta)]([J_0(\beta) - J_4(\beta)] - \kappa[J_0(2\beta) - J_4(2\beta)])\}. \quad (76)$$

Comparisons of approximations (70), (73), and (74) with the exact period (69) are shown in Fig. 3 for $\kappa = 0.1$, Fig. 4 for $\kappa = 0.5$, and Fig. 5 for $\kappa = 0.9$, respectively. Note that the perturbation result (70) gives good approximations only when both κ and β are small. However, in contrast to the perturbation approximation, period formula (73) and especially (74) can give good approximations even for large values of β and κ. Therefore, the two period formulas (73) and (74) are valid not only for strong nonlinearity (large κ) but also for large-amplitude oscillations (large β). Note that our results are not satisfactory if β is very close to π. This may be due to the fact that $\beta = \pi$ is an unstable equilibrium point, called the saddle point, where the corresponding motion is a circle.

Note that alternative approaches, such as the harmonic balance method, the KBM method, and the weighted linearization method, can give the same first approximation as (73). However, when applying the harmonic balance method to get the second or even higher approximations, we have to solve a set of transcendental equations. It seems difficult to obtain (74) by alternative approaches.

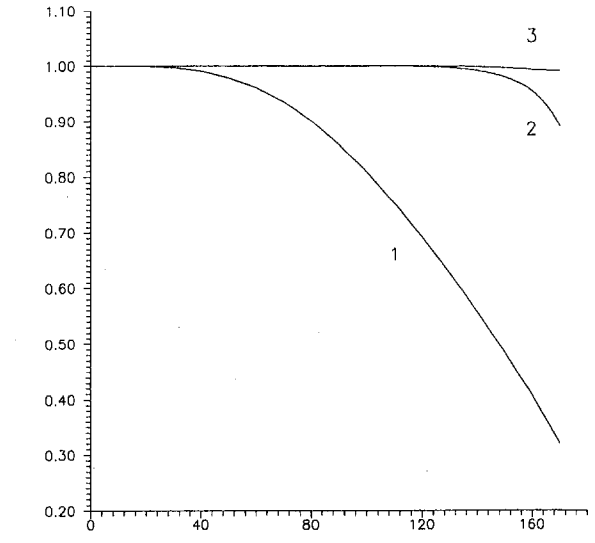


Fig. 4 Comparison of the approximate periods with the exact one (example 2: $\kappa = 0.5$); Horizontal axis: β (degree); Vertical axis: ratio of approximate periods to the exact; Curve 1: T_{pert}/T_{exact}; Curve 2: T_1/T_{exact}; Curve 3: T_2/T_{exact}

3.3 Example 3: Oscillations With Restoring Force $F(u) = u/(1 + u^2)$.

As the third example, let us consider the following oscillation governed by

$$u'' + \frac{u}{1 + u^2} = 0 \quad (77)$$

with two initial conditions $u(0) = \beta$ and $u'(0) = 0$. The exact period T of Eq. (77) is

$$T = 4\int_0^\beta \frac{dt}{\sqrt{\ln(1 + \beta^2) - \ln(1 + t^2)}}. \quad (78)$$

Its first-order perturbation approximation is

$$T_{pert} = \frac{2\pi}{(1 - \frac{3}{8}\beta^2)}. \quad (79)$$

Note that the restoring force is $F(u) = u/(1 + u^2)$. By (28), (29), (47), and (51) we obtain

$$T_1 = \frac{2\pi\beta}{\sqrt{2\left(1 - \frac{1}{\sqrt{1 + \beta^2}}\right)}} \quad (80)$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

and

$$T_2 = \frac{2\pi}{\sqrt{G + \sqrt{G^2 - H}}}, \tag{81}$$

where

$$G = \tfrac{4}{3}\gamma^3 + \tfrac{2}{9}\alpha^2(2 + 3\gamma) - \tfrac{4}{9}\alpha^4(16 - 29\gamma + 7\gamma^3) - \tfrac{16}{9}\alpha^6(5 - 6\gamma + \gamma^3), \tag{82}$$

$$H = -\tfrac{8}{9}\alpha^2\{-4\gamma^3 + \alpha^2(1 - 3\gamma + 8\gamma^3 - 8\gamma^4) + 4\alpha^4(5 - 12\gamma + 6\gamma^2 + 3\gamma^3 - 2\gamma^4) + 16\alpha^6(1 - 2\gamma + \gamma^2)\}, \tag{83}$$

and

$$\alpha = \frac{1}{\beta}, \quad \gamma = \frac{1}{\sqrt{1 + \beta^2}}. \tag{84}$$

Comparisons of the perturbation approximation (79) and the HAM approximations (80) and (81) with the exact period T are shown in Fig. 6. Clearly, the perturbation approximation (79) is valid only for small values of β. It breaks down for large amplitudes of oscillation. However, the HAM approximations (80) and (81) are valid for any possible amplitudes of oscillation.

From (80) and (81) we have

$$\lim_{\beta\to 0} T_1 = \lim_{\beta\to 0} T_2 = 2\pi \tag{85}$$

and

$$\lim_{\beta\to\infty} T_1 \approx \sqrt{2}\pi\beta, \tag{86}$$

$$\lim_{\beta\to\infty} T_2 \approx \left(\frac{3\sqrt{2}}{\sqrt{2 + \sqrt{22}}}\right)\beta. \tag{87}$$

It means that period T of Eq. (77) is equal to 2π for small-amplitude β. However, for large-amplitude oscillations, T is approximately directly proportional to the amplitude of oscillation β. It seems difficult to obtain this qualitative conclusion directly from the exact expression (78).

3.4 Example 4: Oscillations With Restoring Force $F(u) = \alpha u + \lambda u^n$ ($n = 3, 5, 7, \ldots$). As the last example, let us consider a family of nonlinear differential equations

$$u'' + \alpha u + \gamma u^n = 0, \quad (\alpha \geq 0, \gamma > 0, n = 3, 5, 7, \ldots), \tag{88}$$

with two initial conditions

$$u(0) = \beta, \quad u'(0) = 0. \tag{89}$$

The corresponding exact period T is

$$T = 4\int_0^{\pi/2} \frac{d\phi}{\sqrt{\alpha + \left(\frac{2\gamma}{n+1}\right)\beta^{n-1}(1 + \sin^2\phi + \sin^4\phi + \ldots + \sin^{n-1}\phi)}}. \tag{90}$$

From (28), (29), (47), and (51) we obtain two approximations of period T,

$$T_1 = \frac{2\pi}{\sqrt{\alpha + \left(\frac{\gamma}{2}\right)\left(\frac{\beta}{2}\right)^{n-1}\Theta(n+1)}}, \tag{91}$$

$$T_2 = \frac{12\pi\sqrt{\beta}}{\sqrt{\Psi + \sqrt{\Psi^2 - \Pi}}}, \tag{92}$$

where

$$\Psi = 20\alpha\beta + \gamma\left(\frac{\beta}{2}\right)^n[(12 + 24n)\Theta(n+1) + (2 - 14n)\Theta(n+3) + 2n\Theta(n+5)], \tag{93}$$

$$\Pi = 144\alpha^2\beta^2 + 144\gamma\left(\frac{\beta}{2}\right)^n\left\{\alpha\beta[(12n - 2)\Theta(n+1) + (1 - 7n)\Theta(n+3) + n\Theta(n+5)] + n\gamma\left(\frac{\beta}{2}\right)^n[\Theta(n+1)\Theta(n+5) - \Theta^2(n+3)]\right\}, \tag{94}$$

and

$$\Theta(n) = \frac{\Gamma(n+1)}{\Gamma^2\left(\frac{n}{2} + 1\right)} = \frac{n!}{\left[\left(\frac{n}{2}\right)!\right]^2}. \tag{95}$$

Formulas (91) and (92) are valid for any possible amplitudes and give the maximum errors as the dimensionless amplitude $\gamma\beta^{n-1}$ tends to infinity, as shown in Fig. 7, where $3 \leq n \leq 50$. Note that, even for $n = 19$, the maximum error given by (92) is less than ten percent for amplitude $\beta \in [0, \infty)$. All of these

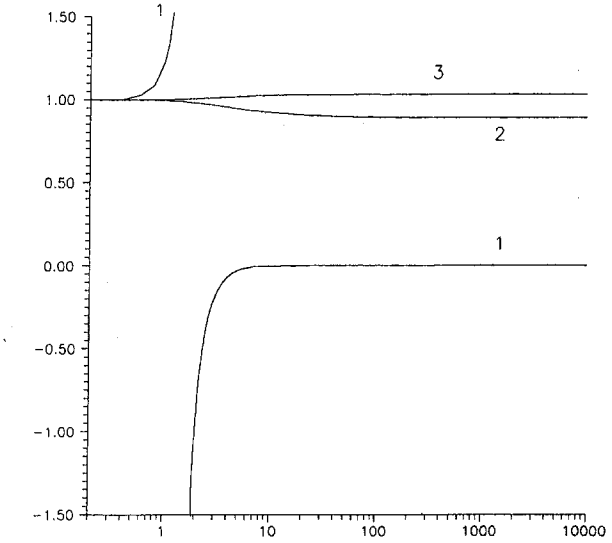


Fig. 6 Comparison of the approximate periods with the exact one (example 3); Horizontal axis: β (degree); Vertical axis: ratio of approximate periods to the exact; Curve 1: T_{pert}/T_{exact}; Curve 2: T_1/T_{exact}; Curve 3: T_2/T_{exact}

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

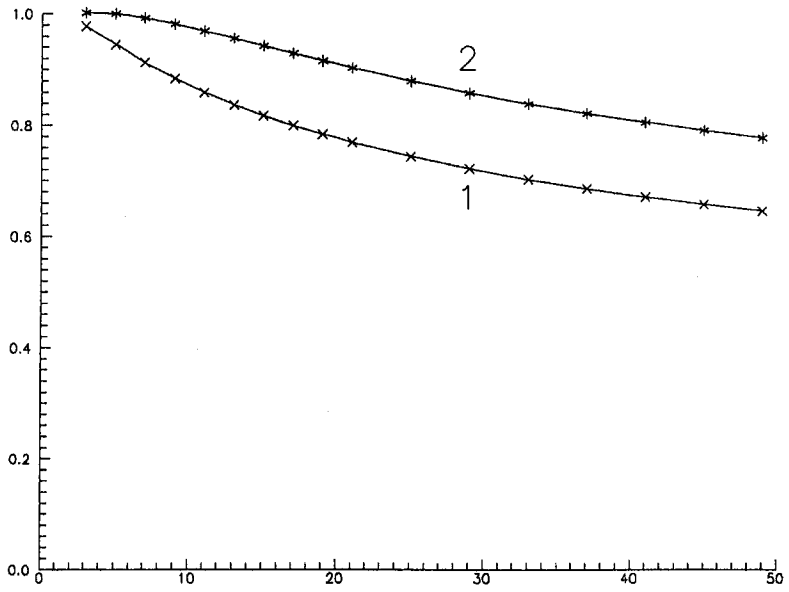


Fig. 7 Maximum errors of example 4 at different values of n as β tends to infinity; Horizontal axis: n; Vertical axis: ratio of approximate periods to the exact; Curve 1: minimum value of T_1/T_{exact}; Curve 2: minimum value of T_2/T_{exact}

illustrate that the validity of the HAM period formulas (47) and (51) for nonlinear oscillations does not strongly depend upon the form of the restoring forces.

In this section we have applied the HAM period formulas (47) and (51) to four typical conservative single-degree-of-freedom oscillation systems. The restoring forces of these four oscillations are rather different. These examples illustrate that, unlike perturbation approximations, the HAM period formulas (47) and (51) are uniformly valid for *any* possible amplitudes of oscillation and can give satisfactory period approximations for oscillations with strong nonlinearity and large amplitudes. These four examples give us reasons to believe that the HAM period formulas (47) and (51) should be valid for many other nonlinear conservative oscillation systems with odd restoring forces.

4 Discussion

In this paper we have applied a new analytical technique, called the Homotopy Analysis Method (HAM), to derive two period formulas (47) and (51) for conservative single-degree-of-freedom oscillation systems with odd nonlinearity. We have further applied these two general formulas to four examples to illustrate that, unlike perturbation approximations, these two formulas are *uniformly* valid for oscillations with *any* possible amplitudes and strong nonlinearity. In other words, the validity of the two period formulas (47) and (51) is independent of small parameters. This is because the HAM is in principle based on homotopy in topology without the assumption of small parameters. Therefore, the validity of the HAM itself does not depend on the presence of small parameters. This point is very important, because the small-parameter assumption brings a lot of restrictions to the perturbation method. Therefore, the HAM can be applied to solve more nonlinear problems than the perturbation method, especially those which do not contain any small parameters.

The basic ideas of the HAM are simple. Some homotopies are first constructed by introducing an imbedding parameter $p \in [0, 1]$ so that unknown functions, together with their periods in case of periodic motions, become also dependent upon this imbedding parameter. These homotopies, or deformations, governed by the zeroth-order deformation equations, can be expressed as Maclaurin series in p. The most important point is that every term in this Maclaurin series, called the kth-order ($k \geq 1$) deformation derivatives at $p = 0$, is governed by a linear equation which can be solved easily. In this way, a nonlinear problem is transformed into an infinite number of linear subproblems. As pointed out by Liao (1992a), this kind of transformation is the key to solve nonlinear problems. Note that the perturbation method relies on the small-parameter assumption to accomplish this kind of transformation. Thus, the validity of the perturbation method itself and its approximate results are naturally dependent upon small parameters. However, unlike the perturbation method, the validity of the HAM and its approximations are independent of small parameters, as shown in this paper. Moreover, the HAM provides us with great freedom to select initial approximations, and through iteration and limiting processes these initial approximations lead to uniformly valid satisfactory results.

Note that the two period formulas (47) and (51) are uniformly valid for the four different examples. We have also applied (47) and (51) to many other nonlinear conservative oscillations with odd nonlinearity and have found that in all the cases they can give satisfactory period approximations in significant parameter ranges. These successful applications of (47) and (51) give us confidence in believing that they should be valid for many other nonlinear conservative oscillations with odd nonlinear restoring forces. As a new attempt to develop a nonlinear analytical technique free from the small-parameter assumption, the HAM seems promising, although it needs further improvements and more applications.

There are other techniques to obtain first approximations for the period of a conservative oscillator, such as the harmonic balance method, the KBM method, and the weighted linearization method. In most cases these alternative techniques give the same first approximations as (47). However, it is difficult to obtain the second or even higher-order approximations by means of these alternatives. For example, the harmonic balance method requires the solution of a set of nonlinear algebraic or even transcendental equations which usually have no exact analytical solutions and must be solved numerically. In comparison with these alternative approaches to obtain the second approximations, the new period formula (51) given by the HAM is both simple and general. It involves only square roots and is valid for any odd restoring forces $F(u)$.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Finally, we would like to point out that the perturbation method is still a useful tool, especially in cases where a proper small parameter is present. What we attempt to develop is a new nonlinear analytical technique in the absence of small parameters. This paper shows one step in this attempt and we still have a long way to go.

Acknowledgments

The authors are grateful to the reviewers for their helpful comments.

References

Agrwal, V. P., and Denman, H., 1985, "Weighted linearization technique for period approximation in large amplitude nonlinear oscillations," *J. Sound Vib.,* Vol. 57, pp. 463–473.

Delamotte, B., 1993, "Nonperturbative method for solving differential equations and finding limit cycles," *Physical Review Letters,* Vol. 70, pp. 3361–3364.

Denman, H., Howard, J., and Liu, Y. K., 1964, "Application of ultraspherical polynomials to nonlinear oscillations (I)," *Quarterly of Appl. Math.,* Vol. 21, p. 325.

Denman, H., Howard, J., and Liu, Y. K., 1965, "Application of ultraspherical polynomials to nonlinear oscillations (II)," *Quarterly of Appl. Math.,* Vol. 22, p. 273.

Duld, A., 1972, *Lectures on algebraic topology,* Springer-Verlag, New York.

Liao, S. J., 1992a, "The Proposed Homotopy Analysis Techniques for the Solutions of Non-linear Problems," Ph.D. dissertation (in English), Shanghai Jiao Tong University, Shanghai, China.

Liao, S. J., 1992b, "A Second-Order Approximate Analytical Solution of a Simple Pendulum by the Process Analysis Method," ASME JOURNAL OF APPLIED MECHANICS, Vol. 59, pp. 970–975.

Liao, S. J., 1992c, "Application of Process Analysis Method in solution of 2D non-linear progressive gravity waves," *Journal of Ship Research,* Vol. 36, No. 1, pp. 30–37.

Liao, S. J., 1995, "An approximate solution technique not depending on small parameters: A special example," *International Journal of Non-linear Mechanics,* Vol. 30, No. 3, pp. 371–380.

Liao, S. J., and Chwang, A. T., 1996, "General Boundary Element Method for Non-linear Problems," *International Journal of Numerical Methods in Fluids,* Vol. 23, pp. 467–483.

Nash, C., and Sen, S., 1983, *Topology and Geometry for Physicists,* Academic Press, London.

Nayfeh, A. H., 1981, *Introduction to Perturbation Techniques,* John Wiley and Sons, New York.

Nayfeh, A. H., 1985, *Problems in Perturbation,* John Wiley and Sons, New York.

Nayfeh, A. H., and Mook, D. T., 1979, *Nonlinear Oscillations,* John Wiley and Sons, New York.

Papy, G., 1970, *Topologie als Grundlage des Analysis-Unterrichts,* Vandenholck & Ruprecht, Goettingen, Germany.

Prigogine, I., and Stengers, I., 1984, *Order out of chaos: man's new dialogue with nature,* Heinemann, London.

Wang, Z. K., and Kao, T. A., 1991, *An introduction to homotopy methods,* Chongqing Publishing House, Chongqing, China.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

D. C. D. Oguamanam
NSERC Postdoctoral Fellow,
Mechanical Engineering Department.

G. R. Heppler
Professor,
Systems Design Engineering Department,
Mem. ASME

University of Waterloo,
Waterloo, ON N2L 3G1, Canada

Geometric Stiffening of Timoshenko Beams

The equations of motion of a prismatic isotropic Timoshenko beam with a tip mass and attached to a rotating hub are derived including the effects of centrifugal forces which appear in the equations of motion as nonlinear functions of the angular speed. The Rayleigh-Ritz method is used to obtain approximate solutions for the cases of a prescribed torque profile and a prescribed hub angular speed profile. In the case of a constant torque input, the effect of the centrifugal forces is not observed until sufficient angular speed is reached. When the system is driven under a constant velocity, it is observed that, for a sufficiently high angular speed, the centrifugal forces reduce the vibration amplitude.

1 Introduction

Structural components such as turbine blades, satellite antennas, helicopter rotors, and some robotic arms are often modeled as rotating flexible beams. These models are distinguished from each other by the methods used in deriving the governing equations of motion, and by the solution technique used thereafter.

Certainly the most frequently used models are based on the Euler-Bernoulli beam theory. For example, Bellezza et al. (1990) use the Euler-Bernoulli beam theory to derive the frequency equations and mode shapes for a flexible rotating beam with a tip mass. They presented their work with application to flexible robots in mind and as a result did not consider the effects due to centrifugal forces on the premise that speeds at which these forces are prominent are rarely attained in practice.

With regard to the frequencies, Hodges (1981) has presented a useful semi-empirical formula for estimating the fundamental frequency of a uniform Euler-Bernoulli beam clamped on a finite hub rotating at a constant angular speed. The formula agrees quite well with the results obtained by Wright et al. (1982) who treated Euler-Bernoulli beams rotating at a constant angular velocity by the method of Frobenius. Wright et al. (1982) obtained exact mode shapes and frequencies for out-of-plane motion of beams that had linearly varying stiffness and mass distributions. For uniform beams their results indicated a decreasing influence of centrifugal stiffening with increasing mode number. Their results also showed, that at all the angular velocities investigated, the first natural frequency decreased with increasing tip mass but, at sufficiently large rotation speeds, the second natural frequency increased with increasing tip mass.

Kojima (1986) has investigated the transient response of a rotating Euler-Bernoulli beam where an approximate solution was obtained by using Galerkin's method and Laplace transforms. The effects of centrifugal forces and hub inertia were ignored and the system was simulated for a prescribed trapezoidally shaped velocity profile.

Within the context of plane sections remain plane, Simo and Vu-Quoc (1987) present the application of a geometrically nonlinear formulation to the dynamic analysis of beams and plates.

Yigit et al. (1988) extended the work of Kojima (1986) by including the effects of centrifugal forces and hub inertia while considering the case of a prescribed torque profile. The results indicate that the linear model, in which it is assumed that the elastic deformation does not affect the rigid-body motion, poorly captures the tip displacement of the beam. Although the tip mass was included during the derivation of the equations of motion, it was excluded in the simulation results that were presented.

Finite element treatments of the problem have been made by Putter and Manor (1978) who studied the effect of centrifugal forces on an Euler-Bernoulli beam and, in an extension of that work, by Hoa (1979) who added the effect of a tip mass. Hoa (1979) (also see Hoa et al., 1980) presented results which showed a decreased frequency of vibration when the beam rotated at low speeds because, at the low speeds, the effect of the tip mass on the inertia of the system dominated the stiffening effect of the centrifugal forces. Conversely, at high speeds it was shown that there would be an increase in the frequency of vibration because the stiffening effect of the centrifugal forces dominated the system as opposed to the inertial effect of the tip mass.

The earliest thorough analytical treatment of the vibration behavior of Timoshenko beams was made by Huang (1961) who used the coupled differential equations presented by Timoshenko (1921) to determine the frequency equations and mode shapes for six common support conditions. He made numerical comparisons between the theories for classical (Euler-Bernoulli) beams, shear beams, Rayleigh beams (Strutt (Lord Rayleigh), 1877), and Timoshenko beams.

Stafford and Giurgiurtiu (1975) developed the equations of motion and frequency equations for a Timoshenko beam that was assumed to be rotating at a constant angular velocity in a fixed plane. Both in-plane and out-of-plane vibrations of the beam where considered in the analysis, which also included the effects of centripetal and Coriolis accelerations. It was shown that as the angular velocity is increased, the inclusion of the transverse shear and rotatory inertia terms becomes increasingly important. Reflecting this finding, Abbas (1986) presented a finite element investigation into the stability of a rotating Timoshenko beam that included the effects of centripetal acceleration and Datta and Ganguli (1990) used a Ritz solution, that employed polynomial trial functions, to treat the problem of a rotating Timoshenko beam that had a damaged section.

Simo (1985) presented a beam formulation that allowed for finite strains and general three-dimensional motion of a beam that included transverse shear effects. Subsequent papers by Simo and Vu-Quoc (1986a, b) explored the case where the beam was restricted to planar motion. The formulation refers all the motion variables to an inertial frame thereby simplifying the inertia terms in the resulting equations of motion. Numerical results obtained from a spatial Galerkin discretization are pre-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Mar. 24, 1997; final revision, Feb. 10, 1998. Associate Technical Editor: N. C. Perkins.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

sented for beams experiencing large rigid-body motions and finite vibration. An alternate mixed variational formulation for the dynamics of moving beams has been presented by Hodges (1990) who includes the possibility that the beams be initially curved or twisted.

A study to assess the merits of using the Euler-Bernoulli, Rayleigh, or Timoshenko beam models to determine the vibration frequencies and mode shapes of a rotating flexible link has been presented by Wang and Guan (1994). The study ignored centrifugal forces and introduced a tip mass only in the Euler-Bernoulli scenario but found that the Timoshenko model is superior for cases where it is necessary to include more than just the first one or two modes or when the beam is short. A position that had earlier been put forward by Huang (1961) and by Timoshenko et al. (1974).

White and Heppler (1995) who have reported the generalized expressions for the frequency equations and mode shapes of a Timoshenko beam which has rigid bodies attached to its ends observed that the contribution of the first moment of the mass of the tip body is negligible when compared to the contribution of the second moment.

We have earlier noted that one of the reasons for the exclusion of the effects of the centrifugal force, as presented by Bellezza et al. (1990), is that this effect may be considered to be insignificant in practice. This is valid for stiff appendages or when the rotational speed is sufficiently slow but the increasing trend towards the use of more flexible linkages calls for a reconsideration of the position, most especially with the current move to drive these appendages at faster speeds.

The centrifugal stiffening effects will appear in the equations of motion as a combination of terms that are nonlinear in the rotation speed and the elastic displacement. In modeling flexible links with various combinations of inertial and geometric nonlinearities, Damaren and Sharf (1995) conclude that the ability to accurately predict the dynamics of fast manoevering manipulators decreases as successively lower-order nonlinear contributions are neglected.

Sadigh and Misra (1995) use Kane's method in obtaining the governing equations of the rotating flexible beam. They highlight the errors introduced by premature linearization and examine three ways by which these errors can be rectified. Their results are in agreement with those of Damaren and Sharf (1995) in that the effectiveness of the model deteriorates with increasing linearization.

The objective of this study is to establish both the governing differential equations for a rotating Timoshenko beam, including the effects of centripetal acceleration, and the set of possible boundary conditions that apply to that equation. We further wish to determine the form of the equations of motion that will arise from a Rayleigh-Ritz treatment of the problem as these will be useful in control systems applications. Finally we wish to make some comparisons between the time responses for a Timoshenko beam and a corresponding Euler-Bernoulli beam. In the sequel we present a treatment of a rotating prismatic isotropic Timoshenko beam that is attached to a rigid hub and which has a rigid body attached at the tip. The centrifugal force components appear as nonlinear functions of the rotation speed. The Rayleigh-Ritz method is used to obtain an approximate solution to the resulting equations of motion and the system is simulated for both a prescribed torque profile and a prescribed velocity profile. The simulation results are compared against results obtained from the corresponding Euler-Bernoulli formulation.

2 Description of the System

A schematic of a flexible beam carrying a mass at the tip and attached to a rotating hub is depicted in Fig. 1. Let the dextral $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$ axes be the basis of the inertial frame $\mathcal{F}_a$, and the $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$ axes be the basis of the frame $\mathcal{F}_b$ which

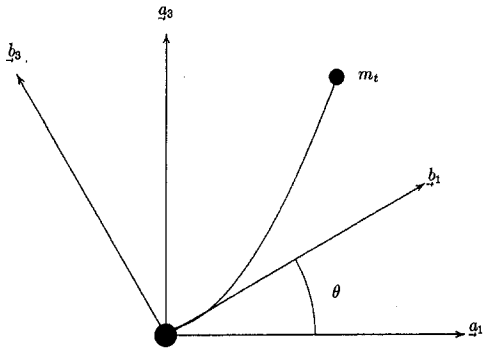


Fig. 1 Schematic of the system

is a body fixed (rotating) frame attached to the beam. The two reference frames are initially parallel. In addition to the usual Timoshenko beam theory assumptions we will assume planar motion (in the $\boldsymbol{a}_1 - \boldsymbol{a}_3$ plane) and small deformations.

3 The Governing Equations

To determine the governing equations we begin by introducing the following assumed displacement field for an infinitesimal beam element of mass dm,

$$\begin{aligned}\boldsymbol{d} &= u(x, y, z, t)\boldsymbol{b}_1 + v(x, y, z, t)\boldsymbol{b}_2 + w(x, y, z, t)\boldsymbol{b}_3 \\ &= (\bar{u}(x, t) - z\psi(x, t))\boldsymbol{b}_1 + \bar{w}(x, t)\boldsymbol{b}_3, \end{aligned} \tag{1}$$

where $u(x, z, t)$ is the axial displacement and $w(x, z, t)$ denotes the transverse displacement. This displacement field embodies the Timoshenko assumptions that all the particles in the same vertical plane are displaced vertically by an equal amount and that, upon deformation, plane sections remain plane but not necessarily normal to the neutral axis. This last assumption is found in the independent bending slope parameter $\psi(x, t)$ which is related to the deflection slope and the transverse shear strain according to

$$\psi(x, t) = \frac{\partial w}{\partial x} - \gamma_{xz}. \tag{2}$$

The position vector $\boldsymbol{r}$ of the infinitesimal beam element dm after deformation may be expressed as

$$\boldsymbol{r} = (x + u)\boldsymbol{b}_1 + (z + w)\boldsymbol{b}_3 \tag{3}$$

and its velocity is given as

$$\begin{aligned}\dot{\boldsymbol{r}} &= \boldsymbol{v} + \boldsymbol{\omega} \times \boldsymbol{r} \\ &= -(z(\dot{\psi} + \dot{\theta}) + \dot{\theta}\bar{w})\boldsymbol{b}_1 + (\dot{\bar{w}} + x\dot{\theta} - z\dot{\theta}\psi)\boldsymbol{b}_3 \end{aligned} \tag{4}$$

where

$$\boldsymbol{v} = \dot{u}(x, t)\boldsymbol{b}_1 + \dot{w}(x, t)\boldsymbol{b}_3 \quad \text{and} \quad \boldsymbol{\omega} = \dot{\theta}\boldsymbol{b}_2.$$

The kinetic energy of the system is composed of three parts: the contribution from the hub motion, the contribution from the motion of the tip mass, and the contribution from the beam. These can be expressed mathematically as

$$\begin{aligned}T = \tfrac{1}{2}\Big(& J_H(\dot{\theta} + \dot{\psi}_0)^2 + m_t v_L^2 - 2\boldsymbol{v}_L \cdot \boldsymbol{c}_{m_t} \times \boldsymbol{\omega} \\ & + \boldsymbol{\omega} \cdot \boldsymbol{J}_{m_t} \cdot \boldsymbol{\omega} + \int_V \rho \dot{\boldsymbol{r}} \cdot \dot{\boldsymbol{r}} dV \Big) \end{aligned} \tag{5}$$

where $\boldsymbol{c}_{m_t}$ is the vector of first moments of the tip mass and $\boldsymbol{J}_{m_t}$ is the inertia dyad for the tip mass. We assume that the x-axis is chosen to be coincident with the centroidal axis of the

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 1 Material properties

Parameter	Parameter values	
	Case I	Case II
L	0.725 m	0.725 m
A	3.175×10^{-4} m^2	3.175×10^{-4} m^2
I	2.67×10^{-10} m^4	13.907×10^{-10} m^4
L/h	228	100
E	6.5×10^{10} Pa	6.5×10^{10} Pa
G	2.8×10^{10} Pa	2.8×10^{10} Pa
κ	0.8509	0.8509
ρ	2.753×10^{-3} kgm^{-3}	2.753×10^{-3} kgm^{-3}
ρAL	0.634 kg	0.634 kg
J_H	1.25×10^{-3} kgm^2	1.25×10^{-3} kgm^2
m_t	0.5 kg	0.317 kg
c_{m_t}	0	0
J_{m_t}	1.25×10^{-3} kgm^2	1.25×10^{-3} kgm^2

beam so that first moments of the beam cross-sectional area are zero.

In addition to the contributions of the elastic strain energy to the total potential energy, there is a contribution from the centrifugal forces. The system total potential energy U may be written as

$$U = \tfrac{1}{2}\int_0^L (EI(\psi')^2 + \kappa GA(\bar{w}' - \psi)^2)dx - \int dW_{cf} \quad (6)$$

where dW_{cf} is the specific work done by the centrifugal forces F_{cf} as per

$$dW_{cf} \triangleq -F_{cf}(ds - dx) \approx -F_{cf}\tfrac{1}{2}(\bar{w}')^2 dx \quad (7)$$

and where dx is the length of an undeformed infinitesimal beam element and ds is the arc length along the neutral axis of the deformed beam (Oguamanam and Heppler, 1995). It may be noted that here we are assuming a hub radius of zero. The total centrifugal force acting on the system has contributions from both the tip mass and the beam and may be expressed as

$$F_{cf} = (m_t L + \tfrac{1}{2}\rho A(L^2 - x^2))\dot{\theta}^2 = \vartheta\dot{\theta}^2. \quad (8)$$

Using (8) in (7), we obtain the appropriate form for the last term in (6).

The virtual work done by the applied torque $\mathcal{M}$ may be expressed as

$$\delta W_{ig} = \mathcal{M}(\delta\theta + \delta\psi_0). \quad (9)$$

The equations of motion are deduced by employing Eqs. (5), (6), and (9) in the extended Hamilton's Principle, and taking the first variation over θ, $\bar{w}$, and ψ. The resulting equations of motion are

$$J_{\text{tot}}\ddot{\theta} + D_{\text{tot}}\dot{\theta} + \mu - \mathcal{M} = 0 \quad (10)$$

$$\rho A(x\ddot{\theta} + \ddot{\bar{w}} - \dot{\theta}^2\bar{w}) - \kappa GA(\bar{w}'' - \psi') - \vartheta\dot{\theta}^2\bar{w}'' + \rho Ax\dot{\theta}^2\bar{w}' + \rho A\dot{\theta}^2\bar{w}' = 0 \quad (11)$$

$$\rho I(\ddot{\psi} + \ddot{\theta} - \dot{\theta}^2\psi) - \kappa GA(\bar{w}' - \psi) - EI\psi'' = 0 \quad (12)$$

where

$$\mu \triangleq J_H\ddot{\psi}_0 + m_t L\ddot{\bar{w}}_L + C_{m_t}(\ddot{\psi}\bar{w}_L - \ddot{\bar{w}}_L\psi_L) + J_{m_t}\ddot{\psi}_L + \int_0^L (\rho Ax\ddot{\bar{w}} + \rho I\ddot{\psi})dx. \quad (13)$$

The general boundary conditions at $x = 0$ are

$$\delta\bar{w}_0 = 0 \quad \text{or}$$

$$\kappa GA(\bar{w}_0' - \psi_0) + (m_t L + \tfrac{1}{2}\rho AL^2)\dot{\theta}^2\bar{w}_0' = 0 \quad (14)$$

$$\delta\psi_0 = 0 \quad \text{or} \quad J_H(\ddot{\theta} + \ddot{\psi}_0) - EI\psi_0' - \mathcal{M} = 0 \quad (15)$$

and at $x = L$ are

$$\delta\bar{w}_L = 0 \quad \text{or} \quad \kappa GA(\bar{w}_L' - \psi_L) - C_{m_t}(2\dot{\psi}_L\dot{\theta} + \dot{\theta}^2 + \psi_L\ddot{\theta}) + m_t(\ddot{\bar{w}}_L - \bar{w}_L\dot{\theta}^2 + L\ddot{\theta} + L\dot{\theta}^2\bar{w}_L') = 0 \quad (16)$$

$$\delta\psi_L = 0 \quad \text{or} \quad EI\psi_L' + C_{m_t}(L\dot{\theta}^2 + 2\dot{\bar{w}}_L\dot{\theta} + \ddot{\theta}\bar{w}_L) + J_{m_t}(\ddot{\psi}_L + \ddot{\theta} - \dot{\theta}^2\psi_L) = 0 \quad (17)$$

where the following definitions apply:

$$J_{\text{tot}} \triangleq J_H + m_t(L^2 + \bar{w}_L^2) + 2C_{m_t}(\bar{w}_L - L\psi_L) + J_{m_t}(1 + \psi_L^2) + \frac{\rho AL^3}{3} + \rho IL + \int_0^L (\rho A\bar{w}^2 + \rho I\psi^2 - \vartheta(\bar{w}')^2)dx \quad (18)$$

$$D_{\text{tot}} \equiv \dot{J}_{\text{tot}}. \quad (19)$$

The equations of motion developed thus far characterize the system when it is driven by a prescribed torque profile. The same derivation method applies when the system is compelled to follow a specified hub angular speed profile and Eqs. (11) and (12) remain applicable but (10) is no longer an equation of motion. Rather, it offers a means of computing the applied torque necessary to achieve the prescribed behavior of $\theta(t)$.

4 Solution Procedure

The equations of motion for the prescribed torque scenario, Eqs. (10), (11), and (12), are coupled integro-partial-differential equations. The uncoupling of these equations is not possible and, even with some assumptions, they are at best difficult to solve. Thus we seek an approximate solution via the Rayleigh-Ritz method. Solution via the Galerkin method is less attractive due to the necessity, in that case, of selecting basis functions which would satisfy the time-varying natural boundary conditions derived above.

The trial functions of the transverse displacement $\bar{w}(x, t)$ and the rotation $\psi(x, t)$ are, respectively, expressed as

$$\bar{w}(x, t) = \mathbf{W}^T(x)\mathbf{p}(t) \quad \text{and} \quad \psi(x, t) = \mathbf{\Psi}^T(x)\mathbf{q}(t) \quad (20)$$

where $\mathbf{W}(x)$ and $\mathbf{\Psi}(x)$ are column matrices of basis functions and $\mathbf{p}(t)$ and $\mathbf{q}(t)$ are vectors of undetermined parameters. A discussion on the selection of the basis functions can be found in Oguamanam and Heppler (1995).

To determine the equations of motion for a prescribed torque profile, we use (20) in the extended Hamilton's Principle and take the variation over $\mathbf{p}$, $\mathbf{q}$, and θ. A matrix representation of the resulting equations of motion may be written as

$$\begin{bmatrix} \mu_{11} & m_{12} & m_{13} \\ m_{12}^T & m_{22} & 0 \\ m_{13}^T & 0 & m_{33} \end{bmatrix}\begin{bmatrix} \ddot{\theta} \\ \ddot{\mathbf{p}} \\ \ddot{\mathbf{q}} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & k_{22} & k_{23} \\ 0 & k_{23}^T & k_{33} \end{bmatrix}\begin{bmatrix} \theta \\ \mathbf{p} \\ \mathbf{q} \end{bmatrix} - \dot{\theta}^2\begin{bmatrix} 0 & 0 & 0 \\ 0 & \Xi_{22} & 0 \\ 0 & 0 & \Xi_{33} \end{bmatrix}\begin{bmatrix} \theta \\ \mathbf{p} \\ \mathbf{q} \end{bmatrix} + \ddot{\theta}C_{m_t}\begin{bmatrix} 0 & c_{12} & c_{13} \\ 0 & 0 & \frac{1}{2}c_{23} \\ 0 & -\frac{1}{2}c_{23}^T & 0 \end{bmatrix}\begin{bmatrix} \theta \\ \mathbf{p} \\ \mathbf{q} \end{bmatrix} + \ddot{\theta}C_{m_t}\begin{bmatrix} 0 & c_{12} & c_{13} \\ -\frac{1}{2}c_{12}^T & 0 & c_{23} \\ -\frac{1}{2}c_{13}^T & -c_{23}^T & 0 \end{bmatrix}\begin{bmatrix} \dot{\theta} \\ \dot{\mathbf{p}} \\ \dot{\mathbf{q}} \end{bmatrix}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

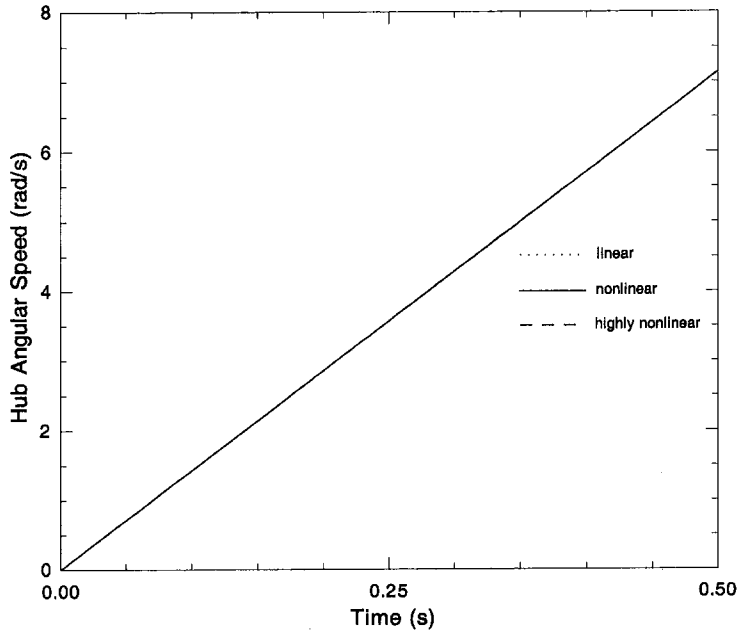


Fig. 2 Timoshenko and Euler-Bernoulli beam, hub angular speed ($\mathcal{M}$ = 4 Nm)

$$+ C_{m_t}\begin{bmatrix}[\ddot{\mathbf{p}}^T \quad \ddot{\mathbf{q}}^T]\begin{bmatrix}0 & \frac{1}{2}c_{23}\\ -\frac{1}{2}c_{23}^T & 0\end{bmatrix}\begin{bmatrix}\mathbf{p}\\ \mathbf{q}\end{bmatrix}\\ \begin{matrix}0\\0\end{matrix}\end{bmatrix}$$

$$+ \ddot{\theta}\begin{bmatrix}[\mathbf{p}^T \quad \mathbf{q}^T]\begin{bmatrix}\Xi_{22} & 0\\ 0 & \Xi_{33}\end{bmatrix}\begin{bmatrix}\mathbf{p}\\ \mathbf{q}\end{bmatrix}\\ \begin{matrix}0\\0\end{matrix}\end{bmatrix}$$

$$+ 2\dot{\theta}\begin{bmatrix}[\dot{\mathbf{p}}^T \quad \dot{\mathbf{q}}^T]\begin{bmatrix}\Xi_{22} & 0\\ 0 & \Xi_{33}\end{bmatrix}\begin{bmatrix}\mathbf{p}\\ \mathbf{q}\end{bmatrix}\\ \begin{matrix}0\\0\end{matrix}\end{bmatrix} = \begin{bmatrix}M\\0\\M\Psi_0\end{bmatrix} \tag{21}$$

where

$$\mu_{11} = \frac{\rho AL^3}{3} + \rho IL + J_H + J_{m_t} + m_t L^2$$

$$m_{12} = \int_0^L \rho Ax\mathbf{W}^T dx + m_t L\mathbf{W}_L^T$$

$$m_{13} = \int_0^L \rho I\Psi^T dx + J_H\Psi_0^T + J_{m_t}\Psi_L^T$$

$$m_{22} = \int_0^L \rho A\mathbf{W}\mathbf{W}^T dx + m_t\mathbf{W}_L\mathbf{W}_L^T$$

$$m_{33} = \int_0^L \rho I\Psi\Psi^T dx + J_H\Psi_0\Psi_0^T + J_{m_t}\Psi_L\Psi_L^T$$

$$\Xi_{22} = \int_0^L \rho A\mathbf{W}\mathbf{W}^T dx + m_t\mathbf{W}_L\mathbf{W}_L^T - \int_0^L \vartheta\mathbf{W}'\mathbf{W}'^T dx$$

$$\Xi_{33} = \int_0^L \rho I\Psi\Psi^T dx + J_{m_t}\Psi_L\Psi_L^T$$

$$c_{12} = 2\mathbf{W}_L^T$$

$$c_{13} = -2L\Psi_L^T$$

$$c_{23} = -2\mathbf{W}_L\Psi_L^T$$

$$k_{22} = \int_0^L \kappa GA\mathbf{W}'\mathbf{W}'^T dx$$

$$k_{23} = -\int_0^L \kappa GA\mathbf{W}'\Psi'^T dx$$

$$k_{33} = \int_0^L \kappa GA\Psi\Psi^T dx + \int_0^L EI\Psi'\Psi'^T dx. \tag{22}$$

The equations of motion for a prescribed velocity profile, which are obtained by taking the variations over $\mathbf{p}$ and $\mathbf{q}$, may be deduced from (21) by deleting the first row and moving the first column to the RHS. In the case of a prescribed hub angular speed the quadratic terms of $\mathbf{p}$ and $\mathbf{q}$ present in (21) are absent because they only appear in the equation that results from taking the variation in θ which is not valid in cases where θ is specified.

5 Numerical Examples

We have chosen to present, in the following examples, results for the nonlinear formulation presented above which we shall refer to as the *highly nonlinear* formulation. In addition we also present results based on the above development but with nonlinear terms involving the product of $\mathbf{p}$ and $\mathbf{q}$ being neglected so that the nonlinearity introduced by the centripetal acceleration is the only nonlinearity contributing to the system response. This formulation shall be referred to as simply *nonlinear*. Finally, the above development with all the nonlinearities neglected (White and Heppler, 1996) is used as the *linear* reference formulation. To provide further insight into the Timoshenko beam model results we have also presented parallel results for the Euler-Bernoulli beam model. The governing equations for the Euler-Bernoulli model may be recovered in the limit as $\kappa G \to \infty$ which results in zero shear strain and $\psi = dw/dx$.

In each example the simulation results have been obtained using at least three elastic modes for each independent variable. The material and geometric properties are outlined in Table 1. The trial functions that are used in this simulation are the pinned-free mode shapes given in White and Heppler (1995) for the Timoshenko model and the pinned-free mode shapes given by Bellezza et al. (1990) for the Euler-Bernoulli beam model. For the purposes of this study, we have ignored the

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

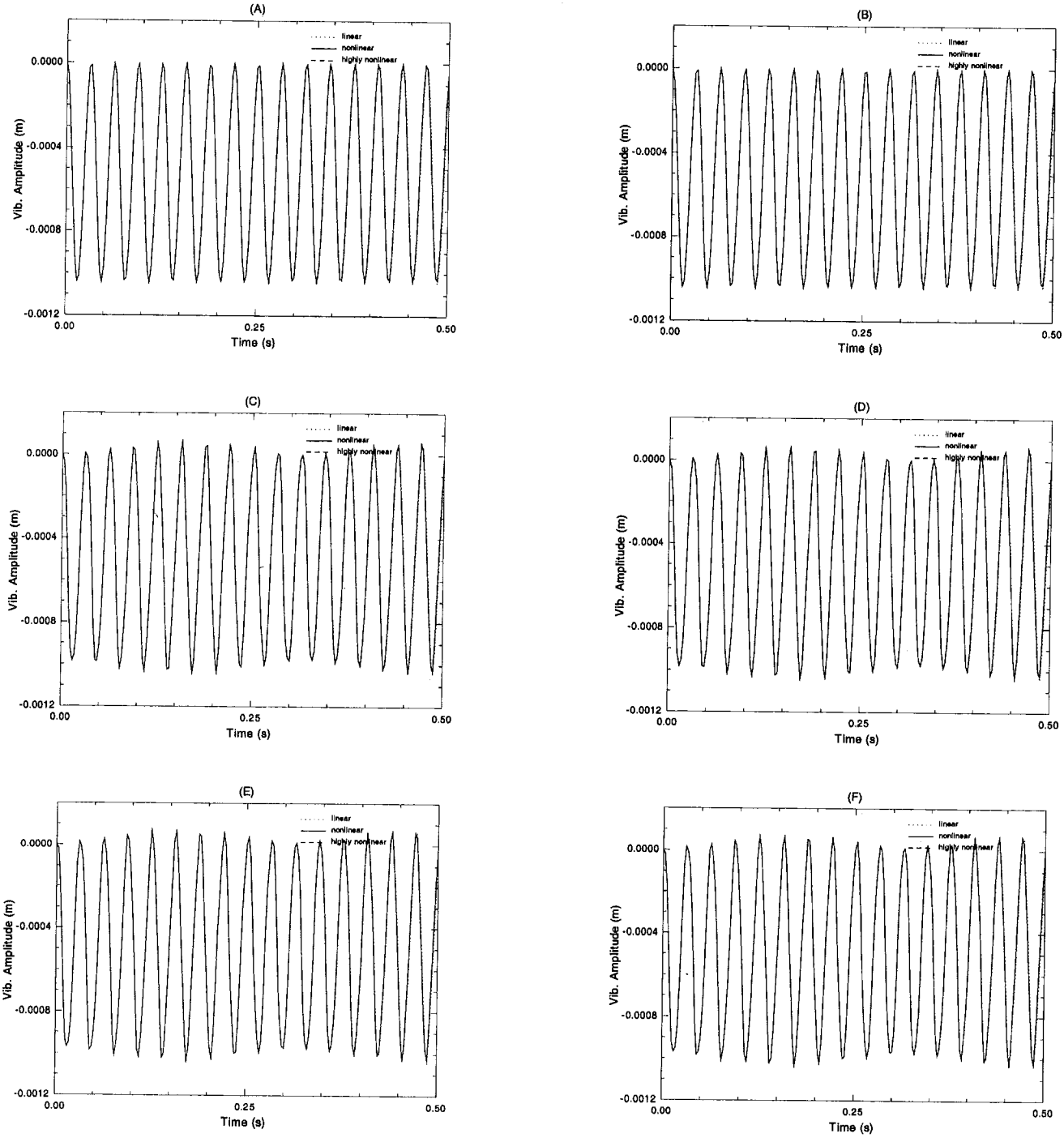


Fig. 3 (*a*) **Timoshenko beam 1 mode,** (*b*) **Euler-Bernoulli beam 1 mode,** (*c*) **Timoshenko beam 2 mode,** (*d*) **Euler-Bernoulli beam 2 mode,** (*e*) **Timoshenko beam 3 mode, and** (*f*) **Euler-Bernoulli beam 3 mode, tip displacement** ($\mathcal{M}$ = 4 Nm) **results**

contributions of the first moment of the tip mass C_{m_t} since its effect is negligible when compared with the effect of the second moment (White and Heppler, 1995). The beam geometries being considered in these examples are relatively slender; they have been selected in this way so that fundamental differences between the Timoshenko and Euler-Bernoulli models may be observed as it is well known that both models predict nearly identical frequencies for slender beams but much different frequencies for short beams.

5.1 Case I: Prescribed Torque Profile. In this set of examples we simulate the system with a constant applied torque of 4 Nm over a time interval of 0.5 s. Numerical simulations using the Rayleigh-Ritz formulation with successively one, two, three, four, and five modes in the case of each of the Timoshenko beam, and Euler-Bernoulli beam models, were performed.

The hub angular speed profiles for the three formulations (i.e., linear, nonlinear, and highly nonlinear) of the Timoshenko beam are depicted in Fig. 2 which has been obtained using the three-mode approximation. The differences between the results that would appear in this figure for the one-mode, two-mode, or three-mode models are insignificant. In terms of the differences in the response due to the degree of nonlinearity in the model it may be observed that the three profiles are indistinguishable on this time interval indicating that the hub has not

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

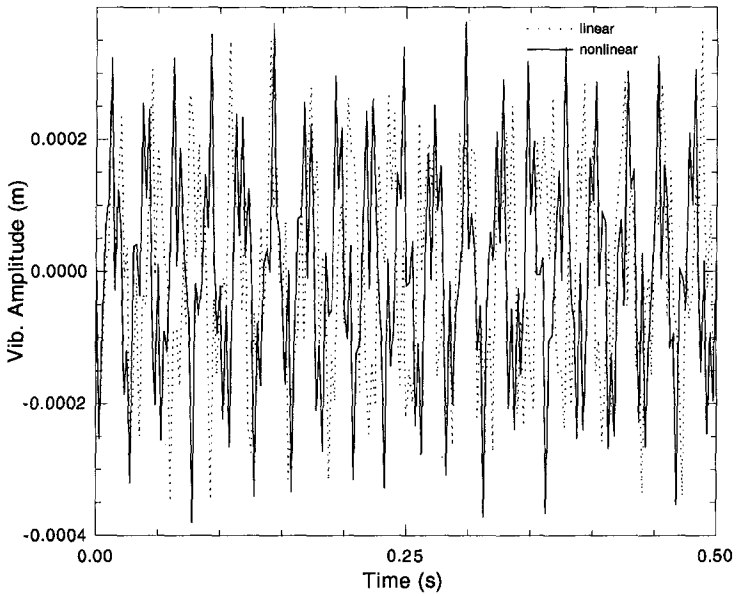


Fig. 4 Timoshenko beam tip displacement ($\dot{\theta}$ = 50 s^{-1}, h/L = 0.01)

yet reached an angular speed that is high enough for the nonlinear effects of $\dot{\theta}$ to be observed. The angular speed is increasing at an constant rate that is consistent with the magnitude of the applied constant torque. It is also evident in Fig. 2 that the elastic motion of the beam does not contribute noticeably to the net velocity of the hub. The above results are indistinguishable from the corresponding results for the Euler-Bernoulli beam model regardless of the number of modes used in the approximation.

The tip-displacement profiles for the linear, nonlinear, and the highly nonlinear scenarios are depicted in Fig. 3 for both the Timoshenko beam model (left column) and the Euler-Bernoulli model (right column). The first row depicts the results for a one-mode approximation, the second row shows the results for a two-mode approximation, and the third row illustrates the results for a three-mode approximation. We have extended this exercise up to five-mode approximations and have found no significant differences between the three-mode approximations and the five-mode approximations.

For both beam models there is a noticeable difference in the responses predicted by the one and two-mode approximations but there is a much less noticeable difference between the two-mode and the three-mode approximations and as mentioned above there are indistinguishable differences between the three, four, and five-mode approximation responses. This suggests that the approximate solution has converged fairly rapidly and that three modes are, for this example, sufficient.

Still with reference to Fig. 3 we can see that in the case of the Timoshenko beam there are no noticeable differences between the two nonlinear formulations and only a slight difference between the nonlinear and the linear formulations after about 0.28 s when the frequency and amplitude from the nonlinear formulations are slightly higher than those of the linear system. This same trend may be observed in the Euler-Bernoulli

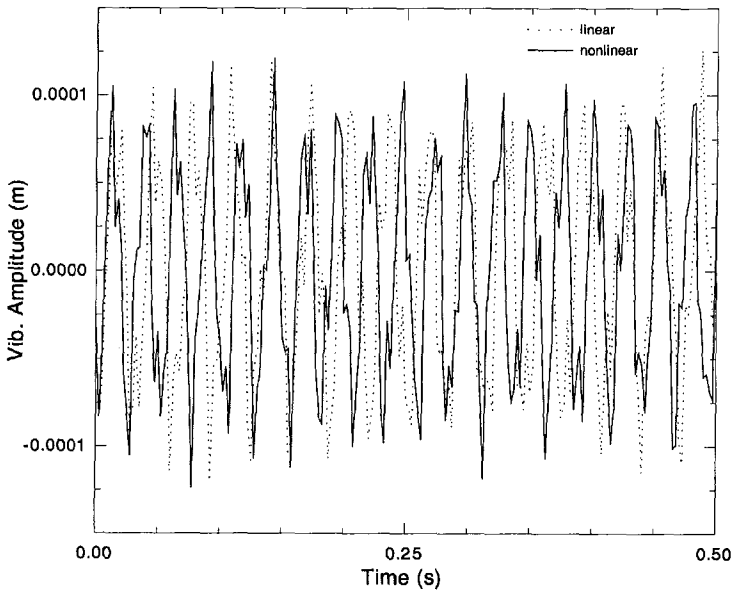


Fig. 5 Euler-Bernoulli beam tip displacement ($\dot{\theta}$ = 50 s^{-1}, h/L = 0.01)

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

beam model results except that the differences between the linear and the nonlinear formulations are a little larger. Between the two models we observe that, in terms of this example, there is little difference between the Euler-Bernoulli model and the Timoshenko model.

Given that the contribution of the centrifugal forces is proportional to the square of the angular speed, the apparent similarity observed in the displacement profiles exhibited by the linear and the nonlinear Timoshenko systems, until a time of ≈ 0.30 s, implies that the speed reached at that time (approximately 3.10 s^{-1} from Fig. 2) was incapable of inducing a noticeable stiffening effect.

The increase in amplitude of the tip displacement in the presence of centrifugal forces is consistent with the findings of Hoa (1979, 1980) who showed that at low speeds, the effect of the tip mass on the inertia of the system dominated over the stiffening effect of the centrifugal forces. In this example we are observing the response when the angular speed is insufficient for the centrifugal forces to be dominant.

5.2 Case II: Prescribed Velocity Profile. The examples in this case examine the system behavior for a prescribed constant hub angular speed of 50 s^{-1}. The material and geometric properties are given in Table 1. The beam used in these examples is thicker than in the previous examples with the ratio of the length of the beam to the height of the beam (L/h) being selected to be 100 and the tip mass is smaller. The results are for three-mode approximations. There are no highly nonlinear formulation results in this case because, for a prescribed hub angular speed, the terms which distinguish that formulation from a nonlinear formulation do not appear in the equations which govern the motion of the beam.

Figure 4 depicts the tip-displacement profiles of the linear and nonlinear Timoshenko beam formulations. It may be observed that the nonlinear system exhibits a deformation with a larger maximum amplitude and a higher frequency than the linear system. The increased frequency is a direct consequence of the effective centrifugal stiffening. The principal effect of going from a one to a two to a three-mode approximation is that the simulations which use more modes display greater peak amplitudes with the greatest change observed between the one-mode approximation results and the two-mode approximation results. The differences in the peak amplitudes between the two-mode and three-mode approximations are noticeable but are much smaller than those that are observed in going from one to two modes.

The overall waveform of both the linear and nonlinear three-mode approximations is complex with significant contributions from the higher modes. The linear system clearly displays inherent beats in the two-mode approximation but this behavior becomes masked when the third mode is added to the approximation.

The differences in the behavior of the Timoshenko models and the Euler-Bernoulli models are easily seen by comparing Fig. 4 to Fig. 5; the former exhibiting close to twice the deflection.

6 Summary

A flexible beam carrying a tip mass and attached to a rotating base has been modeled using the Timoshenko beam theory. The equations of motion were derived using the extended Hamilton's Principle and an approximate solution was achieved via the Rayleigh-Ritz method. The system was simulated with and without the effects of the nonlinearities that arise due to centripetal acceleration being considered and these results were compared to the corresponding results for an Euler-Bernoulli beam model. It was observed that there is a threshold hub angular speed below which centrifugal stiffening effects are negligible not only in terms of the natural frequencies of the system as has previously been established but also in terms of the displacement trajectories. The Timoshenko beam predicts larger deflections at high angular speeds than does the Euler-Bernoulli model.

References

Abbas, B. A. H., 1986, "Dynamic Stability of a Rotating Timoshenko Beam With a Flexible Root," *Journal of Sound and Vibration,* Vol. 108, No. 1, pp. 25–32.

Bellezza, F., Lanari, L., and Ulivi, G., 1990, "Exact Modeling of the Flexible Slewing Link," *Proc. IEEE Int. Conf on Robotics and Automation,* Cincinnati, Vol. 3, pp. 734–739.

Damaren, C., and Sharf, I., 1995, "Simulation of Flexible-Link Manipulators With Inertia and Geometric Nonlinearities," ASME *Journal of Dynamic Systems, Measurement and Control,* Vol. 117, pp. 74–87.

Datta, P. K., and Ganguli, R., 1990, "Vibration Characteristics of a Rotating Blade with Localized Damage Including the Effects of Shear Deformation and Rotary Inertia," *Computers and Structures,* Vol. 36, No. 6, pp. 1129–1133.

Hoa, S. V., 1979, "Vibration of Rotating Beam With Tip Mass," *J. Sound and Vibration,* Vol. 67, pp. 369–381.

Hoa, S. V., Hodges, D. H., and Rutkowski, M. J., 1980, "Comments on 'Vibration of a Rotating Beam with Tip Mass,' " *Journal of Sound and Vibration,* Vol. 72, No. 4, pp. 547–549.

Hodges, D. H., 1981, "An Approximate Formula for the Fundamental Frequency of Uniform Rotating Beams Clamped Off the Axis of Rotation," *Journal of Sound and Vibration,* Vol. 77, No. 1, pp. 11–18.

Hodges, Dewey H., 1990, "A Mixed Variational Formulation Based on Exact Intrinsic Equations for Dynamics of Moving Beams," *International Journal of Solids and Structures,* Vol. 26, No. 11, pp. 1253–1273.

Huang, T. C., 1961, "The Effect of Rotary Inertia and of Shear Deformation on the Frequency and Normal Mode Equations of Uniform Beams With Simple End Conditions," ASME JOURNAL OF APPLIED MECHANICS, Vol. 28, pp. 579–584.

Kojima, H., 1986, "Transient Vibrations of a Beam/Mass System Fixed to a Rotating Body," *J. Sound and Vibration,* Vol. 107, pp. 149–154.

Oguamanam, D. C. D., and Heppler, G. R., 1995, "Centrifugal Stiffening of a Timoshenko Beam," ConStruct Technical Report, TR95-001, University of Waterloo, Waterloo, ON, Canada.

Putter, S., and Manor, H., 1978, "Natural Frequencies of Radial Rotating Beams," *J. Sound and Vibration,* Vol. 56, pp. 175–185.

Sadigh, M. J., and Misra, A. K., 1995, "More on the So-Called Dynamic Stiffening Effect," *J. Astro. Sci.,* Vol. 43, No. 2, pp. 101–125.

Simo, J. C., 1985, "A Finite Strain Beam Formulation. The Three Dimensional Dynamic Problem. Part I," *Computer Methods in Applied Mechanics and Engineering,* Vol. 49, pp. 55–70.

Simo, J. C., and Vu-Quoc, L., 1986a, "On the Dynamics of Flexible Beams Under Large Overall Motions—The Plane Case: Part I," ASME JOURNAL OF APPLIED MECHANICS, Vol. 53, pp. 849–854.

Simo, J. C., and Vu-Quoc, L., 1986b, "On the Dynamics of Flexible Beams Under Large Overall Motions—The Plane Case: Part II," ASME JOURNAL OF APPLIED MECHANICS, Vol. 53, pp. 855–863.

Simo, J. C., and Vu-Quoc, L., 1987, "The Role of Nonlinear Theories in Transient Dynamic Analysis of Flexible Structures," *Journal of Sound and Vibration,* Vol. 119, No. 3, pp. 487–508.

Stafford, R. O., and Giurgiurtiu, V., 1975, "Semi-Analytic Methods for Rotating Timoshenko Beams," *International Journal of Mechanical Science,* Vol. 17, pp. 719–727.

Strutt, J. W., (Lord Rayleigh) 1877, *Theory of Sound,* Macmillan, London.

Timoshenko, S. P., 1921, "On the Correction for Shear or the Differential Equations for Transverse Vibration of Prismatic Bars," *Philosophy Magazine,* Ser. 6, Vol. 41, pp. 744–746.

Timoshenko, S., Young, D. H., and Weaver, W., 1974, *Vibration Problems in Engineering,* 4th Ed., John Wiley and Sons, New York.

Vu-Quoc, L., and Simo, J. C., 1987, "Dynamics of Earth-Orbiting Flexible Satellites With Multibody Components," *Journal of Guidance, Control and Dynamics,* Vol. 10, No. 6, pp. 449–458.

Wang, F. Y., and Guan, G., 1994, "Influences of Rotating Inertia, Shear, and Loading on Vibrations of Flexible Manipulators," *J. Sound and Vibration,* Vol. 171, pp. 433–452.

White, M. W. D., and Heppler, G. R., 1995, "Vibration Modes and Frequencies of Timoshenko Beams With Attached Rigid Bodies," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 193–195.

White, M. W. D., and Heppler, G. R., 1996, "Vibration of Rotating Timoshenko Beams," ASME *Journal of Vibration and Acoustics,* Vol. 118, pp. 606–613.

Wright, A. D., Smith, C. E., Thresher, R. W., and Wang, J. L. C., 1982, "Vibration Modes of Centrifugally Stiffened Beams," ASME JOURNAL OF APPLIED MECHANICS, Vol. 49, pp. 197–202.

Yigit, A., Scott, R. A., and Ulsoy, A. G., 1988, "Flexural Motion of a Radially Rotating Beam Attached to a Rigid Body," *J. Sound and Vibration,* Vol. 121, pp. 201–210.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Transient Bending Wave Propagation in Anisotropic Plates

K.-E. Fällström
Division of Experimental Mechanics

O. Lindblom[1]
Department of Mathematics

Luleå University of Technology,
S-971 87 Luleå, Sweden

In this paper we study transient propagating bending waves. We use the equations of orthotropic plate dynamics, derived by Chow about 25 years ago, where both transverse shear and rotary inertia are included. These equations are extended to include anisotropic plates and an integral representation formula for the bending waves is derived. Chow's model is compared with the classical Kirchoff's model. We also investigate the influence of the rotary inertia. Comparisons with experimental data are made as well.

1 Introduction

In 1877 Lord Rayleigh showed how the addition of rotary inertia contributes to the classical transitional inertia affected the flexural vibrations of isotropic beams. About 40 years later Timoshenko (1921) showed that the effects of shear deformations, previously disregarded, were equally important. It is well known that both effects serve to decrease the computed frequencies because of increased inertia and flexibility of the system.

An extension of the plate theory that account for shear deformation was proposed by Reissner (1945) for the static deflection of plates, and a significant number of papers by others have followed this approach. Mindlin's paper, from 1951, unquestionably made the most profound impact upon the subject. The basic sixth-order system of partial differential equations of motion was derived, relating moments and transverse shears to transverse deflection and bending rotation. The problem of the rectangular plate simply supported on all edges was solved by Mindlin, Schacknow, and Deresiewicz (1956). In 1959 Thorkildsen et al. (1959) treated a stiffened plate as an orthotropic plate for purpose of analysis. The effect of rotary inertia was considered, but shear deformation was ignored.

In 1971 Chow derived differential equations for orthotropic laminated plates from the concepts of Timoshenko's theory. The equations include the effects of transverse shear and rotary inertia.

Many authors have treated the problem to describe the generation and propagation of bending waves in plates during the last years for example El-Raheb (1994), Fair and Chauer (1979), and Scruby and Drain (1990). Most of them have used numerical methods but very few analytical ones.

Recently Lindblom et al. (1997) derived explicit formulas describing the propagation of bending waves in infinite plates. The standard Kirchoff plate equation for anisotropic plates was used. In this paper we make a complementary study. In particular we extend the equations derived by Chow to include anisotropic plates and derive an integral representation formula for the bending waves. In both of these papers theoretical results are compared with experimental results.

This paper is organized in the following way: In Section 2 the differential equations derived by Chow are introduced. In Section 3 a formal solution of a general case is derived. We also consider three special cases where the load is a Dirac pulse in space and time, which are compared with the experimental data. Comparisons between different solutions and the experimental data are made in Section 4. Finally Section 5 contains discussions and some conclusions.

2 The Chow Model for Bending Wave Propagation in Anisotropic Plates

The differential equations derived by Chow (1971) for an orthotropic laminated plate include the effects of both transverse shear and rotary inertia. In the following these differential equations are enlarged to encompass also anisotropic plates, i.e., the L_1-operator, defined below, will contain D_{16} and D_{26}, not both $= 0$. The plate is assumed to be infinite.

The equations of flexural motion for the total out-of-plane displacement $w(x, y) = w_b(x, y) + w_s(x, y)$, $(x, y) \in \Re^2$ of an anisotropic plate are

$$\begin{cases} L_1 w_b + L_2 w_s = \phi \dfrac{\partial^2}{\partial t^2} \nabla^2 w_b, \; t > 0 \\ L_2 w_s + q = \rho h \dfrac{\partial^2}{\partial t^2} (w_b + w_s), \; t > 0 \\ w_b(x, y, 0) = w_s(x, y, 0) = 0 \\ \dfrac{\partial}{\partial t} w_b(x, y, 0) = \dfrac{\partial}{\partial t} w_s(x, y, 0) = 0 \end{cases} \tag{1}$$

where

$$L_1 = D_{11} \frac{\partial^4}{\partial x^4} + 4D_{16} \frac{\partial^4}{\partial x^3 \partial y} + 2(D_{12} + 2D_{66}) \frac{\partial^4}{\partial x^2 \partial y^2} + 4D_{26} \frac{\partial^4}{\partial x \partial y^3} + D_{22} \frac{\partial^4}{\partial y^4}$$

and

$$L_2 = D_{55} \frac{\partial^2}{\partial x^2} + D_{44} \frac{\partial^2}{\partial y^2}.$$

w_b and w_s are, respectively, the bending and shear components of the total out-of-plane displacement w.

D_{ij} = flexural moduli
ρ = mean volume density of the plate
h = thickness of the plate

[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.

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.

Manuscript received by the ASME Applied Mechanics Division, May 19, 1997; final revision, July 2, 1998. Associate Technical Editor: V. K. Kinra.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

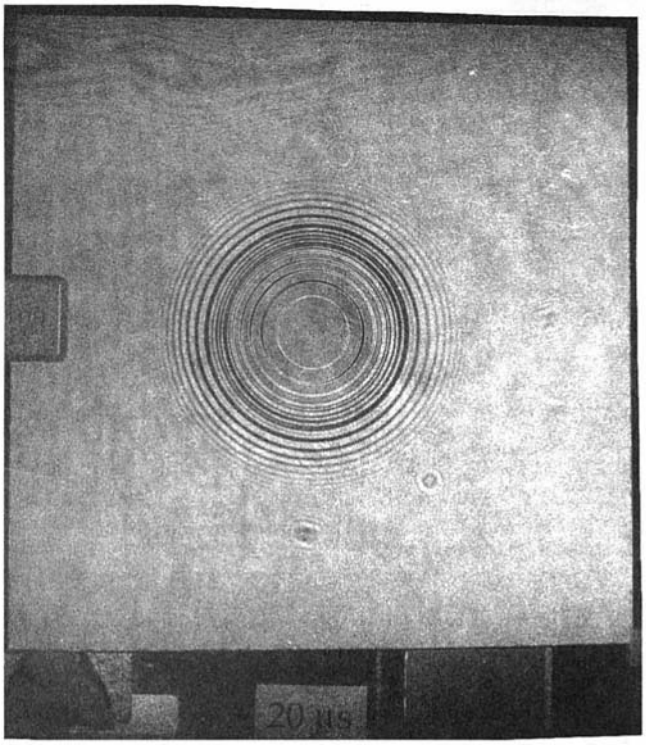


Fig. 1(*a*) The figure shows the interferogram of a steel plate of thickness 0.6 mm at 20 μs after impact

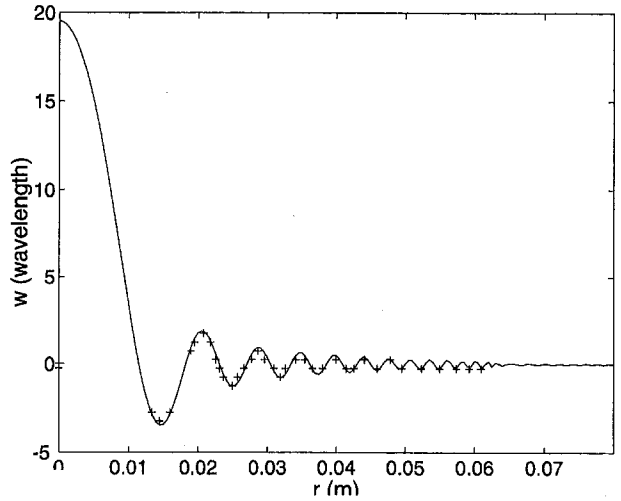


Fig. 1(*b*) A comparison between the general solution (7) (solid curve) and the evaluation of the interferogram in Fig. 1(*a*) (the curve marked with +)

$q = q(x, y, t)$ = the load per unit area
$\phi = \rho h^3/12$
∇^2 = two-dimensional Laplace operator.

3 An Integral Representation Formula for the Solution of (1)

We consider the special case when $q(x, y, t)$ can be separated as $q(x, y, t) = f(t)p(x, y)$.

3.1 A Formal Solution of (1). The influence w_s of the total displacement w is very small and therefore we only consider w_b. Moreover $\hat{p}$ and $\tilde{p}$ denotes the Fourier and Laplace transforms of p, respectively.

PROPOSITION 1. *The solution w_b of the Cauchy problem (1) can formally be written as the following integral representation:*

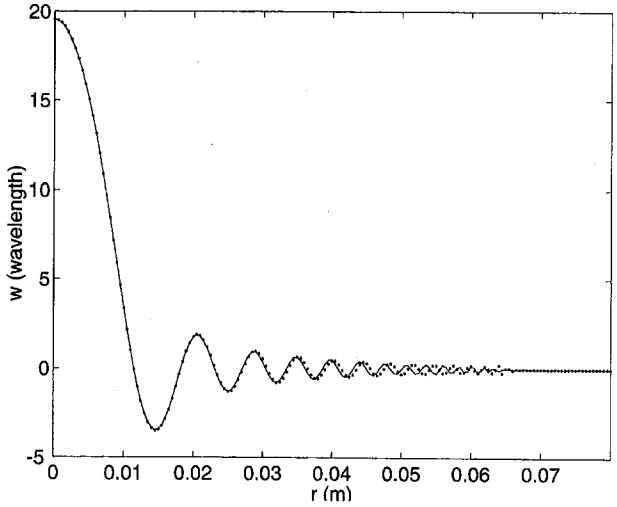


Fig. 1(*c*) A comparison between the general solution (solid curve) and the solution (9) where rotary inertia is neglected (dotted curve)

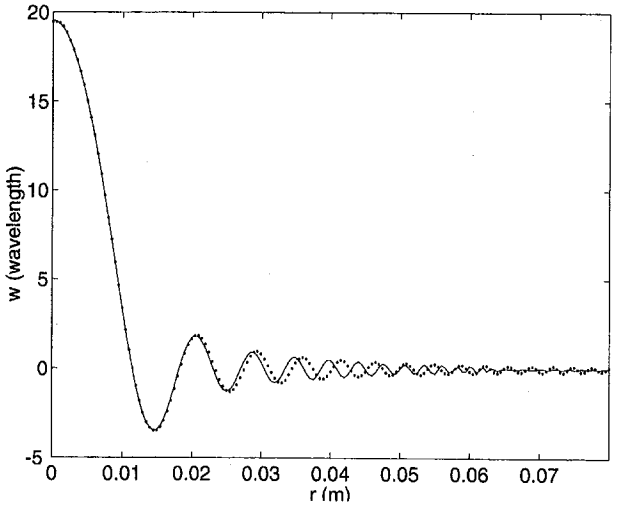


Fig. 1(*d*) A comparison between the general solution (solid curve) and the classical solution (11) (dotted curve)

$$w_b(x, y, t) = \frac{1}{4\pi^2\rho h\phi}\int_0^t f(\tau)d\tau \iint_{\Re^2} \frac{\hat{p}P_2}{P_3(\alpha_2^2 - \alpha_1^2)} \times \left(\frac{\sin \alpha_1(t-\tau)}{\alpha_1} - \frac{\sin \alpha_2(t-\tau)}{\alpha_2}\right) \cos(\xi x + \eta y)d\xi d\eta, \tag{2}$$

where

$$P_1(\xi, \eta) = D_{11}\xi^4 + 4D_{16}\xi^3\eta + 2(D_{12} + 2D_{66})\xi^2\eta^2 + 4D_{26}\xi\eta^3 + D_{22}\eta^4,$$

$$P_2(\xi, \eta) = D_{55}\xi^2 + D_{44}\eta^2,$$

$$P_3(\xi, \eta) = \xi^2 + \eta^2,$$

$$\alpha_1^2 = \frac{a}{2} - \sqrt{\frac{a^2}{4} - b},$$

$$\alpha_2^2 = \frac{a}{2} + \sqrt{\frac{a^2}{4} - b},$$

with

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

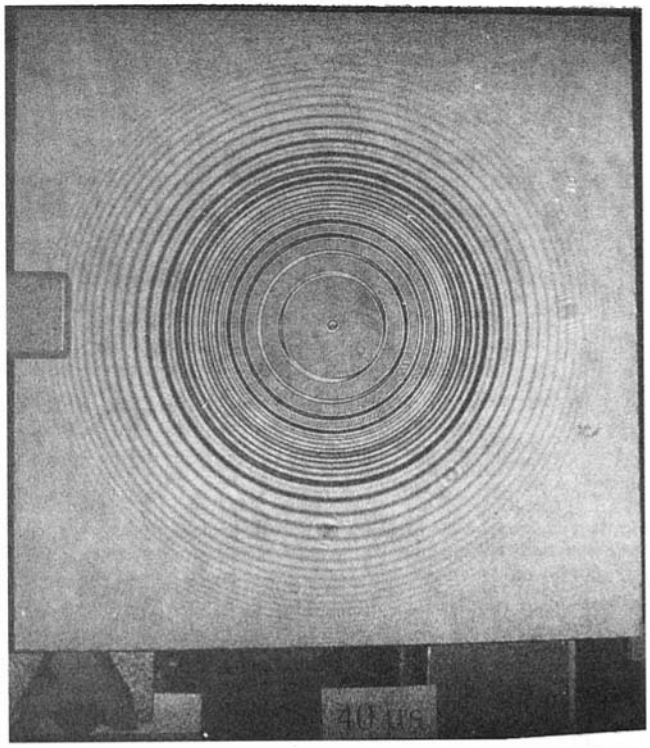


Fig. 2(*a*) The interferogram of the steel plate but now 40 μs after impact

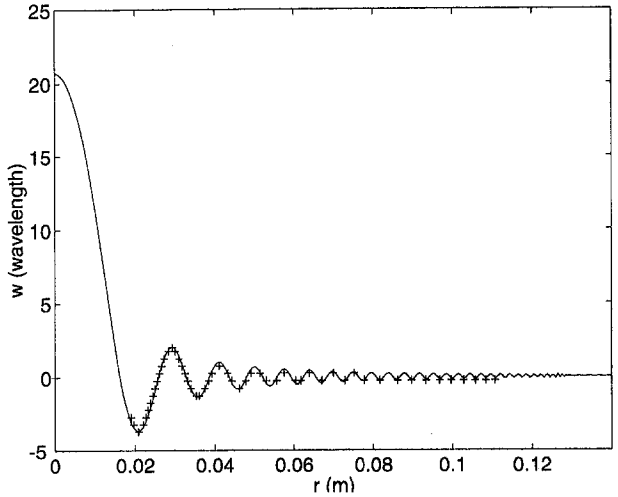


Fig. 2(*b*) The general solution (7) (solid curve) is compared with the experimental results (the curve marked with +)

$$a = \frac{h\rho(P_1 + P_2) + \phi P_2 P_3}{\rho h \phi P_3} \text{ and } b = \frac{P_1 P_2}{\rho h \phi P_3}.$$

For anisotropic plates it holds that $\alpha_1, \alpha_2 \in \Re$.

Proof. The Fourier transformation of (1) yields

$$\begin{cases} P_1\hat{w}_b - P_2\hat{w}_s = -P_3\phi \dfrac{\partial^2}{\partial t^2}\hat{w}_b, \ t > 0 \\ -P_2\hat{w}_s + \hat{q} = \rho h \dfrac{\partial^2}{\partial t^2}(\hat{w}_b + \hat{w}_s), \ t > 0 \\ \hat{w}_b(\xi, \eta, 0) = \hat{w}_s(\xi, \eta, 0) = 0 \\ \dfrac{\partial}{\partial t}\hat{w}_b(\xi, \eta, 0) = \dfrac{\partial}{\partial t}\hat{w}_s(\xi, \eta, 0) = 0 \end{cases} \tag{3}$$

where $\hat{w}(\xi, \eta, t) = \iint_{\Re^2} w(x, y, t)e^{-i(\xi x + \eta y)}dxdy$.

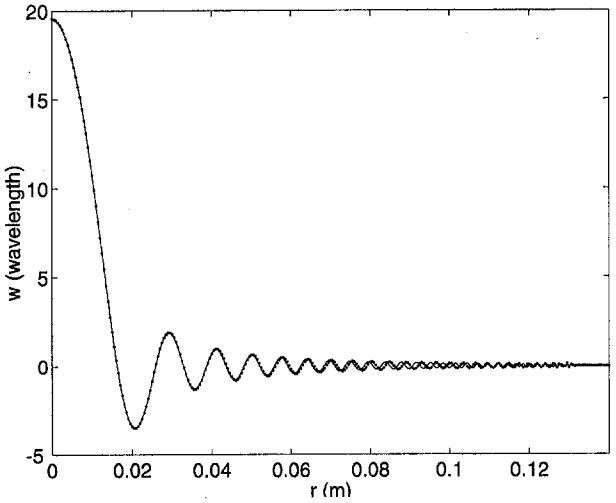


Fig. 2(*c*) A comparison between the general solution (solid curve) and the solution (9) (dotted curve)

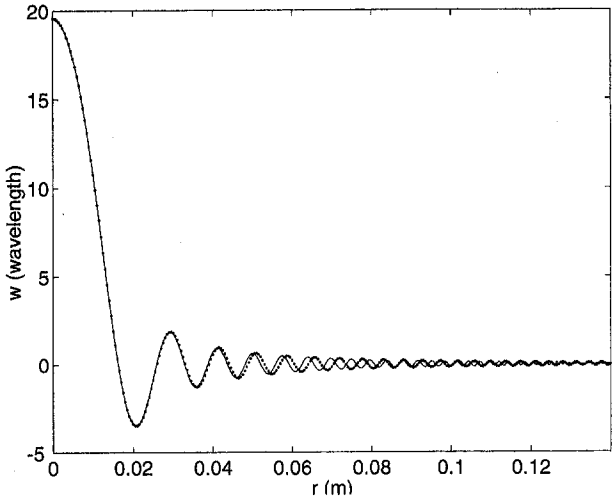


Fig. 2(*d*) A comparison between the general solution (solid curve) and the classical solution (11) (dotted curve)

The Laplace transformation of (3) yields

$$\begin{cases} P_1\tilde{\hat{w}}_b - P_2\tilde{\hat{w}}_s = -s^2 P_3\phi\tilde{\hat{w}}_b, \ t > 0 \\ -P_2\tilde{\hat{w}}_s + \tilde{\hat{q}} = s^2\rho h(\tilde{\hat{w}}_b + \tilde{\hat{w}}_s), \ t > 0 \end{cases} \tag{4}$$

where $\tilde{\hat{w}}(\xi, \eta, s) = \int_0^\infty \hat{w}(\xi, \eta, t)e^{-st}dt$. The solution of (4) with respect to $\tilde{\hat{w}}_b$ is

$$\tilde{\hat{w}}_b = \frac{\tilde{\hat{q}}P_2}{\rho h\phi P_3} \cdot \frac{1}{s^4 + \dfrac{h\rho(P_1 + P_2) + \phi P_2 P_3}{\rho h \phi P_3} s^2 + \dfrac{P_1 P_2}{\rho h \phi P_3}}$$

$$= \frac{\tilde{\hat{q}}P_2}{\rho h\phi P_3} \cdot \frac{1}{\alpha_2^2 - \alpha_1^2} \cdot \left(\frac{1}{s^2 + \alpha_1^2} - \frac{1}{s^2 + \alpha_2^2} \right). \tag{5}$$

From the inverse Laplace transform $L^{-1}\{1/s^2 + \alpha_i^2\} = \sin \alpha_i t/\alpha_i$, $i = 1, 2$, and the convolution theorem it follows that

$$\hat{w}_b(\xi, \eta, t) = \frac{1}{\rho h \phi} \int_0^t \hat{q}(\xi, \eta, \tau) \frac{P_2}{P_3(\alpha_2^2 - \alpha_1^2)} \times \left(\frac{\sin \alpha_1(t - \tau)}{\alpha_1} - \frac{\sin \alpha_2(t - \tau)}{\alpha_2} \right) d\tau$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

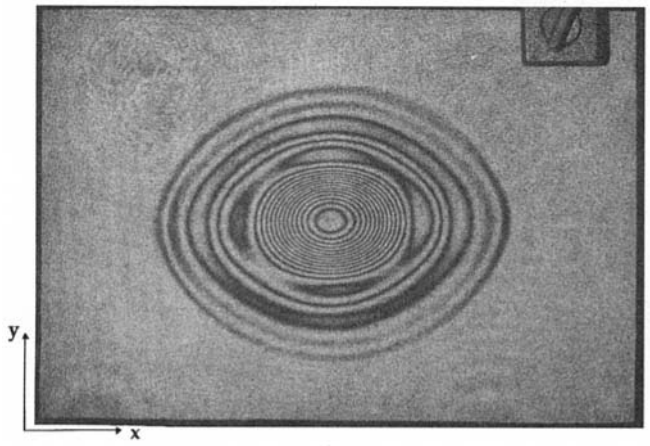


Fig. 3(*a*) The figure shows the interferogram of an orthotropic composite plate of thickness 3.1 mm at 40 μs after impact

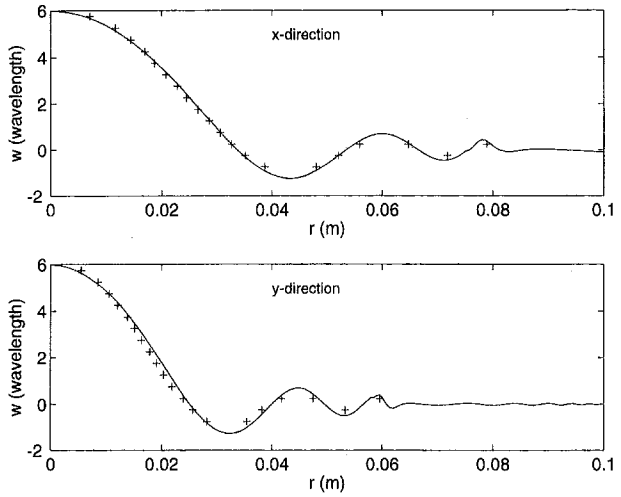


Fig. 3(*b*) A comparison between the general solution (7) (solid curve) in the *x* and *y*-directions and the evaluation of the interferogram in Fig. 3(*a*) (the curve marked with +)

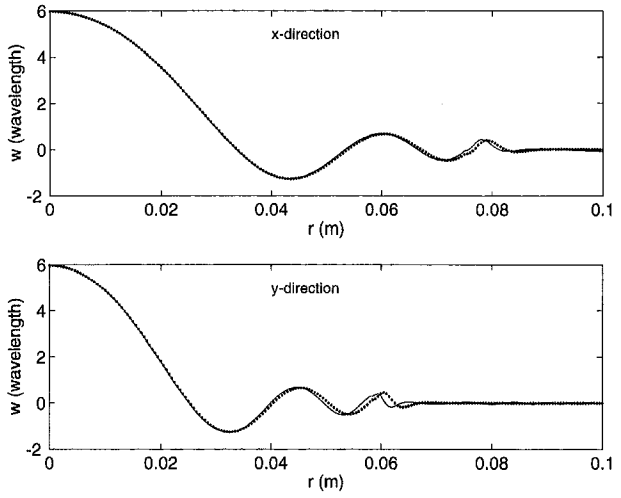


Fig. 3(*c*) A comparison between the general solution (solid curve) and the solution where rotary inertia is neglected (9) (dotted curve)

$$= \frac{\hat{p}}{\rho h \phi} \int_0^t f(\tau) \frac{P_2}{P_3(\alpha_2^2 - \alpha_1^2)} \times \left(\frac{\sin \alpha_1 (t-\tau)}{\alpha_1} - \frac{\sin \alpha_2 (t-\tau)}{\alpha_2} \right) d\tau. \quad (6)$$

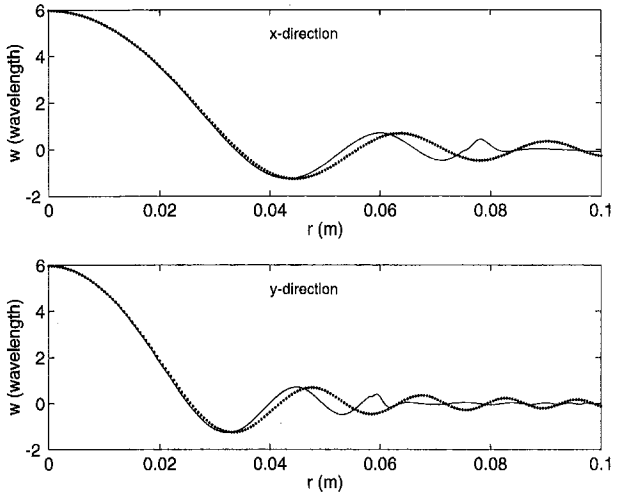


Fig. 3(*d*) A comparison between the general solution (solid curve) and the classical solution (11) (dotted curve)

Apply the inverse Fourier transform on (6) and we obtain (2). The proof is complete. □

3.2 The Special Case Where the Load is the Dirac Delta Function. Let $p(x, y) = \delta(x, y)$ and $f(t) = I\delta(t)$, where, as usual, δ denotes the Dirac delta function. I is the impulse transferred to the plate. Then (1) has the solution

$$w_b(x, y, t) = \frac{I}{4\pi^2 \rho h \phi} \iint_{\mathfrak{R}^2} \frac{P_2}{P_3(\alpha_2^2 - \alpha_1^2)} \times \left(\frac{\sin \alpha_1 t}{\alpha_1} - \frac{\sin \alpha_2 t}{\alpha_2} \right) \cos(\xi x + \eta y) d\xi d\eta. \quad (7)$$

In fact, (7) follows from (2) and the fact that the Fourier transform of $\delta(x, y)$ is $\hat{\delta} = \hat{\delta}(\xi, \eta) \equiv 1$.

Moreover, we note that Fällström et al. (1996, 1997) showed that it is possible to experimentally produce an impact pulse which can be regarded as a remarkable good approximation of a Dirac pulse both in space and time, so this example is very important in the sense that it can be used for comparisons with our experimental data.

It is sometimes convenient to introduce the polar coordinates

$$\begin{cases} x = r \cos(\theta) \\ y = r \sin(\theta) \end{cases} \quad \text{and} \quad \begin{cases} \xi = \sigma \cos(\varphi) \\ \eta = \sigma \sin(\varphi) \end{cases},$$

and rewrite (7) as

$$w_b(r, \theta, t) = \frac{I}{4\pi^2 \rho h} \int_0^{2\pi} \int_0^{\infty} \frac{\sigma}{d} \left(\frac{\sin \gamma_1 t}{\gamma_1} - \frac{\sin \gamma_2 t}{\gamma_2} \right) \times \cos(\sigma r \cos(\varphi - \theta)) d\sigma d\varphi,$$

where

$$\gamma_1^2 = \frac{6S_2}{\rho h^3}(c - d),$$

$$\gamma_2^2 = \frac{6S_2}{\rho h^3}(c + d),$$

$$c = \sigma^2 \left(\frac{h^2}{12} + \frac{S_1}{S_2} \right) + 1,$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

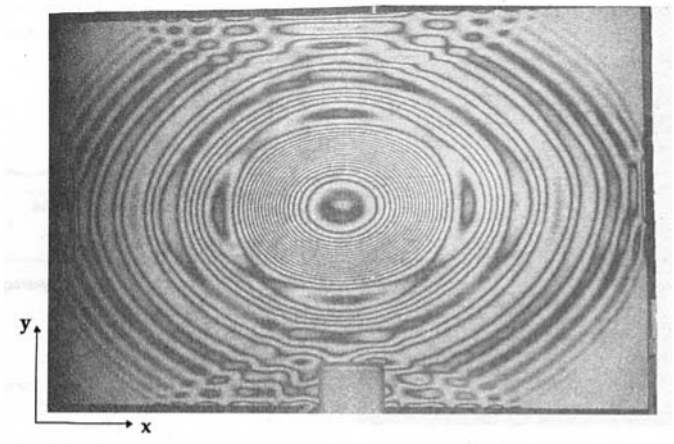


Fig. 4(*a*) The figure shows the interferogram of an orthotropic composite plate of thickness 3.1 mm at 80 μs after impact

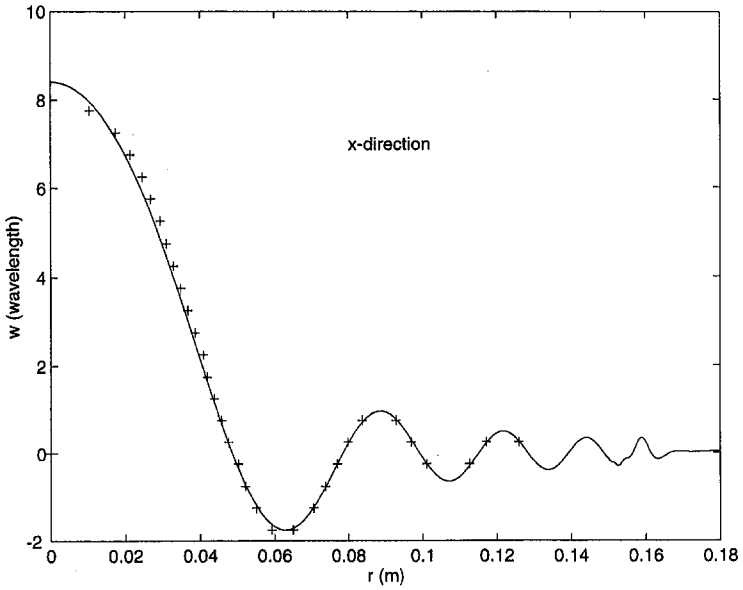


Fig. 4(*b*) A comparison between the general solution (7) (solid curve) in the *x*-direction and the evaluation of the interferogram in Fig. 4(*a*) (the curve marked with +)

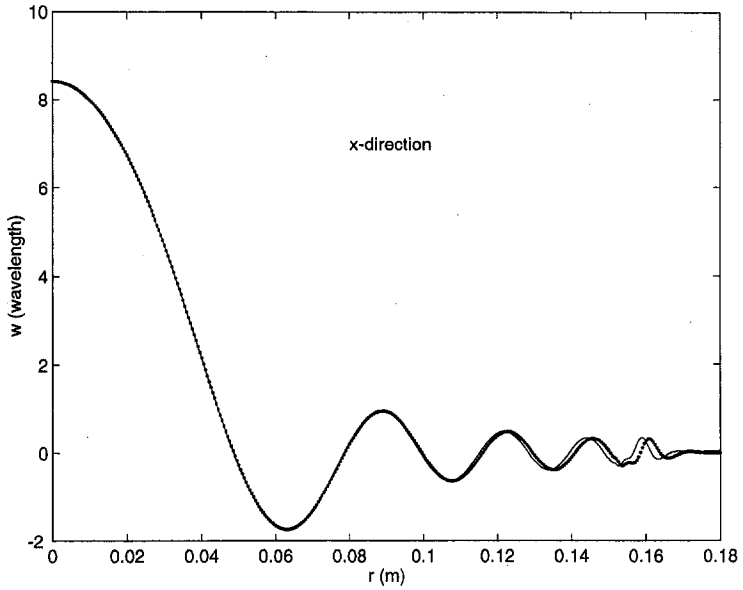


Fig. 4(*c*) A comparison between the general solution (solid curve) and the solution where rotary inertia is neglected (9) (dotted curve)

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

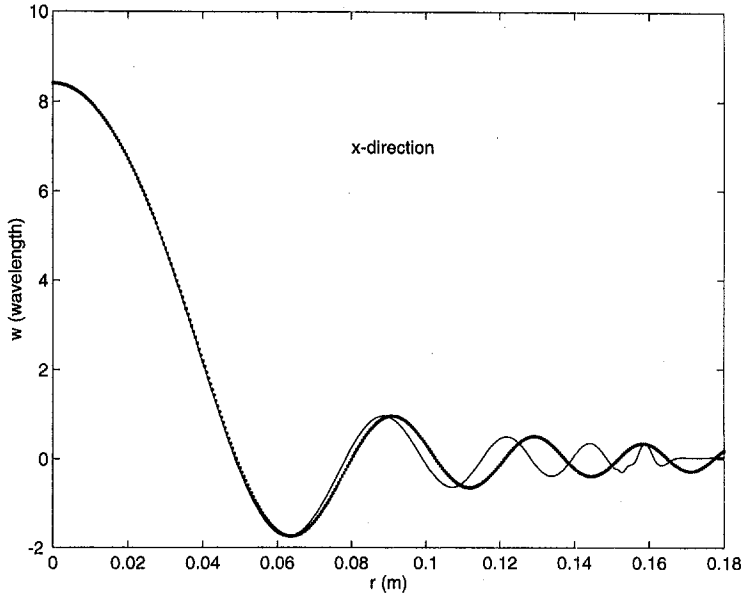


Fig. 4(*d*) A comparison between the general solution (solid curve) and the classical solution (11) (dotted curve)

$$d = \sqrt{\sigma^4\left(\frac{h^2}{12} - \frac{S_1}{S_2}\right)^2 + 2\sigma^2\left(\frac{h^2}{12} + \frac{S_1}{S_2}\right) + 1},$$

$$S_1 = D_{11}\cos^4\varphi + 4D_{16}\cos^3(\varphi)\sin(\varphi) + 2(D_{12} + 2D_{66})\cos^2(\varphi)\sin^2(\varphi) + 4D_{26}\cos(\varphi)\sin^3(\varphi) + D_{22}\sin^4\varphi,$$

$$S_2 = D_{55}\cos^2(\varphi) + D_{44}\sin^2(\varphi).$$

3.3 The Influence of the Rotary Inertia. If the rotary inertia of the plate is neglected, then (1) can be written

$$\begin{cases} L_1 w_b + L_2 w_s = 0, \, t > 0 \\ L_2 w_s + q = \rho h \dfrac{\partial^2}{\partial t^2}(w_b + w_s), \, t > 0 \\ w_b(x, y, 0) = w_s(x, y, 0) = 0 \\ \dfrac{\partial}{\partial t} w_b(x, y, 0) = \dfrac{\partial}{\partial t} w_s(x, y, 0) = 0. \end{cases} \tag{8}$$

Let $q(x, y, t) = If(t)p(x, y)$. We can derive w_b with the same technique as above and obtain

$$w_b(r, \theta, t) = \frac{I}{4\pi^2\rho h}\int_0^{2\pi}\int_0^{\infty} \frac{\sigma}{\sigma^2\dfrac{S_1}{S_2} + 1} \times \frac{\sin\gamma t}{\gamma}\cos(\sigma r\cos(\varphi - \theta))d\sigma d\varphi, \tag{9}$$

where $\gamma^2 = (\sigma^4 S_1/\rho h(\sigma^2(S_1/S_2) + 1))$ and S_1, S_2 are defined at the end of Section 3.2.

3.4 The Solution for the Classical Laminated Plate Equation. If both the rotary inertia and the transverse shear are neglected, then we turn to the case with the classical laminated plate equation. The differential equations of (1) will then be reduced to

$$L_1 w + \rho h \frac{\partial^2}{\partial t^2} w = q. \tag{10}$$

This is the Kirchoff equation, and the solution for $q(x, y, t) = If(t)p(x, y)$ in polar coordinates is (see, e.g., Remark 3.5)

$$w(r, \theta, t) = \frac{I}{4\pi^2\rho h}\int_0^{2\pi}\int_0^{\infty} \sigma\frac{\sin\gamma t}{\gamma} \cos(\sigma r\cos(\varphi - \theta))d\sigma d\varphi, \tag{11}$$

where $\gamma^2 = \sigma^4 S_1/\rho h$ and S_1 is defined at the end of Section 3.2.

4 Comparisons Between Different Solutions

4.1 The Experiments. Double-pulsed holographic interferometry has been used to get interferograms of propagating bending waves. The holographic set-up is a standard one and is described in, for example, Fällström et al. (1997). The light source, a ruby laser, is used both to record and generate the bending waves. The laser pulse is of 30-ns duration and has an energy of 0.2–0.4 J. The part of the pulse used to generate the waves is deflected via a mirror to a positive lens, which focuses the beam onto the surface of the plate. The focused spot is of about 0.4 mm in diameter. The power density will be in order of 109 W/cm^2. Most material will ablate under such intense irradiation.

To increase the transferred impulse and to avoid thermal gradients close to the molten pool (Fällström et al., 1997) a piece of gelatine (water and gelatine powder) is placed on the plate surface at the "impact" center. That is, it is now the ablation of gelatine and not of the plate substrate that will take place. The impact can now be regarded as a Dirac pulse both in space and time (Fällström, 1997).

4.2 Parameters of the Plates. In the following section the effects of transverse shear modulus and rotary inertia of the propagating transient bending waves are treated. To get realistic values of the parameters, reference values are taken from plates treated in Fällström et al. (1997) and Fällström and Jonsson (1991). The comparisons are made with one isotropic, one orthotropic, and one anisotropic plate. The isotropic plate is a steel plate and the orthotropic and the anisotropic plates are made of glass-fiber reinforced epoxy. The parameters for the three plates are:

Isotropic Plate
Size = 250 × 250 mm
Thickness = 0.6 mm

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

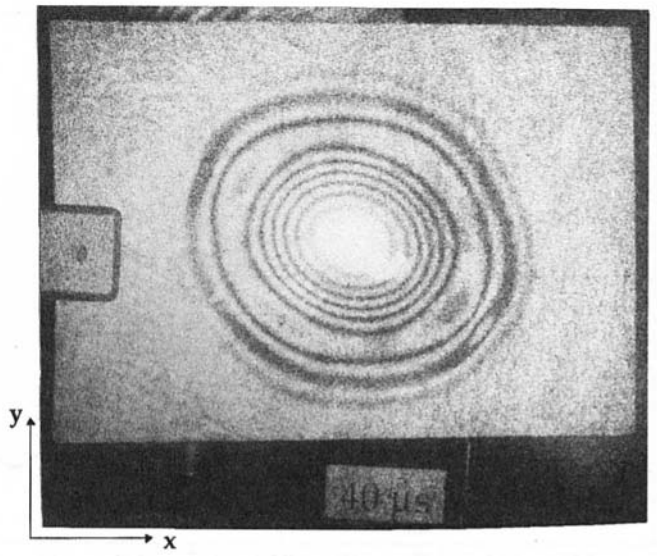


Fig. 5(*a*) The figure shows the interferogram of an anisotropic composite plate of thickness 4.1 mm at 40 μs after impact

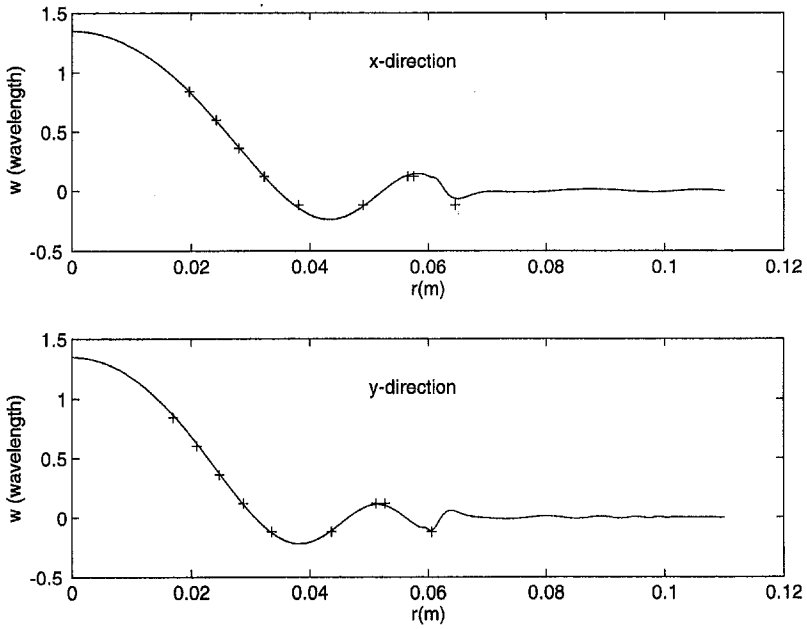


Fig. 5(*b*) A comparison between the general solution (7) (solid curve) in the *x*-direction and the evaluation of the interferogram in Fig. 5(*a*) (the curve marked with +)

Density = 7800 kg/m^3
Young's modulus = 180 GPa
D66 = 1.386 Nm
D11 = D22 = D12 + 2·D66 = 3.386 Nm
D44 = D55 = 12·D66/h^2 N/m

Orthotropic Plate
Laminate sequence = (2.5/−2.5)3
Size = 207.5 × 185 mm
Thickness = h = 3.1 mm
Density = r = 1993 kg/m^3
D11 = 103 Nm
D16 = 0 Nm
D12 = 15 Nm
D66 = 11 Nm
D26 = 0 Nm
D22 = 32 Nm
D44 = 18.6·D66/h^2 N/m
D55 = 11.2·D66/h^2 N/m

Anisotropic Plate
Laminate sequence = (30/−30)3
Size = 210 × 149 mm
Thickness = h = 4.1 mm
Density = r = 1957 kg/m^3
D11 = 135 Nm
D16 = 5.05 Nm
D12 = 10 Nm
D66 = 51 Nm
D26 = −20.35 Nm
D22 = 84 Nm
D44 = 6.0·D66/h^2 N/m
D55 = 5.7·D66/h^2N/m

4.3 Results. Figures 1(*a*) and 2(*a*) show the interferograms for the propagating bending waves in the isotropic plate 20 μs and 40 μs, respectively, after the laser pulse hits the plate. Since it is difficult to resolve the fringes near the impact point

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

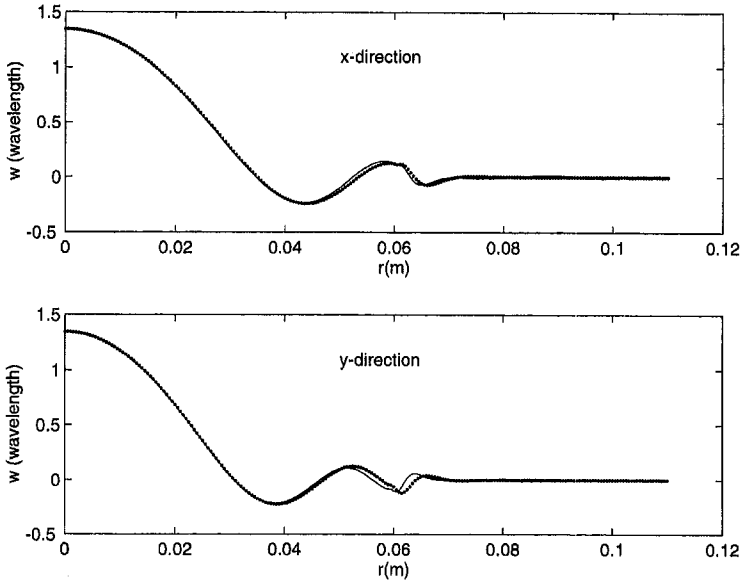


Fig. 5(*c*) A comparison between the general solution (solid curve) and the solution where rotary inertia is neglected (9) (dotted curve)

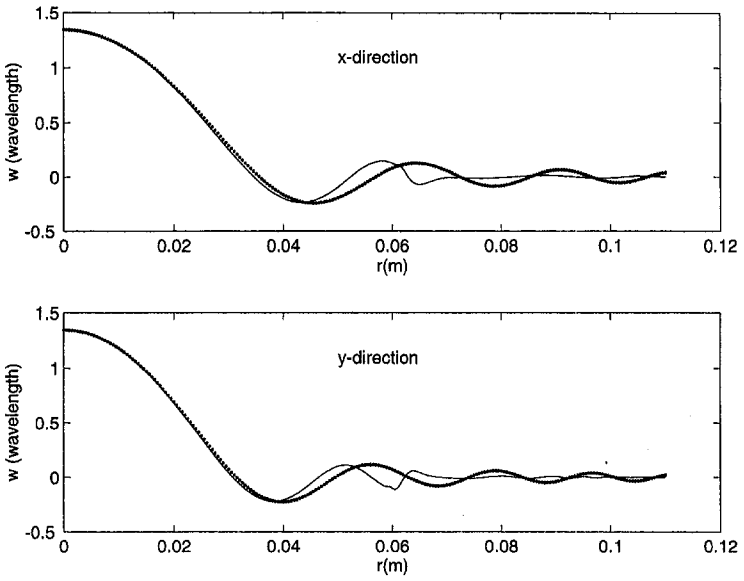


Fig. 5(*d*) A comparison between the general solution (solid curve) and the classical solution (11) (dotted curve)

in the interferograms, the amplitudes of the theoretical curves are normalized to the amplitude of the extremes outside the central peak of the experimental curve. The impulse transferred to the plate is about $0.6 \cdot 10^{-3}$ Ns. Figures $1(b)$ and $2(b)$ show the measured deformation (marked with +) obtained along the center line of the interferograms. Theoretical deformation curves (the solid lines) obtained from Eq. (7) have been added as a comparison with the experimental ones. The conclusion is that the general theory from the foregoing section describes the experimental results very good.

Figures $1(c)$ and $2(c)$ are comparisons between the general solution (7) (solid curves) and the solution (9) where the effect of rotary inertia is neglected (the dotted curves). Finally, Figs. $1(d)$ and $2(d)$ show the general solution (solid curves) compared with the classical solution (11) (the dotted curves).

Figures 3 and 4 show analogous curves comparisons for the orthotropic plate and Fig. 5 the same for the anisotropic plate. Figures 3 and 5 show the propagating bending waves along the x-direction 40 μs after the impact of the plate (parallel with the longer side of the plate) and the y-direction (parallel with the shorter side of the plate) and Fig. 4 shows the same but 80 μs after the impact. As can be observed in Fig. $4(a)$ the wave in the x-direction is almost undisturbed while the wave in the y-direction has interacted with the plate boundaries. Therefore we only study propagation waves in the x-direction. In the x-direction the comparisons between the theoretical and the experimental results have been made only for times where the reflected waves do not disturb the outgoing waves.

5 Discussions and conclusions

The following conclusions can be drawn:

1 The agreement between the general solution (7) and the experimental data is very good both for all three kinds of plates. The theory also shows none or very little amplitude outside the area where the fringes at the interferograms are missing.

2 Near the point of impact the rotary inertia can be neglected (see the (c) figures). Further away from the point of impact the agreement is bad and it is necessary to use the

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

general solution. In conclusion, if one wants to describe all of the transient propagating bending waves it is necessary to use the general theory.

3 If the general solution (7) is compared with the classical one, (11), it is necessary to make a closer analysis of the central peak. For a thin plate the classical theory seems to be applicable to about the second or third extremes if the times after the impact are relatively short. This is not the case for thicker plates (see Fig. 5(*a*)) where a notable discrepancy is found even at the first zero crossing.

4 One parameter of interest when studying propagation bending waves is the phase speed. A flexural wave can be looked upon as a sum of plane waves of the form $e^{i\cdot(k_x\cdot x+k_y\cdot y-\omega\cdot t)}$, where k_x and k_y are the propagation constants in the x and y-directions, respectively, and ω is the angular frequency. The phase velocities are then $v_x = \omega/k_x$ and $v_y = \omega/k_y$. Provided that the load $q = 0$ in Eqs. (1), (8), and (10), the velocities for the different cases can be estimated. Moreover, it is interesting that both for the general case and the one where the rotary inertia is neglected, the velocity will be finite but for the classical case the velocity will be infinite for large propagating constants. This is unphysical and demonstrates a limitation of the classical model used.

5 The figures also show that the wavelength decreases with increasing r, which means that high-frequency waves travel faster than low-frequency ones, as expected for plate waves.

6 It is possible to estimate the values of D_{44} and D_{55}, that is the transverse shear, with the solutions obtained in this paper. These parameters can be estimated by the fact that a change in the value of D_{44} will influence the wave propagation very much in the y-direction but almost nothing in the x-direction. The other way around holds for the D_{55} parameter.

References

Chow, T. S., 1971, "On the Propagation of Flexural Waves in an Orthotropic Laminated Plate and its Response to an Impulsive Load," *J. of Composite Materials,* Vol. 5, pp. 306–319.

El-Raheb, M., 1994, "Transient Waves in a Thick Disc," *J. Acoust. Soc. Am.,* Vol. 95, No. 6, pp. 3287–3299.

Fairand, B. P., and Clauer, A. H., 1979, "Laser Generation of High Amplitude Stress Waves in Materials," *J. Appl. Phys.,* Vol. 50, No. 3, pp. 1497–1502.

Fällström, K.-E., and Jonsson, M. A., 1991, "A Non Destructive Method to Determine Material Properties in Anisotropic Plates," *Polymer Composites,* Vol. 12, pp. 293–305.

Fällström, K.-E., Molin, N.-E., Olofsson, K., Schedin, S., and Wahlin, A., 1997, "Mechanical and Thermal Effects in a Steel Plate Impacted by a Focused Laser Pulse," *Nondestructive Testing and Evaluation,* Vol. 13, pp. 293–307.

Fällström, K.-E., Olofsson, K., Saldner, H. O., and Wahlin, A., 1997, "Dynamic Material Parameters in an Anisotropic Plate Estimated in Phase-Stepped Holographic Interferometry," *Optics and Lasers in Engineering,* Vol. 24, pp. 429–454.

Lindblom, O., Näslund, R., Persson, L.-E., and Fällström, E.-E., 1997, "A Study of Bending Waves in Infinite and Anisotropic Plates," *Appl. Math,* Vol. 42, No. 3, pp. 213–232.

Lord Rayleigh, 1945, *Theory of Sound,* Vols. 1 and 2, Dover, New York (originally published in 1877).

Mindlin, R. D., 1951, "Influence of Rotary Inertia and Shear on Flexural Motions of Isotropic, Elastic Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 18, pp. 31–38.

Mindlin, R. D., Schacknow, A., and Deresiewicz, H., 1956, "Flexural Vibrations of Rectangular Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 23, pp. 430–436.

Reissner, E., 1921, "The Effect of Transverse Shear Deformation on the Bending of Elastic Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 12, p. A-69.

Scruby, C. D., and Drain, L. E., 1990, *Laser Ultrasonics,* Hilger, London, UK.

Thorkildsen, R. L., and Hoppmann, W. H., 1959, "Effect of Rotary Inertia on the Frequencies of Vibration of Stiffened Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 26, pp. 298–300.

Timoshenko, S., 1921, "On the Correction for Shear of the Differential Equations for Transverse Vibrations of Prismatic Bars," *Phil Mag.,* Ser. 6, Vol. 41, p. 742.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

A New Algebraic Rigid-Body Collision Law Based on Impulse Space Considerations

A. Chatterjee
Department of Engineering Science and Mechanics,
Pennsylvania State University,
University Park, PA 16802
e-mail: anindya@crash.esm.psu.edu

A. Ruina
Department of Theoretical and Applied Mechanics
Cornell University,
Ithaca, NY 14853
e-mail: ruina@cornell.edu

We present a geometric representation of the set of three-dimensional rigid-body collisional impulses that are reasonably permissible by the combination of non-negative post-collision separation rate, non-negative collisional compression impulse, non-negative energy dissipation and the Coulomb friction inequality. The construction is presented for a variety of special collisional situations involving special symmetry or extremes in the mass distribution, the friction coefficient, or the initial conditions. We review a variety of known friction laws and show how they do and do not fit in the permissible region in impulse space as well as comment on other attributes of these laws. We present a few parameterizations of the full permissible region of impulse space. We present a simple generalization to arbitrary three-dimensional point contact collisions of a simple law previously only applicable to objects with contact-inertia eigenvectors aligned with the surface normal and initial relative tangential velocity component (e.g., spheres and disks). This new algebraic collision law has two restitution parameters for general three-dimensional frictional single-point rigid-body collisions. The new law generates a collisional impulse that is a weighted sum of the impulses from a frictionless but nonrebounding collision and from a perfectly sticking, nonrebounding collision. We describe useful properties of our law; show geometrically the set of impulses it can predict for several collisional situations; and compare it with existing laws. For simultaneous collisions we propose that the new algebraic law be used by recursively breaking these collisions into a sequence ordered by the normal approach velocities of potential contact pairs.

1 Introduction

Rigid-body simulations involving collisional contact necessarily use some kind of contact constitutive assumptions. Such collisional contact constitutive assumptions fall into three main classes: *full-deformation* laws; *incremental* laws (also called "soft"); and *algebraic* laws (also called "hard"—see Goyal et al., 1994; Walton, 1992). Full-deformation laws are based on continuum mechanics equations for the whole body, most often linear elastodynamic equations. Incremental laws involve some kind of micromechanical model of the deformation and slip in the contact region, with the assumption that the motion of the bulk of the bodies is described by rigid-body differential equations of motion during the collision. Algebraic laws attempt to describe the *net* collisional interaction by algebraic equations that relate precollision and post-collision quantities. Thus, algebraic laws are generally based on less stringent rigidity requirements than incremental laws (see, e.g., Chatterjee and Ruina, 1998a). One interpretation of incremental calculations is that they are a means to calculate a reasonable net impulse and not a literal description of the true contact mechanics. In this sense incremental models are equivalent to algebraic laws where function evaluations involve the solution of differential equations.

Unfortunately no known collision law of any kind is both predictive and accurate, in the sense, say, that Newton's laws, or even linear elasticity or the Navier-Stokes equations can be accurate. For a given pair of bodies, the collisional outcome may well depend on not just the initial velocities, masses, and mass moments of inertia, but also on some combination of contact shape, contact-mechanics, surface chemistry, friction, fracture, and vibration phenomena that are not well understood, especially a priori. Further, in the case of multiple superficially simultaneous collisions, the prediction of the collisional outcome often depends so sensitively on initial conditions that sufficiently accurate initial conditions cannot be expected to be known by a simulator. But even if all modeling ingredients are accurately known, which they generally are not, there are technical difficulties that prevent simple calculation of collisional outcomes. Partial differential equations involving bulk and/or contact deformation are time-consuming to solve (and to set up to solve). There is no reason to believe that, in general, an accurate continuum model can be well approximated by treating the body as rigid everywhere except in a localized quasi-static region describable by ordinary differential equations (as demanded by incremental laws). Finally, there is no reason to expect that the outcome of detailed modeling or exhaustive experimentation has a tractable summarizing description with standard functions or even lookup tables that apply equally well to a wide variety of bodies and their collisions (as is demanded by algebraic collision laws).

Any generally applicable collision law, whether coming from detailed continuum modeling, approximating ordinary differential equations, or summarizing functions will be highly approximate unless applied to a narrow range of collisional situations.

But simulators will simulate collisions with or without an accurate constitutive law. What outcomes should and should not be reasonably allowed in such simulations, and what basic phenomena may one hope to easily capture in such simulations? How can one compare various available collision laws? Can better collision laws be constructed? To help answer these questions it is perhaps useful to describe algebraically and geometrically the basic restrictions that collisional interaction rules might

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, July 15, 1997; final revision, Mar. 20, 1998. Associate Technical Editor: S. W. Shaw.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

be reasonably expected to obey, as we do in Section 3 of this paper. In Section 4 we describe how several presently available collision laws fit into this scheme. In Section 5 we present a new simple algebraic collision law which is a crude description of reality at best, but which overcomes some of the shortcomings in other known laws. And finally in Section 6 we describe a way to use the new law to predict a plausible outcome for multiple simultaneous collisions.

2 Preliminaries

The common basic assumptions of single point contact rigid-body collision theory are that the collisional interaction involves forces that are large compared to body forces and centripetal inertial terms; that the contact forces occur in a region of spatial extent negligibly small compared to body dimensions; and that at times before and soon after the collision velocities of points on the bodies can be accurately calculated using rigid-body relations (using the same positions and orientations before and after the collision). We do not consider here the possibility of non-negligible contact moments about the contact point.[1]

Based on the foregoing assumptions, the net collisional interaction between an arbitrary pair of objects colliding in a known configuration must satisfy impulse-momentum relations of the form

$$\mathbf{P} = \mathbf{M}\cdot\Delta\mathbf{V} = \mathbf{M}\cdot(\mathbf{V}_f - \mathbf{V}_i), \tag{1}$$

where $\mathbf{P}$ is the collisional contact impulse transmitted from one object to the other (we arbitrarily pick one object as the reference object), $\mathbf{V}$ is the velocity of the contact point of the second object relative to the contact point of the reference object, $\Delta\mathbf{V}$ is the change in $\mathbf{V}$ from its precollision value of $\mathbf{V}_i$ to its post-collision value of $\mathbf{V}_f$, and $\mathbf{M}$ is a second-order rank 3 tensor with units of mass, which depends on the mass distributions of both colliding objects. All symmetric positive definite $\mathbf{M}$ have physical realizations in terms of pairs of unconstrained objects of finite mass (Chatterjee and Ruina, 1998b). $\mathbf{M}$ remains rank 3 for typical single-point collisions of linkages consisting of multiple interconnected objects with frictionless constraints/bearings. Given a mechanism's configuration and mass distribution at impending contact $\mathbf{M}$ can be found using standard (non-collisional) rigid-body dynamics (see A1 of Appendix). Derivations of Eq. (1) are outlined in many references (for discussions, see e.g., Smith (1991) or Chatterjee (1997)). Equation (1) incorporates linear and angular momentum balance; these principles can provide no more information. Either $\mathbf{P}$ or $\Delta\mathbf{V}$ describes the effect of the collision on the rigid-body system. For example, if $\mathbf{P}$ is known, then the motion of the system is found by applying $\mathbf{P}$ and $-\mathbf{P}$ to the two contacting bodies.[2] All algebraic rigid-body collision laws start from Eq. (1) (or relations equivalent to it; e.g., see Brach (1991)), and then hypothesize three additional conditions (two in two dimensions) that vary from law to law so that Eq. (1) along with the collision law determine the six unknowns (four in two dimensions) $\mathbf{P}$ and $\Delta\mathbf{V}$. Incremental laws start with the differential form of Eq. (1) (thus requiring stronger rigidity, see Chatterjee and Ruina (1998a)), and then hypothesize specific types of incremental contact behavior that vary from law to law; integration of the resulting differential equations up to a hypothesized termination criterion then determines the outcome of the collision.

When needed, we assume that at least one of the objects is geometrically smooth at the contact point so a well-defined tangent plane exists with normal $\mathbf{n}$ directed out from the reference object in the n or 1 direction. For definiteness, we take the first tangential or 2-direction to oppose the tangential component of precollision contact point relative velocity. The other tangential or 3-direction is orthogonal to the 1 and 2-directions. We use this coordinate system for discussion and graphics. Except when noted, the formulas we present have coordinate-free interpretations. In coordinate form Eq. (1) is

$$P = M(V_f - V_i), \tag{2}$$

where the 3×1 column vectors P, V_i and V_f, and the 3×3 local mass matrix M are the component forms of $\mathbf{P}$, $\mathbf{V}_i$, $\mathbf{V}_f$, and $\mathbf{M}$. Further assumptions about the collisional impulse are described below.

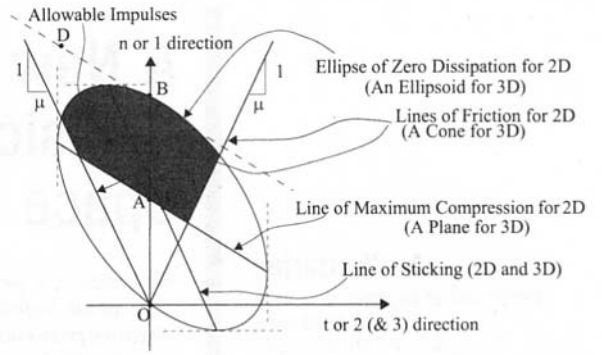


Fig. 1 Two-dimensional schematic of physically permissible region in impulse space for a generic collision

3 The Permissible Region in Impulse Space

The following description of permissible collision impulses contains few new ideas, but is absent from much of the published collision literature (including texts and monographs like Goldsmith (1960) and Brach (1991)). Brogliato (1996) discusses the permissible region in the space of collision parameters. The geometric interpretation (in the space of the collision impulse) that we present below motivates the new collision law, and complements Rubin's recent paper (1998) as well as Brogliato's treatment (both of which attempt to describe the constraints on the collisional interaction in terms of mathematical expressions without the equivalent, explicit geometric description in impulse space).

We now describe various points, lines, and surfaces in impulse space as discussed in part by many (e.g., Goldsmith (1960), Wang and Mason (1992), and Smith (1991)) but perhaps first brought together in Chatterjee (1997).

An equivalent discussion could be presented in contact-relative-velocity space, but the description of friction is not so convenient there.

3.1 Points, Lines, and Surfaces in Impulse Space. See Figs. 1 and 2. Each point in impulse space $P = (P_1, P_2, P_3)$ corresponds to a candidate outcome for a collision with given M and V_i.

The origin $P = 0$ of impulse space corresponds to no collisional interaction and $V_f = V_i$. The normal impulse which brings the normal velocity to zero (perfectly plastic and frictionless) is

$$P_I = -\left(\frac{n^T V_i}{n^T M^{-1} n}\right) n \tag{3}$$

which in our chosen coordinates is $(-V_{i1}/[M^{-1}]_{11}, 0, 0)$. The impulse required to bring the contact points to a stop relative to each other with no bounce and no slip (perfectly plastic and perfectly sticking), is

[1] See Brach (1991) for an example of a nongeneric collision where the contact moment is qualitatively important. Chatterjee and Ruina (1998a) discuss why the contact moment is likely to be unimportant for most collisions.

[2] Special treatment is sometimes required for systems with special constraints such that an eigenvalue of $\mathbf{M}$ is infinite and $\mathbf{P}$ cannot be determined from $\Delta\mathbf{V}$ alone; or bodies with all mass on a straight line so that $\mathbf{M}$ is singular and $\Delta\mathbf{V}$ cannot be determined from $\mathbf{P}$ alone.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

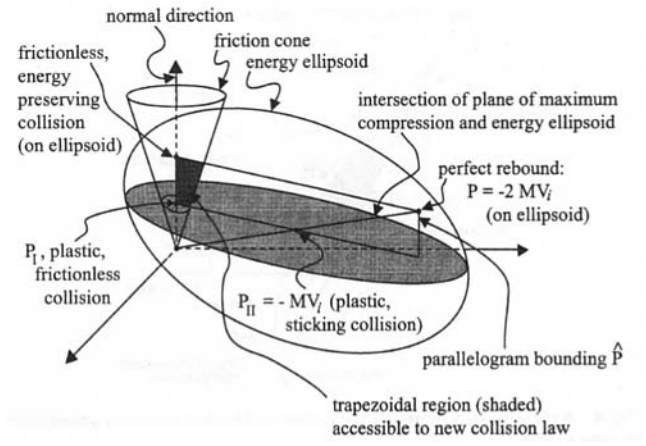


Fig. 2 Schematic diagram showing two-dimensional region in three-dimensional impulse space accessible to the new collision model

$$P_{II} = -MV_i. \tag{4}$$

For any impulse P the relative velocity of the contact points satisfies $P - P_{II} = MV_f$ and we can think of P_{II} as a reference point for impulses, corresponding to the reference state of zero relative contact velocity.

A contact kinetic energy E_c can be defined by

$$E_c \equiv \tfrac{1}{2}V_f^T MV_f = \tfrac{1}{2}(P - P_{II})^T M^{-1}(P - P_{II}). \tag{5}$$

The contact kinetic energy E_c is not a simple sum of energies of the separate bodies. However, changes in E_c due to contact forces are the same as changes in the system kinetic energy. E_c is a natural quantity to consider in collisional energy calculations since it is frame-invariant, unlike the total kinetic energy which is not.

Constant E_c defines a surface which facilitates the geometric description of many collision features (Chatterjee, 1997). Each such E_c surface is ellipsoidal with orientation coinciding with the principal directions of M, and corresponding diameters in inverse proportion to the square roots of the eigenvalues of M. The size and position of the E_c surfaces depends on initial conditions. The center C of each E_c ellipsoid is at P_{II}. An impulse of $P = P_{II}$ leads to zero final E_c and hence the maximum possible collisional dissipation. An impulse of $P = 2P_{II}$ conserves energy and leads to a direct reversal of the contact velocity ($V_f = -V_i$). For a given impulse P the associated contact velocity V is normal to the E_c surface on which P lies. The E_c surface associated with the original motion (or precollision energy) passes through the origin with a normal in the direction of V_i. A collision with no final tangential velocity ($V_f \| n$) and a given energy corresponds to one where $P^T n$ is maximum on the corresponding E_c surface. The set of all collisions with $V_f \| n$ makes up the *line of sticking* (Goldsmith, 1960; Wang and Mason, 1992) which passes through P_{II} and passes through the E_c surfaces where their normals are parallel to n.

For given E_c the set of impulses with no normal separation velocity ($V_f^T n = 0$) is the ellipse on the ellipsoidal E_c surface where the surface normal is perpendicular to n (places where the surface tangent planes contain n). Over the full range of possible values of E_c this makes up the *plane of maximum compression*[3] (see, e.g., Routh (1987), Goldsmith (1960), or Wang and Mason (1992)) given by

$$n^T(M^{-1}P + V_i) = 0,$$

which cuts the E_c surfaces in half and includes both P_{II} (at C) and P_I at A. A frictionless collision with kinematic restitution coefficient $e = 1$ has the impulse $2P_I$ and conserves energy. The energy conserving frictionless collision $2P_I$ kills (using P_I) and negates (using P_I again) the normal approach velocity. Thus, A bisects OB.

The dashed line (or plane) in Fig. 1, drawn through point D and parallel to the plane of maximum compression, marks impulses for which the observed kinematic or Newtonian normal restitution e will be some constant. A specific value of e constrains the impulse transmitted to lie on a plane (a line in two dimensions) parallel to the plane of maximum compression (for which $e = 0$). In Fig. 1 point D is on an $e < 1$ plane but lies outside the initial energy surface and thus corresponds to a net increase in kinetic energy. This construction provides an indication of energy conservation problems with naive algebraic collision laws, such as in Whittaker (1944) or Kane and Levinson (1985).

An identically shaped and oriented family of ellipsoidal surfaces can be drawn in contact force space as level surfaces of $F^T M^{-1} F$. Assuming "force-response rigidity" (Chatterjee and Ruina, 1998a) we can use $F = MA$ (the differential form of Eq. (2)). F and A will not be collinear unless they happen to be along an eigenvector of M, which may not be along the normal or tangential direction. The possible lack of alignment of the M eigenvectors with the contact normal and tangent plane is one source of difficulty in collision modeling since it corresponds to inertial coupling between normal and tangential motions and between tangential motions. This misalignment manifests itself by E_c surfaces being crooked with respect to the normal and contact tangent plane.

For every M and precollision velocity V_i, there is an E_c ellipsoidal surface in impulse space. Conversely, every ellipsoidal surface transversely intersecting the n or 1-axis at the origin corresponds to the collision of some pair of bodies. Thus, for impulse analysis, consideration of arbitrary rigid-body mechanisms is equivalent to considering arbitrary ellipsoidal surfaces as described above.

3.2 Reasonable Restrictions on the Contact Impulse. Here are some natural and common assumptions about collisions:

1 Non-negative dissipation of kinetic energy constrains the impulse to lie on or inside the ellipsoid of the initial E_c surface, so P must satisfy (see, e.g., Smith (1991))

$$(P - P_{II})^T M^{-1}(P - P_{II}) \le V_i^T MV_i. \tag{6}$$

2 Noninterpenetration in the final collision velocity ($V^T n \ge 0$) constrains the impulse to lie "above" (in the normal direction) or on the plane of maximum compression (line of maximum compression, in two dimensions).

3 A non-negative normal impulse (no net tensional impulse) must lie in the upper half-space (upper half-plane in two dimensions): $P^T n \ge 0$.

4 If the contact force obeys the Coulomb friction inequality at all instants of time during the collision, then so must the net contact impulse. Thus P must lie in or on the cone given by $\|P - (n^T P)n\| \le \mu n^T P$. In two-dimensional collisions, this friction cone reduces to a pair of "friction lines."

For positive μ, condition (3) above follows from condition (4).

Conditions (1), (3), and (4) are not fundamental. Mechanical energy could be released by contact chemical reactions or phase transitions in the bodies. There is generally some (usually negligible) contact tension in part of the collision. Coulomb's law is only an approximation. But these restrictions might be reasonable limits for approximate general purpose collision

[3] Routh used the term "greatest compression" to denote the point *during* the collision when, as per his incremental model, the instantaneous normal component of the relative velocity at the contact point becomes zero. That point depends on the path in impulse space followed by the collision. If, say from experiments, only the *net* collision outcome is known, the point of maximum compression cannot be determined. However, the set of all impulses P for which $V_f^T n = 0$ is well defined, and called the plane of maximum compression.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

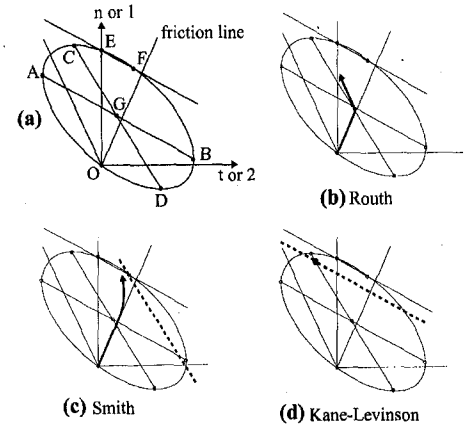


Fig. 3 Sketching regions in two-dimensional impulse space accessible to various collision models

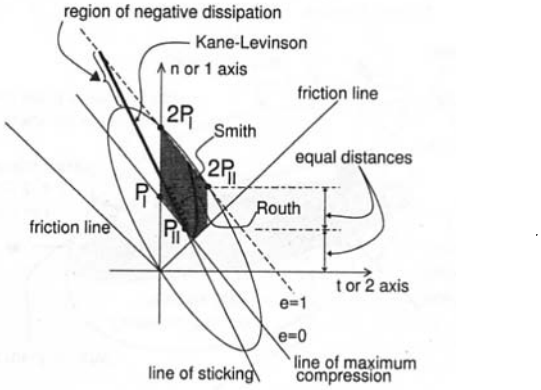


Fig. 4 A generic two-dimensional collision. Shaded region is accessible to new model.

laws. They restrict allowed impulses to a closed, bounded, and convex *permissible* region in impulse space. This *permissible region* is the intersection of half an ellipsoid with a solid cone; for two-dimensional collisions, it is a two-dimensional slice of the corresponding three-dimensional region.

In the absence of additional information about the contact mechanism,[4] every point inside the permissible region represents a plausible collisional outcome. Although we do not expect every permissible impulse to be witnessed in experiments, we know no basic theoretical reason to further restrict the permissible region, nor do we know of sufficiently general experimental data to restrict it on empirical grounds.

All of the collision laws we discuss in this paper use dimensionless collision parameters such as the coefficient of normal restitution. In calculations with any *fixed* parameter values, the collision impulse is then automatically a homogeneous function of degree one in the precollision velocity. P is not simply a linear function of V_i. But, say, doubling V_i while keeping its direction fixed also doubles P while keeping its direction fixed. This velocity scaling is not fundamental, is not obeyed by some real collisions (e.g., spheres; see Goldsmith, 1960), but is apparently obeyed by some other collisions (e.g., vibration dominated interactions; see Stoianovici and Hurmuzlu, 1996).

3.3 Two-Dimensional Geometric Constructions. We now describe how the regions accessible to the different collision laws are represented graphically for two-dimensional collisions. See Fig. 3(*a*).

Pick a normal or 1-direction; and, perpendicular to it, the tangential or 2-direction.

Draw an ellipse whose shape represents a mass matrix M and size represents an initial contact kinetic energy $V_i^T M V_i/2$ (eigenvectors of M are aligned with the principal axes of the ellipse and the ratio of the eigenvalues of M is the square of the aspect ratio of the ellipse). Ellipses of any aspect ratio or orientation are allowed.

On the ellipse, locate the two points where the tangents are vertical (points A and B in the figure) and join these points by a straight line (the line of maximum compression). Locate the points where the tangents are horizontal (points C and D in the figure) and join them by a straight line (this is the line of sticking). The final velocity has a left or right component depending on whether the collision impulse is to the left or right of the line CD.

Pick the impulse origin point O to be on the arc ADB such that the normal to the ellipse is in the direction of the precollision contact velocity. Any point on arc ADB is allowed. If, by convention, the 2-direction is chosen to oppose V_i, then only the points on the portion AD are allowed.

Points A and B correspond to grazing incidence, while point D corresponds to normal incidence. The magnitude of V_i is not determined by the construction.

The intersection of the normal axis with the ellipse at E (or $2P_I$) represents a perfectly elastic, frictionless collision.

Draw friction lines from point O, at angles $\pm\tan^{-1}\mu$ from the normal direction.

The permissible region is now visible as the region between the friction lines, above line AB, and inside the ellipse.

Through point E, draw a line EF parallel to AB. This line represents kinematic restitution $e = 1$. All laws using kinematic restitution $0 \le e \le 1$ must predict an impulse somewhere between the lines AB and EF (possibly extended in both directions).

3.4 Special Situations. For the purposes of comparing collision laws we consider various special collision situations. For these collisions, the curves accessible in impulse space to some well-known one-parameter collision laws, as well as the two-dimensional regions accessible to the new collision law, are shown in Figs. 4 through 8. Details of these constructions will be discussed later in the paper. Below, we briefly comment on aspects relevant to the impulse-space description of this section.

A Generic Collision. The fairly generic collision of Fig. 4 approximately duplicates Fig. 1 but without the definitions of lines and points. Special cases of this collision appear below.

A Collision With Diagonal M. In contrast to the generic case of Figs. 1 and 4, consider a collision where the mass matrix is diagonal, as is the case for collisions of spheres and disks

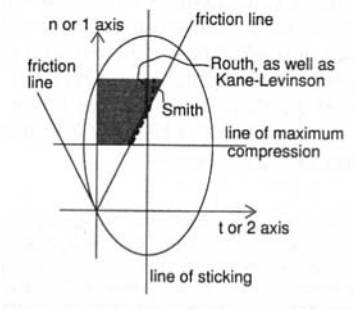


Fig. 5 A collision with diagonal *M* (as for spheres and disks). Shaded region is accessible to new model.

[4] An example of such special information would be if the collision is assumed to respect any special symmetry of M and V_i. The collisions of spheres with no spin component about the n-axis would then be restricted to the n, V_i-plane, a subset of the permissible region. Laws that do not explicitly add a symmetry breaking feature will automatically respect this symmetry, however. Examples include all of the algebraic laws discussed in this paper.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

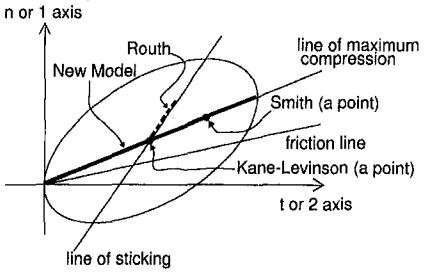


Fig. 6 A tangential or grazing collision. Thick line is accessible to new model (as shown). Note, Routh's method predicts finite normal rebound, but no tangential rebound. New model predicts no normal rebound, but permits tangential rebound.

(see Fig. 5). The lines of maximum compression and sticking are aligned with the coordinate axes.

A Tangential Collision ($n^T V_i \to 0^-$). For grazing or tangential incidence (as shown in Fig. 6), the normal approach velocity is zero, and the energy ellipse is locally tangent to the 1-axis at the origin. If the line of maximum compression slopes appropriately and the friction coefficient is sufficiently high, a collision can occur. Laws based on kinematic restitution do not predict/capture finite normal rebound for grazing collisions.

A Collision With Infinite Friction ($\mu \to \infty$). For collisions where the coefficient of friction is effectively infinite (see Fig. 7), the friction lines are lined up with the tangential axis. Note that for the case shown in the figure, the center of the ellipse, point C, lies *below* the tangential axis. A perfectly dead, no-slip no-bounce collision cannot occur in the case shown since it would require a net negative normal impulse component. Thus, $e = 0$, $\mu = \infty$ does not necessarily mean a dead collision, as some authors have assumed (e.g., Goldsmith, 1960).

A Collision With Unbounded M. Some specially constrained problems have mass matrices M which have one or more eigenvalues equal to infinity as for a pendulum striking a floor (see Fig. 8(a)). This problem has been examined in three dimensions as a special "ideally" constrained (effectively two-dimensional) problem, using Routh's model (e.g., Stronge, 1994). This problem is treated here as the limit of a sequence of problems where the eigenvalue in the direction along the pendulum rod approaches infinity (say with an arbitrarily large point mass replacing the hinge). As the eigenvalue approaches infinity, the initial energy ellipse stretches in the corresponding direction. The portion of it contained inside a suitable finite-sized region approaches the parallel lines shown in Fig. 8(b) (the center of the ellipse is at point C).

3.5 Energy Dissipation and $e \leq 1$. For all single-point *frictionless* rigid-body collisions, it is well known that for normal restitution $e \leq 1$, non-negative dissipation of kinetic energy occurs.

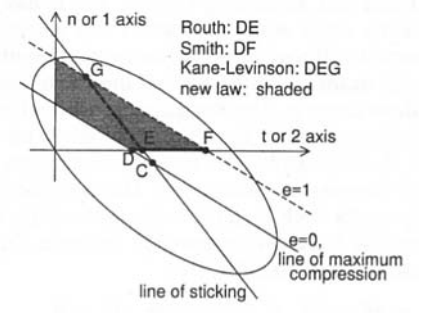


Fig. 7 A collision with infinite friction ($\mu \to \infty$). Shaded region is accessible to new model. Note that if the center of the ellipse, C, were to lie on the t or 2-axis, then Routh's model would predict only a perfectly plastic, sticking collision.

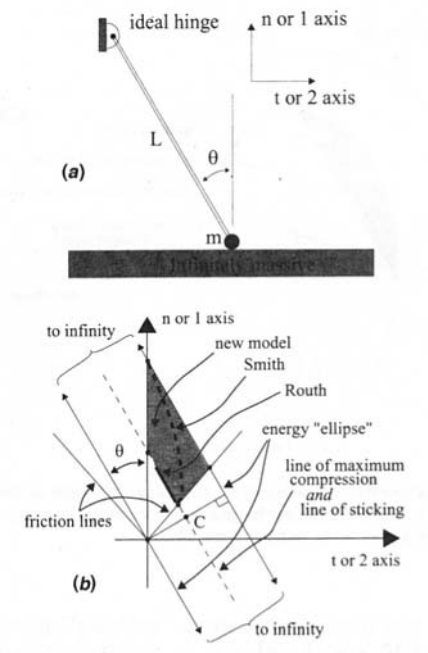


Fig. 8 A two-dimensional pendulum strikes a wall. Modeled by letting one eigenvalue of M approach ∞. Schematic diagram shows a finite portion of energy ellipsoid. Shaded region is accessible to new model. Kane and Levinson's (Whittaker's) model predicts unbounded gain in kinetic energy, does not converge to the kinematically constrained problem, and is not shown.

However, for frictional collisions, the *kinematic* coefficient of normal restitution can be substantially larger than one even in dissipative collisions. Smith and Liu (1992) have shown this using finite element method simulations, and also observed this experimentally, for collisions where M is not diagonal.

M is diagonal for collisions of spheres and disks; such collisions are called "central collisions." Special symmetry-preserving contact deformation behavior can put further limitations on the energy dissipated in such a central collision. Typical experimental data for spheres and disks, too, shows restitution coefficients less than unity. Perhaps due to a combination of these reasons, one occasionally encounters the mistaken opinion that e must be less than unity for central collisions purely from nonnegative energy dissipation considerations. The impulse-space description presented above allows us to briefly address this issue.

Even for diagonal M, energy-dissipating collisions with $e > 1$ can occur, in principle. Figure 9 shows a nearly grazing central collision (diagonal M), with reasonable friction, where the accessible region (gray) is significantly larger than the region allowed by $e \leq 1$ (black). In the figure, the maximum possible value for e is about 17 (it could be made larger). That such collisions are not often observed in practice is due to details of body-specific deformation mechanisms that are more restricted than the general cases allowed in this paper.

Smith and Liu (1992) discuss three well-known definitions of coefficients of normal restitution. Note that the *kinetic* coefficient of restitution e_k is the same as the kinematic coefficient, for diagonal M (central collisions). Thus, $e_k > 1$ does not preclude positive energy dissipation either. The same applies to the *energetic* restitution coefficient, defined in terms of the work done by the normal component of the contact force.

Finally it is also clear that even for collisions with diagonal M, for sufficiently high friction coefficient, it is possible to pick an impulse that corresponds to $e < 1$, lies inside the friction cone, and yet outside the maximum E_c ellipsoid. Typical collision laws do not usually pick such points for central collisions, but some naive laws (e.g., Kane and Levison's law, or Whitta-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

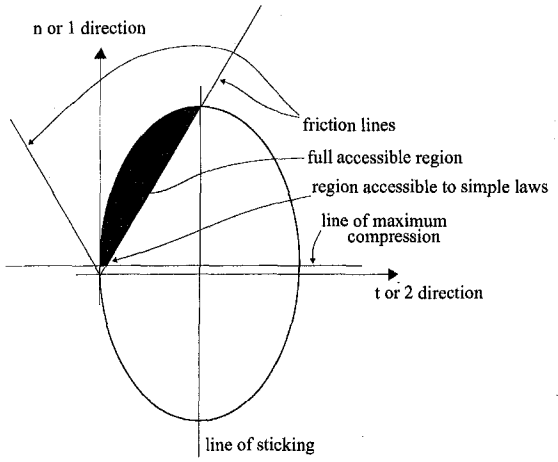


Fig. 9 For a central collision (i.e., diagonal *M*) with sufficiently high friction and sufficiently close to grazing incidence, energy-dissipating collisions can occur with large values of *e*

ker's law) can make such energy "creating" predictions for nondiagonal M, $e < 1$ and $\mu > 0$, as is pointed out in Kane and Levinson's text as well as in appropriate places in this paper.

Thus, for frictional collisions, there is no simple relation between nonnegative energy dissipation and the inequality $e \leq 1$.

4 Some Collision Laws

We now place the impulses predicted by several collision laws in (and out of) the permissible region and also discuss other strengths and weaknesses of these laws.

4.1 Collision Laws That Reach the Entire Permissible Region. One candidate class of collision laws are laws that parameterize the entire permissible region defined by M, V_i, and μ, and no points outside. Here are three ways that this can be done, for example:

1 Arbitrary Parameterization of a Convex Region. Pick a point P_0 inside the permissible region, say its centroid. Then, since the region is convex, all points can be defined by their spherical angles from P_0, say $0 \leq \psi_1 \leq \pi$, $0 \leq \psi_2 \leq 2\pi$, and the fraction of the distance from P_0 to the boundary, say $0 \leq \rho \leq 1$. Thus the outcome of all accessible collisions can be described with the triple (ψ_1, ψ_2, ρ), and no choice of these parameters (within their given bounds) picks a nonpermissible point.

2 Projected Impulses. Parameterize the outcome of the collision by the impulse itself, i.e., let the impulse components be the "collision parameters." If this impulse is outside the permissible region, project it back to the surface along lines leading to an arbitrary point P_0 inside the region (say the centroid). In other words, if P is outside the permissible region, diminish the magnitude of $P - P_0$ until the region boundary is reached (the energy surface, the friction cone, or the plane of maximum compression).

3 Brach's Approach. Consider the following three-parameter law proposed by Brach (1991): a Newtonian or kinematic restitution parameter e for the normal direction, and two *impulse ratio* parameters μ_2 and μ_3 (not to be confused with the coefficient of friction) for the two tangential directions. If all values of $\mu_{2,3}$ were allowed while e was restricted to lie between 0 and 1, then this law would access an infinite slab of finite thickness in impulse space, parallel to the plane of maximum compression. The parameters $\mu_{2,3}$ are understood to be limited within bounds that respect physical considerations. The bounds on these collision parameters are expressed in terms of (configuration-dependent, implicit, nonlinear) inequality constraints expressing non-negative energy dissipation, in addition to friction limits. Brach's laws can capture all permissible impulses (if we allow $e > 1$).

The laws above that access the full permissible region have the merit of being capable of describing all permissible collisions. But we do not know any reason to expect that any of the parameters in these descriptions might be constant over a useful range of collisions (although Brach's proposed impulse-ratio parameter often stays roughly constant for some range of collisions of spheres and disks, if not too wide a range of initial approach angles is considered). The parameters in the above descriptions also do not have any special intuitive or theoretical interpretation.

The utility of a collision law comes not just from its ability to describe a posteriori any given collision once the outcome is known from other sources, but also its ability to predict the outcome of collisions for a range of initial velocities and orientations based either on much collisional data or on separately measurable parameters. The above laws seem lacking in this regard.

Even if not separately measurable and constant, a collision parameter such as a normal restitution coefficient retains some physical interpretation for arbitrary collisions: higher e roughly means that the objects will bounce more. The parameters in the first two laws given above, as well as Brach's impulse-ratio parameters, seem to lack this kind of simple interpretability.

As yet, no collision law we have seen accesses a large three-dimensional subset of the permissible region in impulse space; has collision parameters that have simple physical interpretations; and has collision parameters that are at least roughly constant over some range of some collisions of some bodies for which some data is available.

4.2 Some Algebraic Collision Laws. The permissible region is three-dimensional for three-dimensional collisions and two-dimensional for two-dimensional collisions. Any collision law that has only *one* free parameter (in addition to a friction inequality using μ) can access a one-dimensional subset (curve) of the three-dimensional permissible region. Similarly, any law with two independent free parameters (in addition to a friction inequality using μ) can access a two-dimensional subset of the permissible region. Currently available algebraic laws, except for the full parameterizations of the permissible region as described above, access only a one or two-dimensional region.

We now discuss some popular algebraic collision laws. For illustration we assume that initial motion is to the left so the origin O is on arc AD. The impulses accessible to the laws below are shown in Figs. 4 through 8 for a variety of collisional situations.

4.2.1 Routh's Law. Routh's (1897) law was repopularized recently for two-dimensional collisions by Wang and Mason (1992) and examined in some detail in a three-dimensional context by Bhatt and Koechling (1995a, b). It has also been studied by many other authors. Routh's law assumes infinite tangential contact stiffness, any combination of possibly nonlinear springs and dashpots in the normal direction, and standard friction relations using μ. The neglect of tangential compliance allows expression of the incremental relations in terms of normal impulse P_n instead of time, totally independent of details of the normal compliance mechanism. Denoting tangential and normal components with t and n subscripts, respectively, the tangential impulse P_t (a two-component vector in three dimensions) is incrementally calculated by

$$\text{Either} \quad dP = M \cdot \{dV_n, 0, 0\}^T \quad \text{with} \quad V_t = 0 \quad \text{and} \quad \|dP_t\| \leq \mu dP_n \quad \text{(no slip)}$$

$$\text{or} \quad dP_t = -\mu(V_t/\|V_t\|)dP_n \quad \text{(slip or impending slip)} \qquad (7)$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

where Eqs. (1) and (2) apply to impulse increments throughout (and in cases of impending slip, the unit vector $V_t/\|V_t\|$ is understood to be the direction of impending slip). Routh's law is perhaps *the* classical incremental law. It may be regarded as algebraic in the sense that ODE solving (for given initial conditions) is function evaluation. Also, in two dimensions (where V_t has only one component) the governing differential equations can be solved in closed form. Routh's *kinetic* termination condition was

P_n at end of collision

$$= (1 + e_k)\, P_n \text{ at point of maximum compression}$$

after Poisson (1833).[5] We loosely refer to all laws that use Routh's incremental contact relations Eq. (7), with any termination criterion, as "Routh's law."

Routh's law, in the absence of jamb-associated difficulties (Batlle and Cardona, 1998) can access a one-dimensional curve that always lies in the permissible region for any $0 \le e_k \le 1$ and $\mu > 0$ (see Ivanov, 1992 for a proof of non-negative energy dissipation in three dimensions). It has the ability to predict finite *normal* rebound for tangential or grazing collisions (see, e.g., Wang and Mason, 1992) even for $e_k \le 1$, unlike algebraic laws that use Newtonian or kinematic restitution, which predict zero normal rebound for finite restitution e. Routh's restrictive assumption of no tangential compliance naturally leads to inaccurate predictions as shown in Maw et al. (1981) and Smith and Liu (1992). In particular, Routh's law cannot predict nonzero *tangential* rebound for tangential or grazing collisions nor the slip reversal of a sphere seen in essentially all experiments for at least a small range of nearly normal collisions. The latter effect is easily visible with rubber "super" balls dropped while spinning about a horizontal axis (see, e.g., Johnson (1983) or Garwin (1969)).

See Fig. 3(b). To graph Routh's accessible region for two-dimensional collisions (see Wang and Mason, 1992) fix μ and track P as the normal impulse component P_N increases from 0 (at the start, P is at O, outside the parameter range and outside the permissible region). As P_N increases P moves up along the right friction line until it reaches the line of sticking. At the line of sticking the path has a kink. If the line of sticking makes a smaller angle with the normal than the friction lines do, then sticking is possible and the representative point climbs up along the line of sticking. Otherwise, it climbs up along a line parallel to the left friction line (as shown in the figure).

Depending on the aspect ratio and orientation of the ellipse, the angle of incidence (i.e., location of point O) and the friction coefficient, the intersection with the line of sticking might occur above the line of maximum compression, or even not at all.

The termination criterion, if we use Poisson or kinetic restitution, is that the normal impulse should be $(1 + e_k)$ times the value it had when the representative point reached the line of maximum compression. That point of termination is taken to be the collision impulse.

For grazing collisions Routh's law is the only one discussed in this paper that can predict a collision with nonzero post-collision separation. That is, only Routh's law predicts impulses above the plane of maximum compression for grazing collisions (see Fig. 6).

For infinite friction collisions (see Fig. 7) if the center of the initial energy ellipse happens to lie exactly on the tangential axis, then the region accessible to Routh's law shrinks to a point.

For a collision with unbounded M aligned as in Fig. 8, with large enough friction, Routh's law predicts only a dead collision (no-slip, no-bounce, $P = P_{II}$).

[5] Some ideas of energetic "consistency," and some ambiguities caused by three-dimensional collisions with "jamb," have led to the definition of newer work-based definitions of restitution (see Batlle and Cardona (1998) for discussion and references).

4.2.2 Whittaker's Law. This classical law (see Whittaker, 1944) is stated more clearly in a three-dimensional context in Kane and Levinson (1985), to whom it is often attributed. It uses a coefficient of normal restitution e (Newtonian or *kinematic* restitution) in addition to μ, and therefore accesses a one-dimensional region in impulse space. By this law, either the collision terminates at a condition of sticking, or the impulse is on the friction cone and opposes the post-collision tangential velocity V_{ft}. That is

$$V_f^T n = -(1 + e) V_i^T n \text{ and}$$

$$\text{either } V_{ft} = 0 \quad \text{with} \quad \|P_t\| \le \mu P^T n \quad \text{or} \quad P_t = -\mu \frac{V_{ft}}{\|V_{ft}\|} P^T n.$$

This law requires the solution of nonlinear algebraic equations in three dimensions, cannot predict spin reversal of spheres, cannot predict finite rebound for grazing collisions, and can predict arbitrarily large increases in kinetic energy in a collision for $e \le 1$ and $\mu > 0$ (see Fig. 8).

Whittaker's law is simplest to view geometrically by starting from the coefficient of kinematic restitution e, which specifies one line the collision terminates on (shown dashed in Fig. 3(d)). The intersection of the dashed line with the line of sticking is first located. This point is marked with a black circle in the figure. If the point is between the friction lines, then it is the collision impulse. If it is to the right of the right friction line, then the collision impulse is taken to be the intersection of the right friction line with the dashed line. Conversely, if the point is to the left of the left friction line, then the collision impulse is taken to be the intersection of the left friction line with the dashed line.

As can be seen in Figs. 3 or 4, the intersection of the line of sticking and the $e = 1$ line can easily be outside the initial energy ellipse. This is a direct demonstration of Kane's well known collision energy "paradox." This possibility of generating an increase in system kinetic energy persists for $e < 1$ and reasonable μ so long as the initial E_c surface is skewed as in the figures. The situation is extreme in the context of unbounded M as in Fig. 8 where this law fails completely in that it predicts an unbounded increase of kinetic energy and does not respect the hinge constraint in the limit as an increasing point mass at the bar's end behaves like a hinge.

For diagonal M the Whittaker-Kane-Levinson law is the same as Routh's law (see Fig. 5).

For grazing collisions this law can only predict an impulse of $P = P_{II}$ unless P_{II} is outside the friction cone in which case it predicts zero impulse.

4.2.3 Smith's Law. This law (Smith, 1991) also has one parameter (kinematic restitution e) in addition to μ. Smith's law expresses the tangential impulse direction using a weighted average of pre and post-collision tangential relative velocities. This average is an acknowledgment that the tangential impulse comes from frictional sliding in a variety of directions. Smith's law accesses a one-dimensional curve in the permissible region, cannot predict increases in kinetic energy, can capture spin reversal of spheres, but cannot capture nonreversal of spin for positive μ. It cannot predict finite normal rebound for grazing or tangential collisions, and predicts a fixed amount of tangential rebound for such collisions (the amount of rebound is dependent on μ). Nonlinear algebraic equations need to be solved (solutions exist (Chatterjee, 1997), but uniqueness is unproven at this time).

In Smith's law, the equations describing the three-dimensional collisional interaction are

$$P_n \begin{Bmatrix} 1 \\ -\mu \dfrac{V_{it}\|V_{it}\| + V_{ft}\|V_{ft}\|}{\|V_{it}\|^2 + \|V_{ft}\|^2} \end{Bmatrix} = M \begin{Bmatrix} -(1 + e)V_{in} \\ V_{ft} - V_{it} \end{Bmatrix}, \quad (8)$$

where n and t subscripts denote normal and tangential compo-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

nents. This nonlinear system of equations needs to be solved for P_n and V_{ft}. One or more solutions with $P_n < 0$ frequently exist for three-dimensional collisions with large enough values of μ, and these nonphysical solutions must be rejected.

See Fig. 3(*c*). We can graph the accessible region for Smith's law by parameterizing it with e starting at $e = -1$ (outside the accessible region and parameter range). For e close to -1, the impulse is small, $V_{ft}(e) < 0$, and the impulse ratio is μ. Thus, we draw a curve that climbs from the origin, up the right friction line.

As e increases, P eventually reaches the line of sticking. Now $V_{ft} = 0$. As e increases further, V_{ft} becomes positive, and so the impulse ratio decreases. However, from Eq. (8) it is easy to see that for small V_{ft} the change in the impulse ratio is *quadratic* in V_{ft}. Thus, there is no sudden change in direction (i.e., no kink) and the path of the representative point curves away from the right friction line.

How far and how fast it curves away can be seen from the following. In Fig. 3(*c*), the heavy dashed line is drawn through point F (see Fig. 3(*a*)) and parallel to CD. Just as line EF represents impulses where the normal component of relative velocity is exactly reversed, this dashed line represents impulses such that the *tangential* velocity is exactly reversed. From Eq. (8), the impulse ratio is exactly zero when $V_{fT} = -V_{iT}$, i.e., the tangential relative velocity is exactly reversed. Therefore the curving path of the representative point reaches the normal axis at the point where it intersects the dashed line. Since, in the figure, that intersection occurs beyond point E, it corresponds to $e > 1$ and is not shown.

The collision terminates when the appropriate value of e is reached, which corresponds to reaching a line parallel to AB, somewhere between AB ($e = 0$) and EF ($e = 1$).

For grazing collisions Smith's law can predict only one impulse independent of e (dependent on μ) unless P_{II} is outside the friction cone, in which case Smith's law predicts a zero impulse.

4.2.4 Pfeiffer and Glocker's Law, and Extensions. This three-parameter law (Pfeiffer and Glocker, 1996) is an algebraic law somewhat motivated by Routh's law, in that it assumes that the collisional interaction may be divided into well-defined and distinct compression and expansion phases (see, however, footnote 3). It uses kinetic restitution e_k for the normal direction, and models tangential restitution effects using two other parameters ν and e_t. For single-point contact the law accesses a two-dimensional subset of the permissible region. Unfortunately the law is restricted to two-dimensional collisions (in its 1996 form) and does not seem to generalize to a simple three-dimensional form with tangential restitution. If tangential restitution is not incorporated, this law may be generalized to a single-parameter three-dimensional version (Anitescu and Potra, 1997).

The single-point versions with no tangential restitution, of these laws (Pfeiffer and Glocker, or Anitescu and Potra), are in some ways a combination of Routh's law and the Whittaker-Kane-Levinson law, avoiding the principal difficulties of both. By doing two separate algebraic calculations for compression and expansion, Routh's integration of nonlinear ODE's is avoided. Meanwhile, at the end of each stage, the friction condition assumed is that either there is no slip or the tangential impulse opposes final slip; this extension of the Whittaker-Kane-Levinson law to two stages ensures non-negative energy dissipation. Without tangential restitution, these laws just have one collision parameter (normal restitution) and access a one-dimensional subset of the permissible region.

The primary motivation behind the design of these laws is to pose simultaneous multiple collision problems as tractable linear complementarity problems whose solutions provide plausible collision impulses. However, as mentioned in Subsection 6.2 and discussed in Chatterjee (1998), the linear complementarity conditions are not fundamental properties of rigid bodies during collisions, and so there is no reason to prefer the predictions of these laws over those of other, possibly simpler laws.

4.2.5 A Bilinear Law for Spheres and Disks. Based on experiments, as well as an approximation of results obtained using the theoretical analysis of Mindlin and Deresiewicz (1951), this algebraic law is only applicable to the collisions of objects like spheres and disks where M has eigenvectors aligned with the coordinate axis in our preferred coordinate system. In other words, the law applies only to collisions where M is diagonal in our coordinate system.

The law is discussed in Brach (1991) and has been used by various researchers (see, e.g., Walton, 1992; or Jenkins, 1992). First a normal impulse is calculated using a kinematic restitution coefficient (the same as kinetic restitution, for diagonal M). Then one first assumes that "sticking" occurs, and a candidate tangential impulse is calculated based on the tangential restitution parameter (in this description, the case of "sticking" includes tangential velocity reversal). If the candidate impulse violates the friction inequality, then "slipping" is assumed to occur and the tangential impulse is set equal to μ times the normal impulse. One expression of this law in our chosen coordinates and using kinematic restitution is

$$P_n = -(1 + e)M_{11}V_{in} \quad (\text{recall, } V_{in} < 0)$$

$$P_t = -\frac{V_{it}}{\|V_{it}\|} \cdot \min\{\mu P_n, (1 + e_t)M_{22}\|V_{it}\|\}. \tag{9}$$

This law is consistent with experiments in the following ways. For nearly normal collisions and positive e_t the tangential velocity is reversed (at least in part). Tangential velocity is not reversed for collisions closer to grazing incidence.

As emphasized above, this bilinear law is meant for diagonal M, the case shown in Fig. 5. For this case, this bilinear law is identical to the new law described below, and covers the shaded region shown in the figure.

Note that all the laws described above become identical for frictionless ($\mu = 0$) collisions, irrespective of whether M is diagonal or not.

5 A New Algebraic Collision Law

All of the algebraic laws of which we are aware, such as the laws described above, lack one or more of the following features: the law is applicable to three dimensions, nonreversal of spin of bouncing balls can be predicted, reversal of spin of bouncing balls can be predicted, predictions are in the permissible region, at least a two-dimensional subset of the permissible region is covered, calculation is simple and has unique solutions, and the parameters are easy to interpret.

Towards this end three new collision laws were constructed in Chatterjee (1997) and briefly described in Chatterjee and Ruina (1997). These new laws have all the features described in the paragraph above. The new laws are generalizations to nondiagonal M of the bilinear law of Eq. (9). We present the simplest of these laws here.

The difficulties in generalizing Eq. (9) to arbitrary three-dimensional collisions are that (1) the direction of the frictional impulse needs to be chosen (since there are two independent tangential directions); and (2) for fixed normal restitution e the magnitude of the normal impulse generally depends on the magnitude and direction of the tangential impulse. That the choice of impulse direction cannot be made in a naive manner is suggested by the possible energy increases predicted by Whittaker's (Kane and Levinson's) law. Finally, a form should preferably be picked which leads to easy calculation.

The new law has two restitution parameters e and e_t. The *normal* restitution coefficient $0 \le e \le 1$ (kinematic or Newtonian) describes the degree to which the normal approach velocity

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

is reversed.[6] The *tangential* restitution coefficient $-1 \le e_t \le 1$ is a measure of the extent to which the initial tangential velocity is reversed. Finally, the standard friction coefficient $0 \le \mu \le \infty$ is used to bound the impulse within the friction cone.

The new collision law expresses the collisional impulse as a weighted sum of a perfectly plastic, frictionless collision (P_I), and a perfectly plastic, sticking (i.e., no-slip, no-bounce) collision (P_{II}). Thus the two-dimensional region accessed by the new law is in a plane that is a somewhat natural choice for general collisions (the same plane was independently proposed by Rubin (1998)).

This simple choice of impulse directions may be looked upon as a gross approximation that avoids the difficulties encountered with the three-dimensional implementation of incremental laws (see e.g., the discussion of Routh's law in Bhatt and Koechling (1995a, b)), taking the view, as explained in the Introduction, that even complicated incremental laws are themselves gross approximations with limited applicability. At the same time, this simple choice provides the twin benefits of a simple extra collision parameter as well as explicit linear equations that are easy to solve and have unique solutions (compare with Smith's Eq. (8), which is nonlinear and can also have some physically unrealistic (spurious) solutions which iterative numerical procedures have to recognize and avoid).

The calculation proceeds in two steps. First a normal impulse is calculated which enforces the kinematic restitution condition. Then an additional impulse parallel to the plane of maximum compression is computed which does not affect the normal restitution. The second impulse is computed to achieve the tangential restitution but can be clipped by the friction inequality.

5.1 "Sticking" Calculation of Candidate Impulse: No Friction Limitations. The candidate impulse, $\hat{P}$, is taken to be

$$\hat{P} := (1 + e)P_I + (1 + e_t)(P_{II} - P_I), \quad \text{where} \quad 0 \le e \le 1, -1 \le e_t \le 1. \tag{10}$$

We prove below that the impulse $\hat{P}$ is guaranteed to satisfy the energy and noninterpenetration criteria (conditions 1 and 2 of Section 3.1). Our proof consists of the following observations.

The set of possible impulses $\hat{P}$, as given by Eq. (10), is a parallelogram with four vertices within or on the boundaries of the convex half-ellipsoidal region permitted by energy and interpenetration conditions.

P_I (for $e = 0$, $e_t = -1$) is a perfectly plastic frictionless collision. $2P_I$ (for $e = 1$, $e_t = -1$) is an energy conserving frictionless collision. $2P_{II}$ (for $e = 1$, $e_t = 1$) corresponds to a collision with perfect rebound, $V_f = -V_i$, hence, no interpenetration occurs and energy is conserved. For the fourth vertex, consider applying the corresponding impulse $2P_{II} - P_I$ (for $e = 0$, $e_t = 1$) as a sequence of two impulses, $2P_{II}$ and $-P_I$. After $2P_{II}$ has acted, the energy is the same as before while the velocity is reversed, becoming $-V_i$. The impulse $-P_I$ corresponds to a perfectly plastic, frictionless collision for that (new) initial condition. Thus the fourth vertex corresponds to a collision terminating on the plane of maximum compression and inside the initial energy surface.

Thus the interior of the parallelogram described by Eq. (10) lies entirely in the region permitted by energy and interpenetration conditions (end of proof).

5.2 "Slipping" Check of Candidate Impulse: Consideration of Friction. The candidate impulse, $\hat{P}$, is of the form $(1 + e)P_I + c(P_{II} - P_I)$, for a positive number $c = 1 + e_t$. If $\hat{P}$ lies inside the friction cone, then all criteria of Subsection 3.2 are satisfied, and we take $\hat{P}$ to be the impulse transmitted in the collision.

Note that $(1 + e)P_I$ corresponds to a frictionless collision and so always lies inside the friction cone. If $\hat{P}$ lies *outside* the friction cone, then there is some constant $\kappa \ge 0$ such that the impulse

$$P^* := (1 + e)P_I + \kappa(P_{II} - P_I) \tag{11}$$

lies exactly *on* the surface of the friction cone. In such cases, we take P^* to be the impulse transmitted in the collision. The calculation is straightforward.

$$\text{If } \|\hat{P} - (n^T\hat{P})n\| > \mu n^T\hat{P} \tag{12}$$

$$\text{then } \kappa = \frac{\mu(1 + e)n^TP_I}{\|P_{II} - (n^TP_{II})n\| - \mu n^T(P_{II} - P_I)} \tag{13}$$

$$\text{else } \kappa = 1 + e_t \tag{14}$$

(note that the check is sufficient even for the problematic cases: (a) when $n^T\hat{P} < 0$, i.e., when $\hat{P}$ has a negative normal component, and (b) when P_{II} and hence $\hat{P}$ is along the normal direction). Finally, obtain P^* from Eq. (11).

Geometrically, the preceding calculation is equivalent to projecting the candidate impulse $\hat{P}$ on to the friction cone along a line in impulse space in the $P_I - P_{II}$ direction. This shows that for frictionless collisions the present law reduces to the other laws reviewed here, with Newtonian restitution (as may also be seen by putting $\mu = 0$ into Eq. (13)). Since the impulse $(1 + e)P_I$ is known to be inside the friction cone, and since the calculation of Eq. (13) is only attempted when $\hat{P}$ lies *outside* the friction cone, existence and uniqueness of solutions are assured from geometrical considerations.

One might also construct collision laws that predict approximately the same impulse as suggested above, but with a smoothened transition between "sticking" (no projection required) and "slipping" (projection required). We skip such considerations for simplicity.

To summarize, the collision calculation consists of Eqs. (3), (4), and (10), followed by Eqs. (12) through (14), and finally Eq. (11). For clarity, the calculations described above are given in MATLAB code in Appendix A2. The two-dimensional region in three-dimensional impulse space that is accessible to the new collision law is shown schematically for a generic collision in Fig. 2. See also Fig. 4.

Since the corresponding calculations for the case of spheres and disks are exactly equivalent to those in the bilinear law referred to earlier, this law *also* reduces to that bilinear law whenever the mass matrix M is diagonal (as for collisions of disks and spheres; see Fig. 5).

For grazing collisions (Fig. 6) the region accessible to the new law shrinks to a line that includes the coverage of the other three-dimensional algebraic laws except for Routh's law. Rightly or wrongly, the new law does not predict the normal rebound of Routh's law for the frictional locking type of situation shown. On the other hand, the new law can predict tangential rebound, which Routh's law cannot.

For collisions with infinite friction (see Fig. 7) the new law covers the other laws considered, perhaps reflecting an ability to deal with a broader range of tangential compliances.

Finally, the new law makes reasonable predictions for a constrained problem with sufficiently high friction, with the constraint modeled as the limiting case of M having a large eigenvalue (see Fig. 8).

5.3 Composite Collision Laws. A simple way to generalize the new law, to (say) allow normal rebound in grazing collisions, is to use it in a weighted sum or interpolation with one or more other laws each of which generates impulses in the permissible region. Since the permissible region in impulse

[6] We have chosen to keep e between 0 and 1 for simplicity. However, this is not a fundamental restriction, and there can be kinematic restitution greater than 1 in some frictional collisions (see Subsection 3.5). If a specific application requires it, e can be allowed to be greater than 1 but assurance of staying in the permissible region is lost and needs to be checked on a case-by-case basis.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

space is convex, the interpolated impulses would also be in the permissible region.

For example, we might introduce an interpolation parameter σ, and define a "newer" collision law as follows:

> Given M, V_i, μ, and e, calculate the collision impulse P_R as predicted by Routh's law, and the impulse P_{new} as predicted by the new law. Using the parameter $0 \leq \sigma \leq 1$, now compute the collision impulse to be $P := \sigma P_{\text{new}} + (1 - \sigma)P_R$.

The extreme of $\sigma = 0$ returns Routh's law, while $\sigma = 1$ makes our Routhless prediction.

At present there does not seem to be enough data to suggest that a good fit with such a combined law would have a usefully greater accuracy than either law alone. But such an interpolation will yield in effect a three-parameter collision law that will be able to capture phenomena such as simultaneous normal *and* tangential rebound in tangential collisions. For general three-dimensional collisions, such a composite law will access a three-dimensional subset of the accessible region.

6 Simultaneous Multiple Collisions

Although simultaneous collisions of independently moving objects seem nongeneric, the collision of one object with another that already is in contact with still other objects is common in practice. And although single-point collisions of seemingly ideal mechanisms fit the formalism described in the paper thus far, they may also be treated as simultaneous collisions. Hence a recent surge of interest in predicting the outcome of simultaneous collisions.

The first step in generalizing the impulse space approach is straightforward. Recall that the impulse-momentum relations for a single-point collision are of the form $P = M(V_f - V_i)$, equivalent to $WP = V_f - V_i$ (where $W = M^{-1}$). It is well known (see, e.g., Pfeiffer and Glocker, 1996) that a similar equation holds for multiple collisions, which we write as

$$WG = U_f - U_i, \tag{15}$$

where the column matrix G contains the components of the impulses P at all k locations, and the column matrices U_f and U_i contain the components of the relative velocities V_f and V_i at all k locations. In what follows, we will loosely call G an impulse, and the U's relative velocities. In general W will be symmetric and positive semidefinite. Often, W will be singular, corresponding to the possibility of self-equilibrating impulse sets (say, equal and opposite impulses along the line joining the two contact points on a body that simultaneously touches two immovable objects). Thus the permissible region will often be infinite in extent unless new restrictive assumptions are added.

6.1 The Ill-Posed Nature of Simultaneous Multiple Collisions. In some sense, a rigid-body collision law is a distinguished limit of the properties of a deformable body. The limit is distinguished in that the limiting behavior depends not just on some single measure of deformation going to zero, but also on ratios of deformation moduli (say tangential to normal compliance). For single-point collisions the relevant ratios may not vary too wildly from one collision to another similar collision, hence the hope of deterministic collision laws with some predictive utility. For multiple impacts, however, the key ratios in the distinguished limit include, say, the ratios of (very small) separation distances of various pairs of contacting points. These ratios can differ radically for macroscopically similar-looking situations.

Most notably, all collision laws give vastly different results if the order of near-simultaneous collisions is changed by minute changes in initial conditions, for typical configurations. But since "simultaneous" is only a coarse-grained description of a complex event that does have structure in time, variously ordered near-simultaneous collisions are all different possibilities for the same "simultaneous" collision. Similarly, the outcome of approximately simultaneous collisions also depends sensitively on the relative compliances of the various contacts and the relative degree of rigidity of the various bodies. These issues have been discussed by Ivanov (1995).

The ill-posed nature of "simultaneous" collisions is not a mathematical fine point that can be removed by any formalism, no matter how elaborate. The ill-posedness of some multiple-point collisions is a fundamental physical property of the collisions, not a consequence of an unsuitable modeling approach. The outcome of a "pool break" will vary from game to game, heedless of the unique outcome predicted by any simultaneous-collision law. The classic example of such a situation is the longitudinal impact of three spheres discussed in detail by Ivanov (1995). Another example is a horizontal rod that falls onto a horizontal, frictionless table, making contact at its two endpoints, studied in detail by Goyal et al. (1998a, 1998b).

The authors of papers on multiple impacts are aware of these difficulties. The primary motivation behind such studies is often to develop a mathematically consistent framework for simulations. For example, Moreau (1983) says,

> ". . . physical situations undoubtedly exist, where this concept gives a reasonably accurate description of reality. But, to the author's opinion, the main interest of the concept lies in the internal mathematical consistency of the resulting evolution problem."

He then adds that an existing, simple and consistent theory may more easily allow more realistic empirical corrections.

We summarize the simultaneous collision problem as follows: Inherently inescapable ill-posedness notwithstanding, simulators will simulate simultaneous collisions and there is a need for simultaneous-collision laws that are as reasonable as possible.

6.2 The Linear Complementary Problem Formulation and Multiple Collisions. Although we do not adopt such an approach here, we briefly mention the linear complementarity problem or LCP (see Cottle et al., 1992) approach to multiple collisions.

For simplicity, consider what a frictionless, plastic collision might mean in a multiple collision context. We cannot require that the normal components of relative velocities at all contact locations be exactly zero at the end of such a collision (else, some of the normal impulses required may be negative).

Several authors adopt the *complementarity conditions* that at each location, either the normal velocity comes to zero and there is a non-negative impulse, or there is zero impulse and the post-collision normal velocity is non-negative (see, e.g., Löstedt, 1982; Baraff, 1989; Pfeiffer and Glocker, 1996; Anitescu and Potra, 1997). For a plastic frictionless multiple collision this assumption guarantees a possibly nonunique but always physically permissible G_I.

LCP-based methods have been extended to deal with frictional multiple collisions with nonzero normal and tangential restitution in two dimensions (Pfeiffer and Glocker, 1996), and without tangential restitution in three dimensions (Anitescu and Potra, 1997).

Note that the realism of such complementarity conditions is not a fundamental property of rigid-body collision dynamics, although similar complementarity conditions are fundamental properties of smooth motions (Chatterjee, 1998). We know of no reason to prefer the *predictions* of an LCP-based law to those of any other law.

6.3 Extension of the New Law to Multiple Collisions. One approach to multiple collisions might be to repeat the approach we have taken to single point collisions. One might construct a permissible region in the higher-dimensional im-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

pulse space of G and then seek a parameterization of a reasonable lower-dimensional subset. Using a linear combination of a perfectly plastic frictionless collision with a perfectly plastic sticking collision (one from the solution set of each would need be selected) might still be a reasonable choice. The details of the projection on to the frictional cone(s) would be more involved. This approach might result in a versatile collision law, depending on the modeling choices made. We have not pursued this, or other similar higher dimensional approaches, thus far.

6.4 Using the New Law in a Sequence of Simple Collisions. In the first three sections of this paper we have described a basis for defining reasonable single-point collision laws and have presented one such law. The classical approach to multiple collisions, as discussed among others by Ivanov (1995) and Baraff (1989), is to treat them as a sequence of collisions.

Well-known drawbacks to this approach are that (a) it involves indeterminacy in the sequence to be used (Ivanov, 1995), and (b) it can lead to an infinite set of collision calculations to complete one macroscopic simultaneous collision (Baraff, 1989). Both of these drawbacks are discussed below.

Let us consider the use of our new law in multiple collisions, broken up into a sequence of single collisions. For k "simultaneous" contacts, we use $2k$ collision parameters and k friction parameters. If frictionless hinges are treated as places of simultaneous collision they need no parameters (they have plastic, sticking collisions). The collection of bodies involved in the simultaneous collision are considered as separate bodies. Whether mechanisms are broken into separate bodies or not is a modeling choice (discussed separately below).

We propose that "simultaneous" collisions should be ordered by normal approach velocity. After r single-point collisions (where $r \geq 0$ starts at 0), we pick the contact location where the *normal* component of relative velocity is the lowest (i.e., we find the contact location with the greatest normal approach velocity). If this turns out to be a non-negative number (i.e., separating or sliding but not interpenetrating), then the collision is over. If it is negative, then we carry out a single-point collision calculation for that location *ignoring* the other locations (i.e., assuming that no impulses occur at the other locations). Once the collision impulse is determined, its effect on all the relative velocities at all the contact locations is found, using Eq. (15). This gives the initial velocities for the next stage of the calculation, and the process is repeated.

Having two potential contact points with the same approach velocity *is nongeneric* and can only occur if there is special symmetry. Such situations include the horizontally falling rod (Goyal et al., 1998a), two identical spheres hitting a third sphere from two sides, with identical velocities (Ivanov, 1995) and a "pool break" with an exactly centered initial shot. In this nongeneric case three options available to the simulator are (1) Change the initial conditions in the simulation (within a reasonable space of initial uncertainty) to break the symmetry, (2) order the "equal-approach-velocity" collisions arbitrarily (say, by taking the dot product of the position with a predefined non-special direction), or (3) following Ivanov's recommendation for such problems randomize the collision order so that the genuine indeterminacy in the experiments would then be represented by indeterminacy in the collision law.

If the process described above terminates in a finite number of single-point calculations, then we will have a prediction that is reasonable, obeys basic restrictions, and has all the merits of the single-point law used.

However, the prescription may *not* terminate in a finite number of collisions.

In such cases, so long as rare energy conserving conditions are not met[7] a single-point collision law dissipates a nonzero fraction of the system's kinetic energy at every collision. Thus, for $e < 1$, if the number of collisions is infinite, the system will *exponentially* (in the number of collisions—macroscopic time is not advanced) approach a state of zero interpenetration velocity at all locations due to the dissipation in each collision. Either by extrapolation or by simple truncation, the end of the collision can still be found in a reasonable time. Since the possibility of infinitely many collisions in finite time is unavoidable even in single-point collisions (as for the classic case of a ball bouncing vertically on the floor, with $e < 1$) series truncation or extrapolation is already a necessary part of general rigid-body simulations with collisions.

For the nonphysical cases where one wants to model energy conserving collisions, it is possible to construct examples where the sequence of collisions continues indefinitely with no energy dissipation (we do not mean infinite sequences of separate collisions, but sequences where the system never changes its configuration and the individual impacts are spaced apart by infinitesimal time intervals). Thus for energy conserving collisions our proposed sequencing does not necessarily generate a prediction.

The extension to multiple impacts that we have described above has, by construction, most of the attractive features that originally motivated the single-point version. In particular,

1 it is algebraic, and for each location has two parameters (other than friction) with approximate but clear physical interpretations.
2 it is simple, in that its algorithmic requirements are limited to basic matrix algebra and linear calculations.
3 existence of solutions is guaranteed whenever $e < 1$, and even for $e = 1$ in all but some nongeneric collision configurations. Except for systems with special symmetry, uniqueness of solutions is guaranteed. The nonuniqueness that remains is in keeping with the physics.
4 the law reduces exactly to the single-point version if there is only one contact.

We mention that simulations using the new law in a sequence of collisions with between five to ten bodies and about 30 "simultaneous" contacts have been conducted by B. Mirtich (personal communication), with promising results.

6.5 Mechanism Constraints as Simultaneous Impacts. Although the colliding pendulum of Fig. 8 fits in the formalism of single-point collisions with one eigenvalue of M approaching infinity, it may also be viewed as a simultaneous collision at both rod ends. Within the sequencing approach we propose for "simultaneous" collisions, this case would be treated as a collision of a free-floating rod with the ground followed by a sticking collision of the other end with the hinge. This in turn may or may not lead to an infinite sequence as described.

For collisions of mechanisms there is then the choice of whether to treat the mechanism as a set of bodies or as an intact mechanism. The intact-mechanism approach leads, at least for single-point collisions, to a single direct collision calculation but is perhaps not likely to be accurate (since the mechanism constraints cannot be expected to be closely obeyed[8] *during* the collisions). On the other hand, the independent body approach frequently leads to an infinite series of calculations (which needs to be truncated or extrapolated).

7 Concluding Comments

The set of reasonably permissible impulses for single-point collisions has been described. The properties and shortcomings of some known collision laws have been described, in part using the impulse space description. A new algebraic collision law

[7] Namely (i) $e = 1$ and $\mu = 0$, (ii) $e = 1$, $e_t = 1$, $n^T P_{II} > 0$ and μ sufficiently large, or (iii) $e = 1$, and MV_I directed exactly along the normal direction (so that P_{II} is normal.)

[8] If the compliance effects in the bearings are comparable to the deformation effects during the collision, the constraint is not ideal.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

has been presented for single-point impacts that is automatically in the "permissible" part of the impulse space, is capable of predicting a fairly wide range of phenomena, and has unique solutions that are easy to calculate even in three dimensions. A generalization to multiple impacts by means of sequenced collisions has been proposed.

We are hopeful that the impulse considerations and new law will facilitate more straightforward rigid-body simulations with impacts.

We conclude with a few natural open questions. What part of the permissible region can be accessed by reasonable incremental (microcontact) laws that involve some tangential compliance? What part of the permissible region is accessed by various experimental situations? Does the proposed sequencing, which renders many simultaneous collision situations determinate, predict outcomes that match those observed in practice?

References

Anitescu, M., and Potra, F. A., 1997, "Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems," *Nonlinear Dynamics,* Vol. 14, pp. 231–247.

Baraff, D., 1989, "Analytical methods for dynamic simulation of non-penetrating rigid bodies," *Computer Graphics,* Vol. 23, No. 3, pp. 223–232.

Batlle, J. A., and Cardona, S., 1998, "The Jamb (Self-Locking) Process in Three-Dimensional Rough Collisions," ASME JOURNAL OF APPLIED MECHANICS, Vol. 65, pp. 417–423.

Bhatt, V., and Koechling, J., 1995a, "Partitioning the Parameter Space According to Different Behaviors During Three-Dimensional Impacts," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 740–746.

Bhatt, V., and Koechling, J., 1995b, "Three-Dimensional Frictional Rigid-Body Impact," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 893–898.

Brach, R. M., 1991, *Mechanical Impact Dynamics: Rigid Body Collisions,* John Wiley and Sons, New York.

Brogliato, B., 1996, *Nonsmooth Impact Mechanics: Models, Dynamics and Control,* Springer, London, UK.

Chatterjee, A., 1997, "Rigid Body Collisions: Some General Considerations, New Collision Laws, and Some Experimental Data," Ph.D. thesis, Cornell University, Ithaca, New York.

Chatterjee, A., 1998, "Note on the realism of complementarity conditions in rigid body collisions," manuscript in preparation.

Chatterjee, A., and Ruina, A., 1997, "Some new algebraic rigid body collision laws," *Applied Mechanics in the Americas,* Vol. 4, Proceedings of the Fifth Pan American Congress of Applied Mechanics, San Juan, Puerto Rico, pp. 403–406.

Chatterjee, A., and Ruina, A., 1998a, "Two Interpretations of Rigidity in Rigid-Body Collisions," ASME JOURNAL OF APPLIED MECHANICS, Vol. 65, pp. 894–900.

Chatterjee, A., and Ruina, A., 1998b, "Realizability of Arbitrary Local Mass Matrices in Single-Point Rigid Body Collisions," *Proceedings of the International Symposium on Impact and Friction of Solids, Structures and Intelligent Machines,* Ottawa, Canada, June, to appear.

Cottle, R. W., Pang, J.-S., and Stone, R. E., 1992, *The Linear Complementarity Problem,* Academic Press, San Diego, CA.

Garwin, R. L., 1969, "Kinematics of an ultraelastic rough ball," *American Journal of Physics,* Vol. 37, No. 1, pp. 88–92.

Goldsmith, W., 1960, *Impact: The theory and physical behavior of colliding solids,* Edward Arnold, Ltd., London, UK.

Goyal, S., Papadopoulos, J. M., and Sullivan, P. A., 1998a, "The Dynamics of Clattering, I: Equations of Motion and Examples," ASME *Journal of Dynamic Systems, Measurement and Control,* Vol. 120, pp. 83–93.

Goyal, S., Papadopoulos, J. M., and Sullivan, P. A., 1998b, "The Dynamics of Clattering, II: Global Results and Shock Protection," ASME *Journal of Dynamic Systems, Measurement and Control,* Vol. 120, pp. 94–102.

Goyal, S., Pinson, E. N., and Sinden, F. W., 1994, "Simulation of dynamics of interacting rigid bodies including friction I: General problem and contact model," *Engineering with Computers,* Vol. 10, pp. 162–174.

Ivanov, A. P., 1992, "Energetics of a collision with friction," *Journal of Applied Mathematics and Mechanics,* Vol. 56, No. 4, pp. 527–534.

Ivanov, A. P., 1995, "On multiple impact," *Journal of Applied Mathematics and Mechanics,* Vol. 59, No. 6, pp. 887–902.

Jenkins, J. T., 1992, "Boundary Conditions for Rapid Granular Flow: Flat, Frictional Walls," ASME JOURNAL OF APPLIED MECHANICS, Vol. 59, pp. 120–127.

Johnson, K. L., 1983, "The bounce of 'superball,'" *International Journal of Mechanical Engineering Education,* Vol. 11, No. 1, pp. 57–63.

Kane, T. R., and Levinson, D. A., 1985, *Dynamics: Theory and Applications,* McGraw-Hill, New York.

Lötstedt, P., 1982, "Mechanical systems of rigid bodies subject to unilateral constraints," *SIAM Journal of Applied Mathematics,* Vol. 42, No. 2, pp. 281–296.

Maw, N., Barber, J. R., and Fawcett, J. N., 1981, "The Role of Elastic Tangential Compliance in Oblique Impact," ASME *Journal of Lubrication Technology,* Vol. 103, pp. 74–80.

Moreau, J. J., 1983, "Standard inelastic shocks and the dynamics of unilateral constraints," *Unilateral problems in structural analysis* (International Centre for Mechanical Sciences, Courses and Lectures, No. 288), G. D. Piero and F. Maceri, eds., Springer-Verlag, New York, pp. 173–221.

Pfeiffer, F., and Glocker, C., 1996, *Dynamics of rigid body systems with unilateral constraints* (Wiley Series in Nonlinear Science), A. H. Nayfeh and A. V. Holden, series eds., John Wiley and Sons, New York.

Poisson, S. D., 1833, *Traité de Mécanique,* 2nd Ed., Bachelier, Paris.

Routh, E. J., 1897, *Dynamics of a System of Rigid Bodies,* 6th Ed., Macmillan and Co., London.

Rubin, M. B., 1998, "Physical Restrictions on the Impulse Acting During Three-Dimensional Impact of Two "Rigid" Bodies," ASME JOURNAL OF APPLIED MECHANICS, Vol. 65, pp. 464–469.

Smith, C. E., 1991, "Predicting Rebounds Using Rigid-Body Dynamics," ASME JOURNAL OF APPLIED MECHANICS, Vol. 58, pp. 754–758.

Smith, C. E., and Liu, P.-P., 1992, "Coefficients of Restitution," ASME JOURNAL OF APPLIED MECHANICS, Vol. 59, pp. 963–969.

Stoianovici, D., and Hurmuzlu, Y., 1996, "A Critical Study of the Applicability of Rigid-Body Collision Theory," ASME JOURNAL OF APPLIED MECHANICS, Vol. 63, pp. 307–316.

Stronge, W. J., 1994, "Swerve During Three-Dimensional Impact of Rough Bodies," ASME JOURNAL OF APPLIED MECHANICS, Vol. 61, pp. 605–611.

Walton, O., 1992, *Particulate Two-Phase Flow,* M. C. Roco, ed., Butterworth-Heinemann, Boston, Chapter 25.

Wang, Y., and Mason, M. T., 1992, "Two-Dimensional Rigid-Body Collisions With Friction," ASME JOURNAL OF APPLIED MECHANICS, Vol. 59, pp. 635–642.

Whittaker, E. T., 1944, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies,* 4th Ed., Dover, New York.

APPENDIX

Details of Some Calculations for Single-Point Impacts

A1 Calculating the Mass Matrix M. The mass matrix M is a property of the configuration of the rigid-body system at impending contact. How M is most easily calculated in practice depends on what one is doing. In the worst case a rigid-body simulator can find the three columns of M^{-1} by doing four noncollisional rigid-body calculations. This method is based on the fact that the relative acceleration between the contact points on the two bodies is the sum of two parts: a part that includes centripetal terms as well as the effects of other applied forces but not the effect of equal and opposite contact forces, and another part that is strictly linear in the equal and opposite contact forces. The calculation described below can be performed in any orthonormal coordinate system.

First, in the reference calculation, the two impending contact points are allowed to move as they would if they were not in contact but all the noncollisional forces in the system were acting. The difference between the acceleration vectors of the two contact points is the reference relative acceleration.

Next, a unit force is applied in the positive x_1-direction, on the contact point of body 2, and in the negative x_1-direction on the contact point of body 1. The two impending contact points are allowed to move as they would if they were not in contact but all the noncollisional forces in the system as well as the new forces in the x_1-direction were acting. The difference between the acceleration vectors of the two contact points is calculated. Subtracting from this the reference relative acceleration, the first column of M^{-1} is obtained.

A similar calculation, but with the equal and opposite forces acting along the positive and negative x_2, and then x_3, directions, gives the second and third columns of M^{-1}. The matrix M^{-1} thus obtained is inverted to give M.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

A2 MATLAB Code for Collision Calculation.

```
function P = collrule( M, V_i, n, mu, e, e_t)
% This matlab function calculates the collision impulse P, as
% predicted by the algebraic model proposed by Chatterjee and Ruina.
% Note: single quote (') means transpose
% M is a 3X3 symmetric positive definite matrix, V_i is 3X1,
% n is 3X1 with n'*n = 1,
% mu >= 0, 0 <= e <= 1, -1 <= e_t <= 1
% Assume that n'*V_i <= 0 (else, no collision occurs)

P_I = - n*n'*V_i /( n'*inv(M)*n );      % perfectly plastic, frictionless:          Eq. 3
P_II = - M*V_i;      % perfectly plastic, sticking:                               Eq. 4
P_hat = (1+e)*P_I + (1+e_t)*(P_II - P_I);      % candidate impulse:               Eq. 10
if norm(P_hat - n*n'*P_hat) > mu*n'*P_hat      % friction check:                  Eq. 12
  kappa = mu*(1+e)*n'*P_I/(norm(P_II - n*n'*P_II) - mu*n'*(P_II - P_I));      %  Eq. 13
else
  kappa = 1 + e_t;      %                                                        Eq. 14
end

P = ( 1+e )*P_I + kappa* (P_II - P_I);      % the collisional impulse:           Eq. 11
```


Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

An Energy-Based Coefficient of Restitution for Planar Impacts of Slender Bars With Massive External Surfaces

Y. Hurmuzlu
Mechanical Engineering Department,
Southern Methodist University,
Dallas, TX 75275
Assoc. Mem. ASME

A new method to solve the collision problems of slender bars with massive external surfaces is developed. The proposed solution accounts for the effect of impact induced vibrations and multiple collisions on the post-collision velocities of the impacting members. The approach is based on representing the vibrational energy of the bars during the collision process in terms of a nondimensional parameter, termed the elastic energy percentile. The elastic energy percentile is expressed as a simple scalar function of the drop angle and a nondimensional parameter, which encapsulates the bar geometry, material, and the stiffness of the contact surface. The elastic energy percentile is then used to develop a new momentum-based solution method. The method relies on a revised energetic coefficient of restitution that resolves the effect of impact induced vibrations on the post-collision velocities of the impacting bars. The assumptions used in the theoretical development and the outcomes predicted by the proposed method were verified by conducting a set of experiments using several bars with varying geometric and material properties.

1 Introduction

The coefficient of restitution remains to be the most controversial, and arguably, the most important constant that is used in the solution of collision problems. Following its inception by Newton (1686), and after undergoing several revisions regarding its definition, widely accepted solutions of impact problems remain to be incomplete without a proper definition of the restitution coefficient. Investigators in the field have proposed alternative methods of solution that do not necessitate the use of this constant. Methods such as the ones based on Hertz contact theory (Maw et al., 1976; Marghitu and Hurmuzlu, 1996), or discrete, lumped spring-dashpot based models (Stoianovici and Hurmuzlu, 1996) are few that fall in this classification. The appeal of the restitution coefficient, however, lies in the remarkably simple way of resolving the interaction between the colliding bodies. This simplicity also enables the use of momentum-based methods in the solution, which are significantly less complex than other methods of approach.

The concept of the coefficient of restitution has evolved progressively since its introduction by Newton as velocity ratio. Routh (1897) used Poisson's Hypothesis and introduced an impulse based coefficient followed by Stronge (1990) who introduced the energetic coefficient of restitution. Hurmuzlu and Marghitu (1995) and Marghitu and Hurmuzlu (1996) conducted thorough comparisons of the outcomes using the three definitions and concluded that, pending further experimental evidence, the energetic coefficient yields the most consistent results.

Stoianovici and Hurmuzlu (1996) conducted a set of experiments with slender bars falling on a massive surface. This problem was selected because it was benchmark example that was widely used in the related literature. The bars were dropped from various heights, and the pre-impact conditions were selected such that the three definitions of the coefficient of restitution yield identical results (i.e., no tangential velocity reversals at the contact point). The study confirmed several key aspects of rigid-body theory that is widely used in solving collision problems (Chatterjee and Ruina, 1997). For example, it was shown that Coulomb's law of friction remained generally valid during the collision process. In addition, for the velocity range that was considered in the experiments, the coefficient of restitution did not depend on the incidence velocity. Yet, the most surprising result was the unusual variation (as high as 80 percent of the 0–1 interval) in the coefficient of restitution as the inclination of the bars were varied from vertical to horizontal. They attributed this variation to the vibrational energy that was trapped in the bar as a result of the impact event. The authors also developed a discrete model to explain and model this behavior. They demonstrated that the outcomes predicted by the model matched the experimental results. Their study cast further questions on the practical utility of using a constant coefficient of restitution that depended on local contact properties only. The authors concluded that, in its present form, the coefficient of restitution, had a very limited applicability even when the underlying conditions of the theory were met (relatively rigid colliding bodies and low-impact speeds). Thus, the solution of impact problems, using the rigid-body impact laws can be possible for a wide range of problems when proper definitions of the coefficient of restitution are developed (Brogliato, 1996; Chatterjee and Ruina, 1997).

The goal of the present study is to generalize the concept of coefficient of restitution such that it can be applied to a wider range of impact problems. Our goal is to amend the definition of the coefficient of restitution such that it incorporates the effect of internal vibrations in planar collisions of slender members. We seek to obtain a simple algebraic form that can be used with the standard momentum based methods. We impose the requirement that the new coefficient is consistent with the classical one when the effect of vibrations diminish.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Feb. 17, 1997; final revision, Aug. 28, 1997. Associate Technical Editor: V. K. Kinra.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

In the ensuing article we will focus on the free collisions of bars with massive external surfaces. We will propose a new method to accurately predict the post impact velocities of these bars. Finally, we will verify the practical utility of the method by comparing the computed outcomes with the experimental results.

2 Experiments

The experimental setup consists of an electromagnetically actuated dropping device, a computerized high-speed video system (for kinematic data acquisition), and slender steel bar specimens with reflective markers. Here, we only describe the general features of the experimental study and its outcome as pertains to the present article, the reader is referred to Stoianovici and Hurmuzlu (1996) for further detail.

The objective of the experiments was to collide the slender bars with an external massive surface while controlling the pre-collision normal impact velocity and orientation of the bar. This was achieved by constructing a dropping device with an electromagnetic release circuit. The device was designed to drop the bars from various heights with given inclination angles. In addition, the pre-collision motion of the bar was kept planar to realize two-dimensional collisions. During each experiment, the motion of three reflective markers, attached to the bars as shown in Fig. 1, were captured by using a high-speed camera system (1000 frames/sec). Then, the kinematic data was transferred to a personal computer to compute the pre and post-collision velocities.

The experiments were conducted with six bar lengths, ranging from 63.5 mm to 600 mm, and two diameters (see Fig. 1). During the first set of experiments each bar was dropped from four different heights and inclination angles (θ) that varied from 5–90 degrees with five-deg increments. During the second set each bar was dropped from two heights with angular orientations identical to the first experiment set. The main difference between the two sets was that the first was conducted with $\frac{1}{2}''$ (12.7 mm) and the second was conducted with $\frac{1}{4}''$ (6.4 mm) diameter bars.

As noted in Stoianovici and Hurmuzlu (1996), the most striking observation from the results of the first set of experiments was that the kinematic coefficient of restitution varied significantly as a result of varying the incidence angle of the bar. In addition, within the incidence velocity range used for the experiments (1–4 m/s), the variation of the coefficient with velocity was negligible. This can be observed from Fig. 2, which depicts the kinematic coefficient of restitution obtained from experiments conducted with bar #3 for four incidence velocities.

3 Preliminaries

3.1 Discretized Analytical Model. Having observed the unusual variation in the coefficient of restitution an analytical model was developed to study the physical phenomena that causes this behavior (Stoianovici and Hurmuzlu, 1996). The model was based on the discretization of the bar into n rigid segments interconnected by a pair of viscoelastic links. Each link consists of two spring/damper combinations with stiffness k and damping coefficients b, respectively, placed symmetrically at a distance of $a/2$ from the neutral axis. These parameters were calculated as

$$k = \frac{EA(n-1)(3n-1)}{6Ln}, \quad \text{and} \tag{1}$$

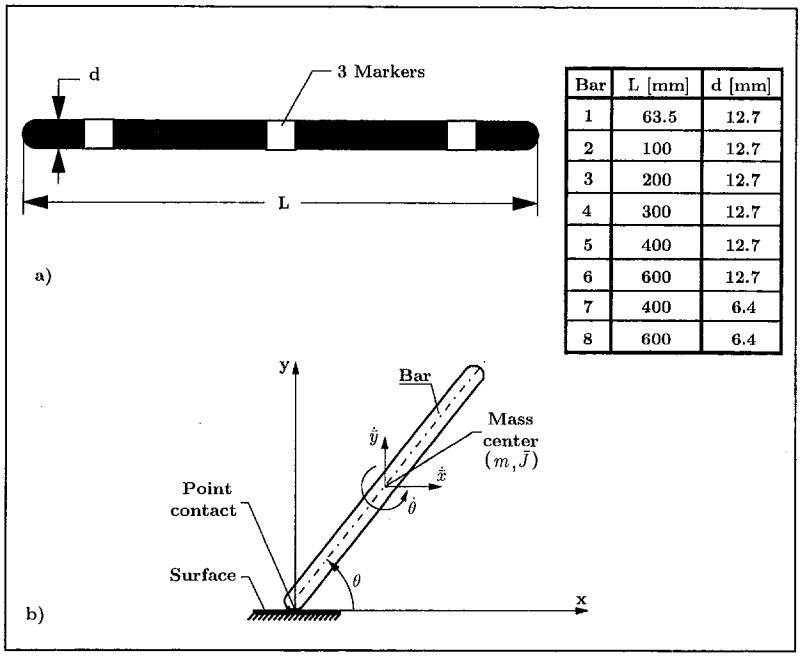
$$a = \sqrt{\frac{12I_z(7n-5)}{7A(3n-1)}} \tag{2}$$

where, I_z is the cross-sectional moment of inertia, $E = 2.1 \times 10^{11}$ N/m^2 (for steel) is the Young's modulus, and A (m^2) is the cross-sectional area of the bar. The vertical ground contact was modeled as

$$Q_n = (K - B\dot{y})|y|, \tag{3}$$

where y and $\dot{y}$ are the vertical displacement and velocity of the contacting end of the discretized bar, respectively, K is the surface stiffness, B is the ground damping coefficient. The horizontal ground contact was modeled using Coulomb's dry friction model with experimentally estimated static and dynamic friction coefficients of $\mu_s = .1$ and $\mu_k = .075$, respectively.

Numerical values of K and B were estimated experimentally as 11×10^7 N/m and $B = 2.0 \times 10^7$ Ns/m^2, respectively. The



Bar	L [mm]	d [mm]
1	63.5	12.7
2	100	12.7
3	200	12.7
4	300	12.7
5	400	12.7
6	600	12.7
7	400	6.4
8	600	6.4

Fig. 1 The bar and the system of coordinates: (*a*) geometry of the bar and marker configuration, (*b*) system of coordinates

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

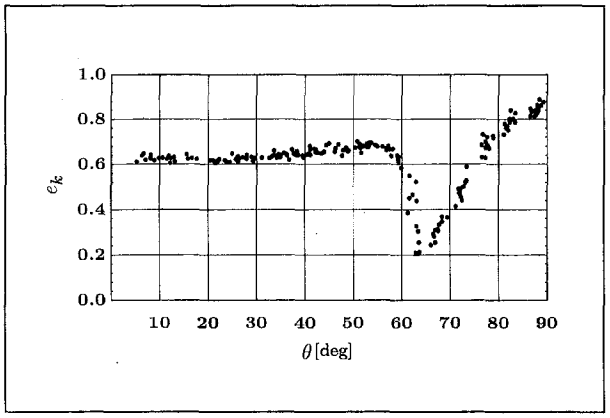


Fig. 2 Experimental kinematic coefficients of restitution for bar #3

parameter b was set equal to 100 N s/m, whereas k and a were computed using $n = 9$ and the particular length and diameter of the bar that was being considered.

In order to correlate the analytical and experimental outcomes, an analogous rigid bar was defined as having the same dimensions and material properties of the elastic bar (see Fig. 1). Kinematics of center of mass of the elastic and the rigid bars were set equal to one another. The angular acceleration of the rigid bar was computed from

$$\ddot{\theta} = \frac{m}{2\bar{J}}\{\ddot{\bar{x}}[(L-d)\sin\theta + d/2] - \ddot{\bar{y}}(L-d)\cos\theta)\} \quad (4)$$

where $\ddot{\bar{x}}$ and $\ddot{\bar{y}}$ are the central accelerations of the rigid and elastic bars, $\bar{J}$ is the centroidal moment of inertia, and m is the bar mass. The differential equations of motion of the elastic bar were numerically integrated to compute the motion of the mass center. This solution was substituted into Eq. (4) to obtain the angular acceleration of the rigid bar. Then, the motion of the rigid bar was computed by integrating Eq. (4) and the corresponding linear accelerations of the center of mass. This process was needed because the experimental data reflects the general rigid-body motion of the bar. The camera resolution was not sufficient to capture the vibrational motion of the bar. As shown in Stoianovici and Hurmuzlu (1996), the outcome of the analytical model almost perfectly matched the results of the first set of experiments that were described earlier in the article. Next we will use this numerical procedure in order to further analyze the dynamics of the collision of the slender bars considered in this study.

3.2 Internal Energy Storage. In this paper we classify the energy content of the bar through use of the following four definitions of nondimensional energy categories:

1 The percentile kinetic energy of the *rigid bar*, E_r:

$$E_r = \frac{\frac{1}{2}[m(\dot{\bar{x}}^2 + \dot{\bar{y}}^2) + \bar{J}\dot{\theta}^2]}{E_r^0} \quad (5)$$

where, $\dot{\bar{x}}$ and $\dot{\bar{y}}$ are the linear velocities of the center of mass (see Fig. 1) and E_r^0 is the total energy at the onset of the collision.

2 The percentile elastic energy stored internally by the bar, E_i:

$$E_i = \frac{\sum_{j=1}^{n} K_j + \sum_{j=1}^{n-1} P_j}{E_r^0} - E_r \quad (6)$$

where, K_j is the kinetic energy of the jth segment and P_j is the potential energy stored in the springs between the jth and $(j+1)$th segments.

3 The percentile potential energy of the surface spring, E_s:

$$E_s = -\frac{\frac{K}{4}(y|y| - y^2)}{E_r^0}. \quad (7)$$

4 The energy dissipated by the surface friction and damping, E_f:

$$E_f = 1 - [E_r + E_i + E_s]. \quad (8)$$

This term was evaluated as the difference between the total energy of the bar at the onset of the collision (100 percent) and the remaining three energy percentiles $(E_r + E_i + E_s)$.

3.3 An Expanded Definition of Micro Collision. Previously, it has been shown that (Stoianovici and Hurmuzlu, 1996) an impact that appears as a single collision to the naked eye may actually consist of a succession of multiple micro collisions. Figure 3 depicts the simulated vertical tip displacement of the elastic bar y, during the first 260 μ-sec of the collision of bar #4 for a vertical drop velocity of -1 m/s. For a drop angle of $\theta = 55$ deg, the tip of the bar separates from and reestablishes contact with the surface. In the second case ($\theta = 58$ deg), however, the tip of the bar remains in contact with the surface throughout the same time period. In both cases, the vertical velocity of the tip of the bar reverses direction during this time interval. Based on this on this observation, which is a frequent occurrence during elastic collisions, we define the following conditions for *initiation* of a micro collision:

$$\{y = 0 \quad \text{and} \quad \dot{y} < 0\} \quad \text{or} \quad \{y < 0 \quad \text{and} \quad \dot{y} = 0\}. \quad (9)$$

Similarly, a micro collision *terminates* when,

$$\{y = 0 \quad \text{and} \quad \dot{y} > 0\} \quad \text{or} \quad \{y < 0 \quad \text{and} \quad \dot{y} = 0\}. \quad (10)$$

Note that, the conditions are not ambiguous although their second parts are identical. Because, when $y < 0$ termination of a micro collision marks the start of the next one. We will use these conditions in the remainder part of the present paper to determine the number of micro collisions during a given impact period.

3.4 Critical, Sub, and Super-Critical Collisions. The exchange between the components of the energy play a key role in the excessive variation of the kinematic coefficient of restitution. Figure 4 depicts this exchange during four distinct

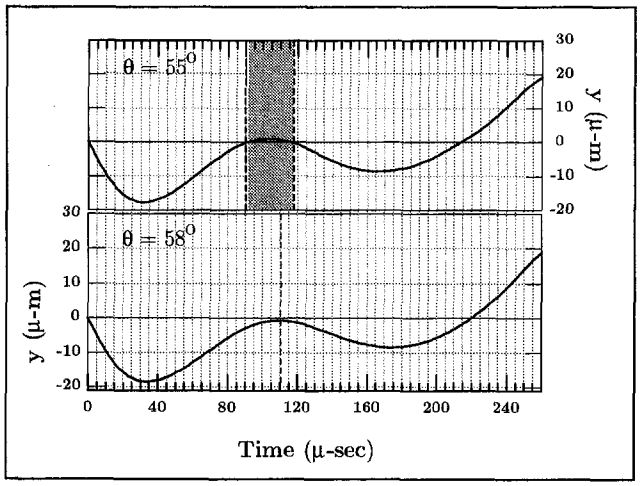


Fig. 3 Tip displacement of bar #4 during two noncentral collisions

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

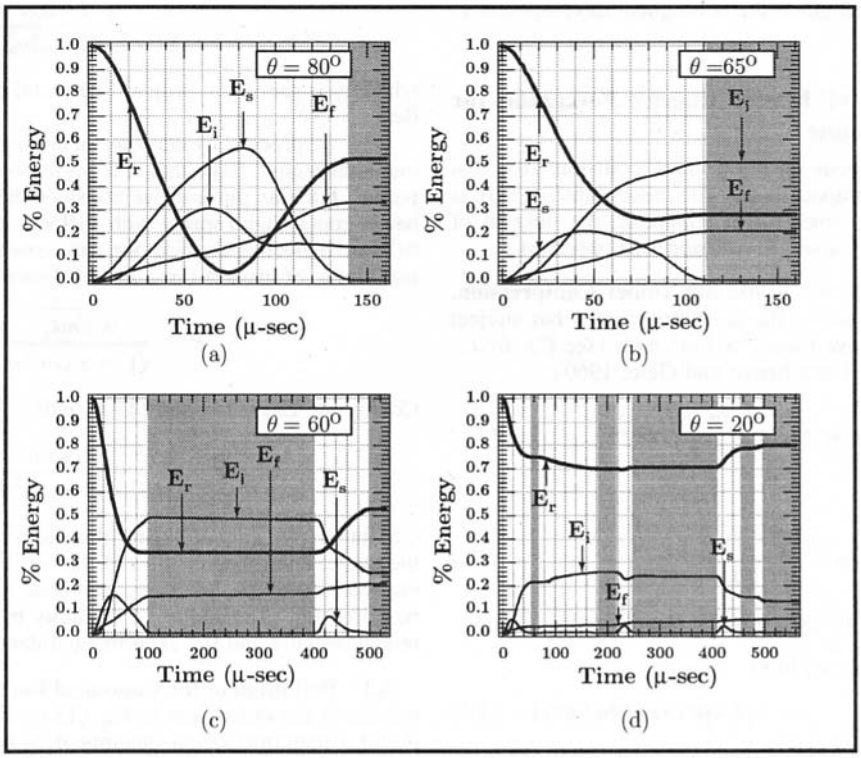


Fig. 4 Energy percentiles during the collision period

drop angles (progressively more horizontal drop from Figs. $4(a)$–$4(d)$. We define the collision that leads to the maximum internal energy storage as the *critical collision* (Fig. $4(b)$). Whereas, collisions that occur for higher and lower drop angles are termed as *super* and *subcritical collisions,* respectively (see Figs. $4(c)$ and $4(d)$). We observe that for a supercritical collision (Fig. $4(a)$), the bar comes into contact with the surface only once (the unshaded region). For subcritical collisions, however, multiple contacts are observed (Figs. $4(c)$ and $4(d)$).

Figure 5 depicts the elastic energy percentiles and number of multiple collisions for bar #3. We can clearly observe from the figure that E_i has a maximum at $\theta \approx 64$ deg (the critical angle). In addition, the transition from one to two micro collisions takes place at this angle. We also observe that the energy percentile diminishes for the centroidal collision at $\theta = 90$ deg. We should note that the behavior depicted for bar #3 is consistently observed in the energy percentiles of all bars that were considered in the present investigation. Based on these observations, we impose the following two conditions on the elastic energy percentile:

(a) There is a transition from one to two micro collisions at the critical point, where the elastic energy percentile has a global maximum.

(b) The internal elastic energy storage diminishes at the $\theta = \pi/2$.

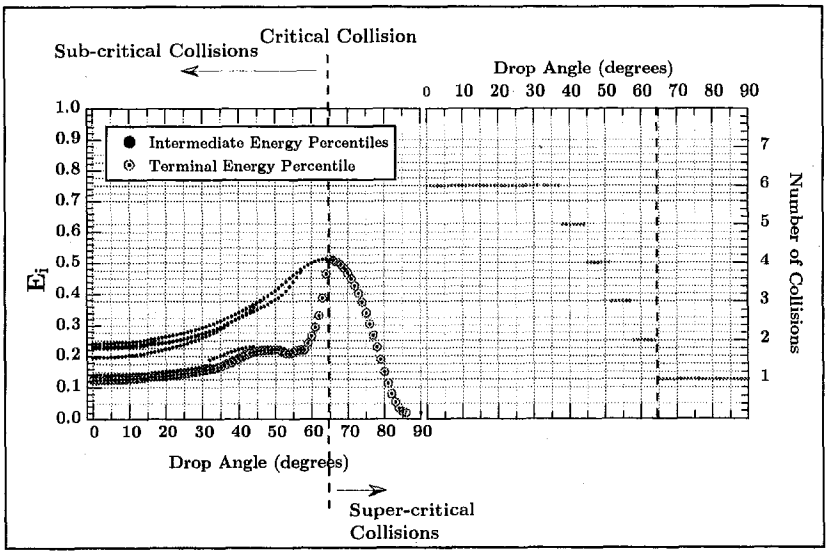


Fig. 5 Elastic energy percentiles and number of collisions

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

These conditions will guide the subsequent development in the article.

4 Approximation of Elastic Energy Percentile for Supercritical Collisions

In this section we focus on the primary collisions: the first micro collision of the impact event. The main objective here is to approximate E_i for supercritical collisions, the interval of drop angles where only a single collision was observed.

4.1 Elastic Energy of a Static Bar Under Compression. The differential equation for the deflection u of a bar subject to a nonaxial compressive force P at both ends (see Fig. 6(a)) can be written as (see Timoshenko and Gere, 1960):

$$\frac{d^2u}{dx^2} + \kappa^2 u = -\frac{P\cos\theta(L-x)}{EI_z} \tag{11}$$

where

$$\kappa = \sqrt{\frac{P\sin\theta}{EI_z}}. \tag{12}$$

Assuming fixed ends, we can solve for the deflection u as

$$u = \cot\theta[-L + x + L\cos(\kappa x) - L\cot(\kappa L)\sin(\kappa x)]. \tag{13}$$

Now, the elastic energy that is stored in the bar due to the force can be found as

$$V_e = \frac{1}{2}EI_z\int_0^L \frac{d^2u}{dx^2}\,dx = \left[\frac{P^2L^3\cos^2\theta}{4EI_z}\right]\left[\frac{1-\sin(2\kappa L)/(2\kappa L)}{\sin^2(\kappa L)}\right]. \tag{14}$$

For relatively rigid beams (κL is small) the right-hand side of Eq. (14) can be approximated as

$$V_e = \left[\frac{P^2L^3\cos^2\theta}{4EI_z}\right]\overbrace{\left[\frac{1-\sin(2\kappa L)/(2\kappa L)}{\sin^2(\kappa L)}\right]}^{\approx 2/3} \approx \frac{P^2L^3\cos^2\theta}{6EI_z}. \tag{15}$$

Assuming that this relationship is also valid for an impulsive force P, we may approximate the elastic energy percentile for a vertically dropped bar as

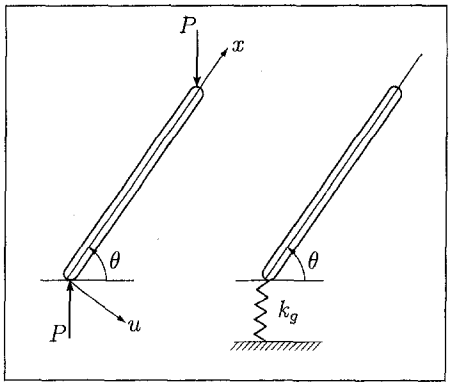


Fig. 6 (a) Statically compressed bar, (b) motion while in contact with a ground spring

$$\epsilon = \frac{V_e}{E_r^0} = \frac{P^2L^2\cos^2\theta}{3EI_z\rho A v_0^2}, \tag{16}$$

where v_0 is the initial drop velocity and ρ is the density of the bar.

Now we proceed by deriving an approximate relationship for the compressive force that acts on the bar during the collision period. For this purpose we consider the dynamics of a rigid bar in contact with spring with stiffness k_g at one end (see Fig. 6(b)). Neglecting friction one can write the expression for the magnitude of ground force P_s as follows:

$$P_s = \frac{v_0\sqrt{mk_g}}{\sqrt{1+3\cos^2\theta}}. \tag{17}$$

Combining Eqs. (16) and (17) yields

$$\epsilon = \frac{k_gL^3}{3EI_z}\frac{\cos^2\theta}{1+3\cos^2\theta}. \tag{18}$$

Obviously, we cannot expect this form to faithfully represent the elastic energy percentiles of the bars considered in the present work. We can, however, formulate a canonical expression based on Eq. (18) and test its validity by using the discretized model initially and the experimental data eventually.

4.2 Definition of the Canonical Form. The main characteristic of the expression in Eq. (18) is that it depends on cos θ and a nondimensional quantity $\eta \equiv k_gL^3/(EI_z)$. We define the following canonical form to approximate E_i^1 (the superscript denotes micro collision number):

$$\tilde{E}_i^1(\theta) = \frac{r^2}{(1-r^2)^2 + r^2/\epsilon^*} \tag{19}$$

where r is the angular ratio defined as

$$r = \frac{\cos(\theta)}{\cos(\theta_c)}. \tag{20}$$

This equation casts $\tilde{E}_i^1$ in terms of two nondimensional parameters r and ϵ^*. The constants ϵ^* and θ_c are assumed to depend on contact properties (k_g), modulus of elasticity (E), and bar geometry (L and I_z). The form of Eq. (19) is tailored such that the following conditions are met:

$$\tilde{E}_i^1(\pi/2) = 0 \quad \text{and} \quad \max_{0\le\theta\le\pi/2}\tilde{E}_i^1 = \tilde{E}_i^1(\theta_c) = \epsilon^*.$$

The primary difference between the expressions given by Eqs. (18) and (19) are the denominator terms. Equation (18) satisfies condition (b) of Section 3.4, yet it fails to satisfy condition (a). We meet this condition by adding the fourth-order term to the denominator and grouping in terms of the two nondimensional parameters such that the critical energy percentile at $\theta = \theta_c$ is a maximum point.

Now, if our approximation is a valid one, $\tilde{E}_i^1$ should approximate E_i^1 for all drop angles once the two constants ϵ^* and θ_c are properly chosen. Figure 7 depicts the E_i^1 values for bar #3 computed from the discretized model and Eq. (6). As shown on the figure, the constants θ_c and ϵ^* can be estimated as 64 deg and 0.521, respectively, from the critical impact point. The resulting canonical form produces an excellent approximation for all drop angles in the range $0 \le \theta \le \pi/2$. The validity of the canonical form was further established when we observed that the accuracy of the approximation for the particular geometry of bar #3 was equally preserved for other bar geometries as well.

4.3 The *Hr* number. The objective of this subsection is to use the discretized numerical model to represent the two parameters θ_c and ϵ^* as a function of the nondimensional param-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

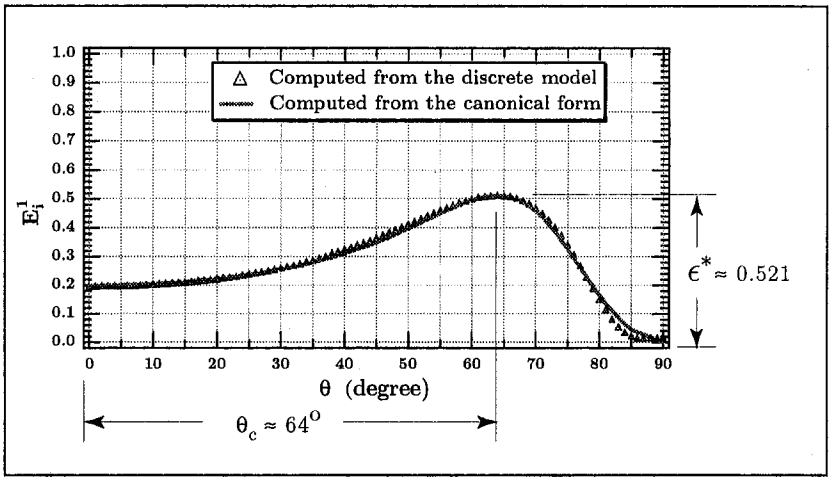


Fig. 7 Approximation of the primary elastic energy percentile for bar #3

eter η. One may compare Eqs. (18) and (19) to derive the following inferential relationship:

$$\cos^2 \theta_c \epsilon^* \sim 1/\eta.$$

Based on this relationship we propose the following nondimensional parameter, which we call the *Hr* number in this article.

$$Hr \equiv 1/\eta^{(1/3)} = \left(\frac{EI_z}{k_g L^3}\right)^{1/3}.$$

We now plot (see Fig. 8) the critical parameters $\cos \theta_c$ and ϵ^* versus the *Hr* number for the bars considered in this study by varying L and d and letting $E = 2.1 \times 10^{11}$ and $k_g = 11 \times 10^7$. One can observe from the figure that the parameters linearly depend on *Hr*, and using a linear fit can be approximated as

$$\cos \theta_c = 0.25 + 2.61Hr \tag{21}$$

and

$$\epsilon^* = 0.67 - 2.31Hr. \tag{22}$$

Stoianovici and Hurmuzlu (1996) defined a slenderness ratio as $\gamma = L/d^{1.3815}$. They showed that the critical angle and magnitude computed for the same values of γ were the same. This ratio was computed empirically from the results of their numerical simulation. In the present paper one can observe that the *Hr* number includes the slenderness ratio for bars with circular cross sections ($Hr \sim d^{1.3333}/L$). Based on this observation, now we investigate whether the critical parameters remain the same for identical values of the *Hr* number, which includes material and contact properties as well as geometric factors. For this purpose we compute the critical parameters for a bar with a fixed $Hr = 0.0673$ (bar #3), and vary L and d, L and E, and L and k_g, respectively. The resulting parameters are shown in Fig. 9. As we can observe from the figure, the critical angle

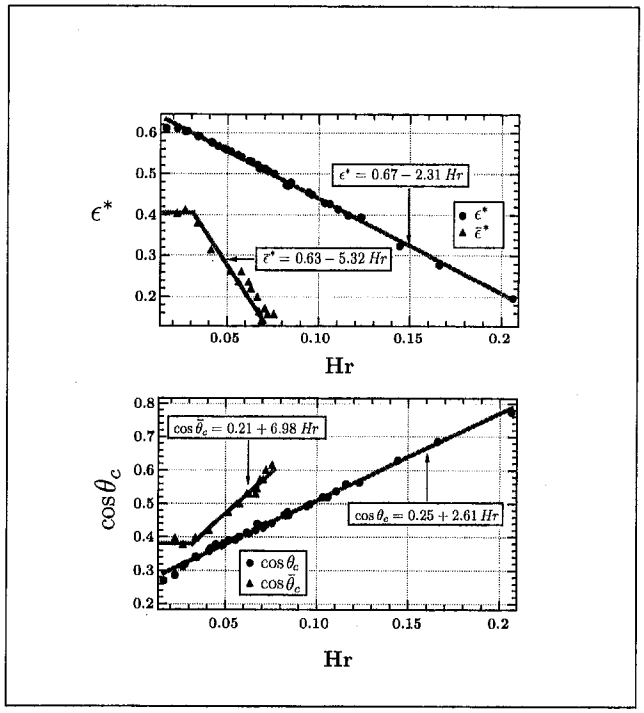


Fig. 8 Critical parameters

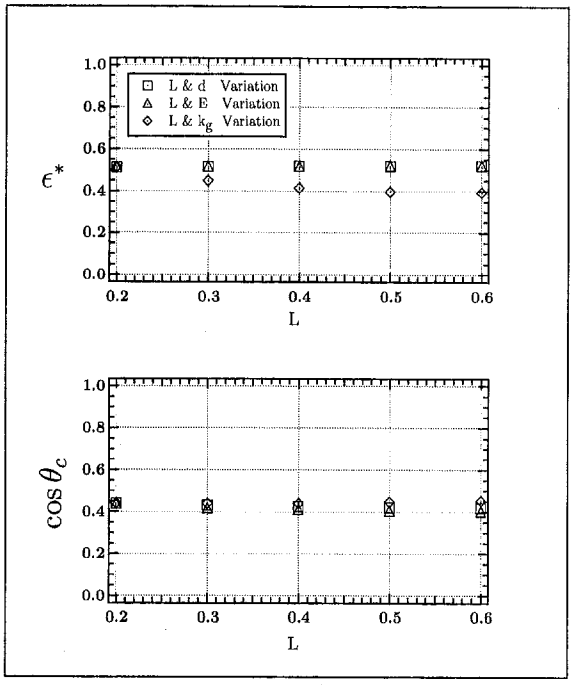


Fig. 9 Variation of the critical parameters at $H_r = 0.0673$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

remains unchanged for all variations. The critical magnitude, however, remains unchanged for the first two variations but varies significantly for the third. Thus, we may conclude that for a given drop surface (fixed k_g), the critical parameters are function of the Hr number only. In addition, two bars dropped on the same surface will have the same critical parameters if and only if they have the same Hr number. Thus, the approximations given in Eqs. (21) and (22) (obtained for a specific E) can be used for bars with with different elasticity moduli.

So far, we have shown that the elastic percentiles at the critical angles are invariant with respect to the Hr number when k_g is fixed. This characteristic should be preserved for all supercritical collisions based on the specific form of Eq. (19). For subcritical collisions, where we have multiple collisions, we cannot expect the invariance principle to be automatically satisfied. If it is valid, however, then this would provide a guideline to develop expressions for the approximation of elastic energy percentile for subcritical collisions. Figure 10.*a* depicts the terminal energy percentiles (i.e., E_i at the end of all micro collisions) computed from the discretized model for five bar lengths: $L = 0.2, 0.3, 0.4, 0.5, 0.6$ m with $E = 2.1 \times 10^{11}$ and $k_g = 11 \times 10^7$. In each case the bar diameter was selected such that Hr remained constant ($Hr = 0.0673$). As we can observe from the figure, the energy percentiles are generally invariant to the parameter variations of the bars for all drop angles when the Hr number is fixed. One may note that there is an observable variation for drop angles in the range 40 deg $< \theta <$ 50 deg. We neglect this minor variation since our objective here is to obtain a simple representation of the elastic energy that arises from a complex sequence of collisions. In addition, the results were obtained through the discrete model, which itself was an approximation. Figure 10(b) depicts the terminal percentiles computed for the previous six bar lengths and $Hr = 0.0673$ with $d = 0.0127$ and $k_g = 11 \times 10^7$. In this case the bar diameter was fixed and the Young's modulus E was adjusted to achieve constant Hr. Again, we observe that the elastic energy percentiles can be considered as invariant for a constant Hr number when the bar parameters are varied.

5 Approximation of the Elastic Energy Percentile for Subcritical Collisions

The subcritical collisions are significantly more complex and may involve a succession of multiple collisions. To study this problem, we follow an empirical approach. Figure 11 depicts three plots of energy percentiles for gradually increasing Hr values. We identify two significant regions of subcritical collisions where the elastic energy exhibits two distinct behaviors. In the first region, which starts from the critical point (shaded areas in the figures) we observe a sudden drop in elastic energy accompanied by the appearance of a second micro collision. Subsequently, a succession of more collisions take place in a second region that extends from the appearance of a third micro collision to a horizontal drop angle. We approximate the energy percentile in the first subcritical region as

$$\tilde{E}_i^2(\theta) = \frac{1/\epsilon^*}{[(1 - r^2)^2 + r^2/\epsilon^*]^2}. \tag{23}$$

This relationship is based on our original canonical form and satisfies the following matching condition at the critical point:

$$\tilde{E}_i^2(\theta_c) = \tilde{E}_i^1(\theta_c) = \epsilon^*. \tag{24}$$

In addition, the expression is tailored to preserve the invariance principle that was defined in the previous section. The resulting approximated energy percentiles are depicted in Fig. 11 as the dark gray curves on the respective plots.

As the drop angle is decreased, we move on to the second region, where a multitude of micro collisions appear. We observe that the terminal energy percentile $\tilde{E}_i^3$ (the elastic energy percentile at the end of all micro collisions) follows a similar

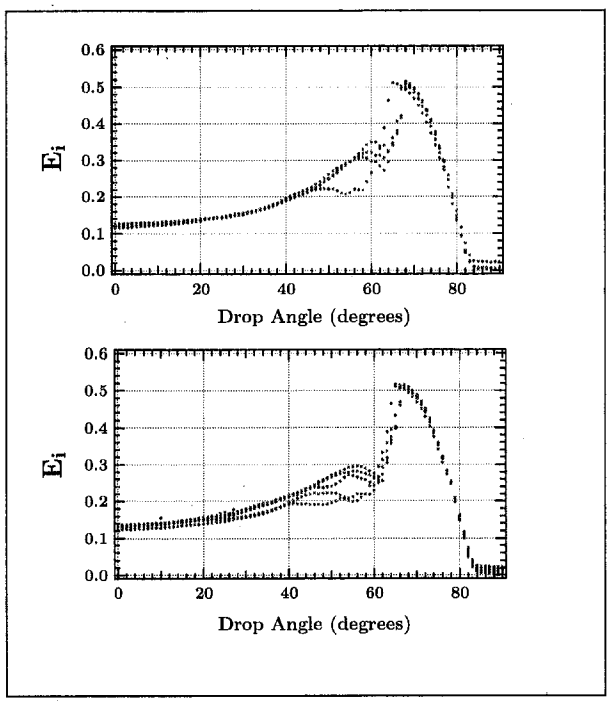


Fig. 10 Variation of the elastic energy percentiles at $H_r = 0.0673$

pattern to the one observed for the primary percentiles. Based on this observation we define the following canonical form for $\tilde{E}_i^3$ in the second region of subcritical collisions:

$$\tilde{E}_i^3(\theta) = \frac{\bar{r}^{\alpha}\bar{\epsilon}^{*\,1-\alpha/2}}{[(1 - \bar{r}^2)^2 + \bar{r}^2/\bar{\epsilon}^*]^{\alpha/2}}, \tag{25}$$

where $\bar{r}$ is the angular ratio defined as

$$\bar{r} = \frac{\cos(\theta)}{\cos(\bar{\theta}_c)}. \tag{26}$$

The parameters $\bar{\theta}_c$ and $\bar{\epsilon}^*$ are analogous to the previously defined parameters θ_c and ϵ^*, respectively. Again, the form of the expression given in Eq. (25) is based on the one given in Eq. (19). Furthermore, it satisfies the following conditions:

$$\tilde{E}_i^3(\pi/2) = 0 \quad \text{and} \quad \max_{0 \le \theta \le \pi/2} \tilde{E}_i^3 = \tilde{E}_i^3(\bar{\theta}_c) = \bar{\epsilon}^*.$$

Next, we follow an approach similar to the one followed for supercritical collisions. We identify the critical parameters $\bar{\theta}_c$ and $\bar{\epsilon}^*$ from numerical results obtained form the discretized model. Then, we tune the parameter α such that the profile given by Eq. (25) produces a good fit for all drop angles in the second region of the subcritical impacts of bar #3. Accordingly, we obtain $\alpha = 3$ as the value of the parameter. The approximate energy percentiles for four bar geometries are depicted as the gray curves in Fig. 11.

Finally, as we have done for the supercritical collisions, we plot the parameters $\bar{\theta}_c$ and $\bar{\epsilon}^*$ against Hr (see Fig. 8). One can observe from the figure that the parameters linearly depend on Hr in a certain interval and saturate for $Hr < 0.031$. The resulting expressions for the critical parameters can be written as

$$\cos \bar{\theta}_c = \min [0.21 + 5.32Hr, 0.38] \tag{27}$$

and

$$\bar{\epsilon}^* = \min [0.63 - 6.98Hr, 0.41]. \tag{28}$$

6 The Coefficient of Vibrational Energy Loss

In this section we combine the results of the previous two sections to obtain a single nondimensional coefficient $e_G(\theta,$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

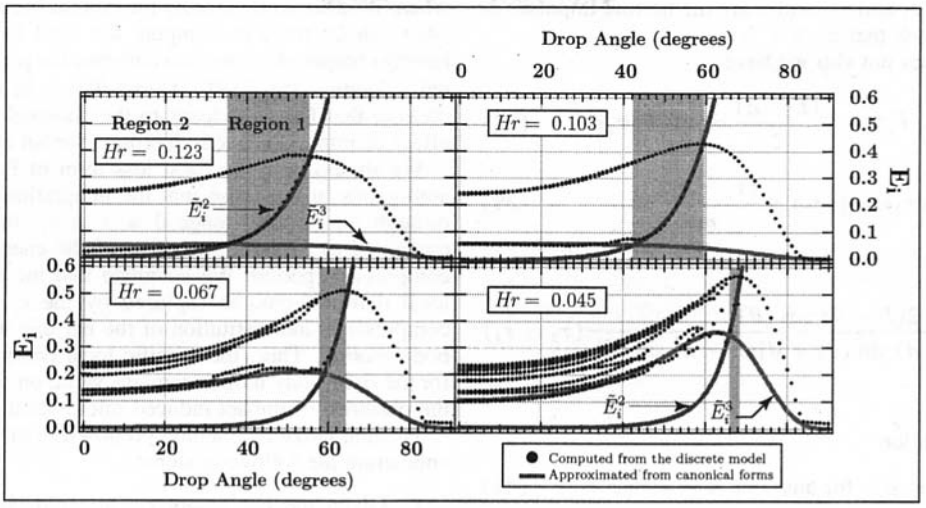


Fig. 11 Elastic percentile regions for subcritical collisions

Hr), which represents the vibrational energy that remains in the bar at the end of the collision process. Using Eqs. (19), (20), (21), (22), (23), (25), (26), (27), and (28) we may write

$$e_G(\theta, Hr) = \begin{cases} \dfrac{r^2}{(1-r^2)^2 + r^2/\epsilon^*} & \text{if } \theta \geq \theta_c \\ \max\left[\dfrac{1/\epsilon^*}{[(1-r^2)^2 + r^2/\epsilon^*]^2}, \right. & \\ \left. \dfrac{\bar{r}^3/\bar{\epsilon}^{*1/2}}{[(1-\bar{r}^2)^2 + \bar{r}^2/\bar{\epsilon}^*]^{3/2}} \right] & \text{otherwise} \end{cases} \tag{29}$$

where

$$r = \frac{\cos(\theta)}{\cos(\theta_c)} \quad \text{and} \quad \bar{r} = \frac{\cos(\theta)}{\cos(\bar{\theta}_c)}$$

$$\cos\theta_c = 0.25 + 2.61Hr \text{ and}$$

$$\cos\bar{\theta}_c = \min[0.21 + 6.98Hr, 0.38]$$

and

$$\epsilon^* = 0.67 - 2.31Hr \text{ and } \bar{\epsilon}^* = \min[0.63 - 5.32Hr, 0.41].$$

Equation (29) represent a remarkably simple expression to represent the complex behavior of the bar as a result of multiple micro collisions.

7 Revising the Coefficient of Restitution

An apparent conclusion of the preceding analysis is that the energy loss during collision has two main components. The first component is the energy loss due to local deformations, which is the one captured by the classical definition. The second component is due to the vibrational energy that remains in the bar after the impact. A distinguishing characteristic between these two forms of energy loss is the vibrational energy can be partially recovered by the bar as a result of subsequent collisions while the local loss is permanent.

Based on this line of thinking, we propose the following form for a new energy-based coefficient of restitution for collision of bars,

$$e = e_L \oplus e_G, \tag{30}$$

where e_L is the loss factor that represents the local losses at the collision point. The term e_G accounts for the global loss that is due to the vibrational energy. We will treat e_L as the energetic coefficient of restitution as defined in Strong (1990). The overall energy ratio is in turn computed as the "sum" of e_L and the factor e_G to account for local as well as global energy losses due to impact. As it will become clear in the proceeding subsections, the two coefficients are not simply added, but their effects on energy are superposed.

7.1 Equations of Motion. The centroidal velocities of the rigid bar during a single point impact with a massive external surface while slipping can be written as

$$\dot{\bar{x}}(\tau_2) = \dot{\bar{x}}(\tau_1) - \frac{\mu_k(\tau_2 - \tau_1)}{m}\nu \tag{31}$$

$$\dot{\bar{y}}(\tau_2) = \dot{\bar{y}}(\tau_1) + \frac{(\tau_2 - \tau_1)}{m} \tag{32}$$

$$\dot{\theta}(\tau_2) = \dot{\theta}(\tau_1) - \{(L-d)\cos(\theta) + \mu_k\nu[(L-d)\sin(\theta) + d]\}\frac{(\tau_2 - \tau_1)}{2\bar{J}} \tag{33}$$

with

$$\nu = \begin{cases} -\text{sgn}[F_r] & \text{if } |F_r| > \mu_s \text{ and } \dot{x}_t(\tau_1) = 0 \\ \text{sgn}[\dot{\bar{x}}(\tau_1)] & \text{if } \dot{x}_t(\tau_1) \neq 0, \end{cases} \tag{34}$$

and the contact force ratio F_r given as

$$\frac{F_x}{F_y} = F_r = \frac{(L-d)^2\sin(2\theta)/2}{(L-d)\sin(\theta)[(L-d)\sin(\theta) + d] + \dfrac{4\bar{J}}{m}}, \tag{35}$$

where the horizontal and vertical velocities of the contact point are given by

$$\dot{x}_t(\tau) = \dot{\bar{x}}(\tau) + \frac{(L-d)}{2}\sin(\theta)\dot{\theta}(\tau) \tag{36}$$

$$\dot{y}_t(\tau_i) = \dot{\bar{y}}(\tau) - \frac{(L-d)}{2}\cos(\theta)\dot{\theta}(\tau), \tag{37}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

respectively. Here, τ_1 and τ_2, and τ are the normal impulses at the contact point such that $\tau_2 > \tau > \tau_1$.

When the bar does not slip we have

$$\dot{\bar{x}}(\tau_2) = \dot{\bar{x}}(\tau_1) - \frac{(L-d)}{2}\sin(\theta)\dot{\theta}(\tau_2) \tag{38}$$

$$\dot{\bar{y}}(\tau_2) = \dot{\bar{y}}(\tau_1) + \frac{(\tau_2 - \tau_1)}{m} \tag{39}$$

$$\dot{\theta}(\tau_2) = \dot{\theta}(\tau_1) - \frac{2(L-d)\cos(\theta)}{4\bar{J} + m[(L-d)\sin(\theta) + d](L-d)\sin(\theta)}(\tau_2 - \tau_1) \tag{40}$$

subject to the condition

$$|F_r| \le \mu_s \quad \text{for any} \quad \tau_1 < \tau < \tau_2. \tag{41}$$

The horizontal and vertical velocities of the contact point are given by

$$\dot{x}_t(\tau) = 0 \tag{42}$$

$$\dot{y}_t(\tau) = \dot{\bar{y}}(\tau) - \frac{(L-d)}{2}\cos(\theta)\dot{\theta}(\tau), \tag{43}$$

respectively.

Equations (31) to (37) and (38) to (43) can be used sequentially (see Marghitu and Hurmuzlu, 1996), depending on the slip conditions, to solve for the three variables $\dot{\bar{x}}(\tau)$, $\dot{\bar{y}}(\tau)$, and $\dot{\theta}(\tau)$ for the impact interval $0 \le \tau \le \tau_f$. The normal impulse τ_f, that marks the end of the collision process will be computed in the next subsection by considering the two types of energy losses mentioned earlier.

7.2 Solution of the Impact Problem. The energy loss due to local deformation can be calculated by considering the work done by the vertical component of the contact force as follows:

$$W_L = \int_{y_{t_1}}^{y_{t_2}} F_y dy_t = \int_0^{\tau_f^*} \dot{y}_t d\tau \tag{44}$$

since $F_y dt = d\tau$. We postulate that this loss can be computed by following a procedure that is identical to the classical approach. Specifically, we use the energetic coefficient of restitution that was proposed in Strong (1990). Accordingly, we first compute the impulse at maximum compression τ^* (the vertical impulse for which $\dot{y}_t = 0$). Then, the energy loss can be computed as

$$W_L = (1 - e_L^2)\left|\int_0^{\tau^*} \dot{y}_t(\tau)d\tau\right|. \tag{45}$$

The energy loss due to vibrational energy, on the other hand, can be expressed as

$$W_G = e_G E_r^0. \tag{46}$$

Writing the energy balance for the collision process yields

$$\frac{m}{2}[\dot{\bar{x}}^2(\tau_f) + \dot{\bar{y}}^2(\tau_f)] + \frac{\bar{J}}{2}\dot{\theta}^2(\tau_f) = E_r^0 - W_L - W_f - W_G$$

$$= (1 - e_G)\left\{\frac{m}{2}[\dot{\bar{x}}^2(0) + \dot{\bar{y}}^2(0)] + \frac{\bar{J}}{2}\dot{\theta}^2(0)\right\}$$

$$- (1 - e_L^2)\left|\int_0^{\tau^*} \dot{y}_t(\tau)d\tau\right| - \mu_k\left|\int_0^{\tau_f} \dot{x}_t(\tau)d\tau\right|, \tag{47}$$

where W_f is the work done by the ground friction force. Equation (47) can be used to compute the final impulse τ_f. Finally, having computed τ_f, one can compute the post collision centroidal velocities $\dot{\bar{x}}(\tau_f)$, $\dot{\bar{y}}(\tau_f)$, and $\dot{\theta}(\tau_f)$. In addition, one may observe that Eq. (47) leads to the classical solution when the effect of impact induced vibrations vanish (i.e., $e_G = 0$).

We should note the local loss term in Eq. (47) may seem ambiguous in the sense that the integration is not carried out through the impulse range $0 \le \tau \le \tau_f$. Instead, this term is computed as a specific fraction of the energy loss during the compression period. We postulate that the energy loss due to local deformations, as captured by the e_L, is caused by the compression and restitution of the bar due to its general rigid-body motion. Thus, the specific local energy loss formulation for the rigid-body motion remains valid, on the average, despite the presence of impact induced microvibrations.

To summarize the method proposed in the present article we enumerate the following steps:

1 Given the bar geometry, material, and contact surface properties (L, I_z, E, and k_g) and the drop angle θ, use Eq. (29) to compute e_G.

2 Use Eqs. (31) to (43), the pre-impact velocities, the coefficients of friction μ_s and μ_k, and the coefficients of restitution e_L and e_G to obtain Eq. (47) in terms of the impulse at the end of the collision process, τ_f.

3 Use τ_f to compute the post-impact velocities.

8 Experimental Verification

Now we use the procedure described in the previous section to compute the post-collision velocities of the bars that were considered in the experimental study. First we estimated the local coefficient of restitution from the vertical drop of bar #3 as $e_L = 0.86$. The resulting kinematic coefficients of restitution for the six bars that were tested are plotted in Fig. 12. As can be observed the procedure duplicates the experimental outcome in an accurate fashion. The approximation somehow suffers for very slender bars, especially in the subcritical region. This can also be partly attributed to the experimental error in acquiring the data for very short and very slender bars. We may note that for bars #7 and #8 the impact-induced vibrations were visible to the naked eye.

Finally, we conducted two sets of additional experiments to verify the invariance principle that we have developed in the article. For this purpose we used a steel bar with a diameter 6.35 mm and a length of 240 mm such that it had the same Hr number as bar #6. We also used an aluminum bar ($E = 7.1 \times 10^{10}$) with a diameter of 12.7 mm and a length of 430 mm, which produces the same Hr number as bar #6. We conducted additional experiments to check the effect of the dropping surface. In these experiments, bar #6 was directly dropped on the concrete floor of the laboratory. Results of the experiments are depicted in Fig. 13. It can be seen that the experimental data qualitatively and quantitatively demonstrate the equivalence of the three bars and the validity of the solutions obtained by using the methods proposed in this article.

9 Conclusions

In a recent article, Stoianovici and Hurmuzlu (1996) has shown that impact-induced vibrations in free-drop experiments of slender bars lead to significant variations in rebound velocities. They have shown that these variations were mainly due to vibrational energy trapped in the bars after the collision process. They concluded that the classical definition of the coefficient of restitution had limited applicability in solving impact problems of slender bars. The objective of this paper is to generalize the concept of coefficient of restitution such that it can be used in the solution of a wider range of impact problems. In this work we develop a new method to solve planar collision problems of

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

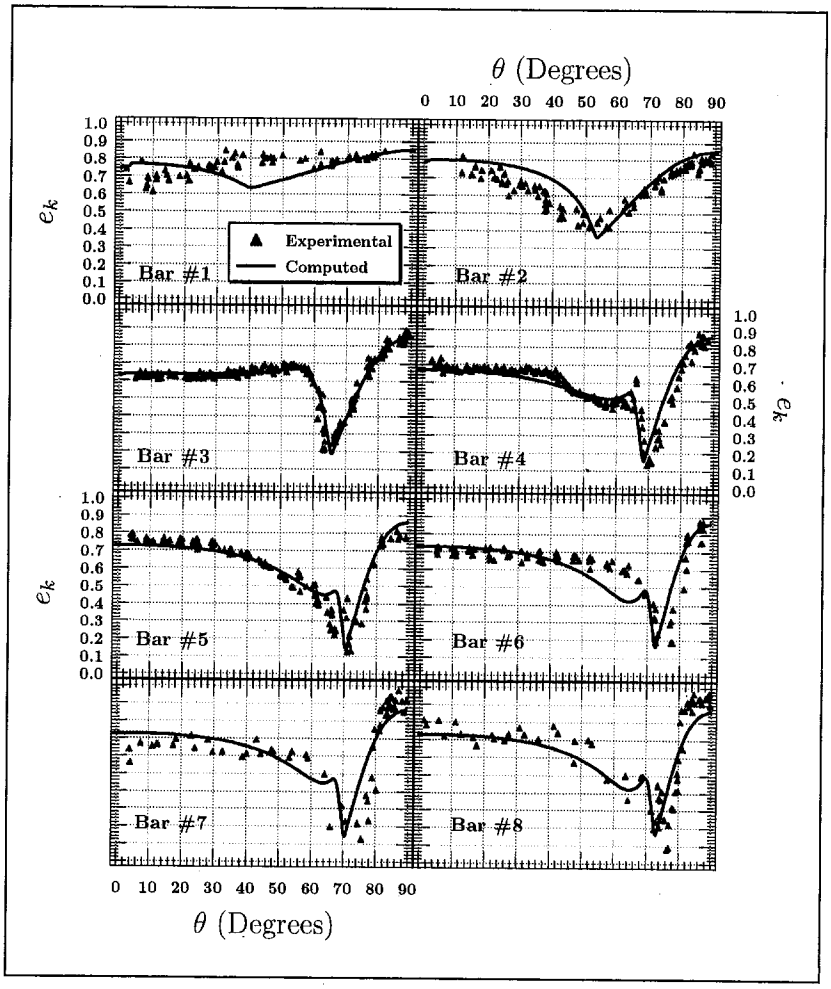


Fig. 12 Experimental and computed coefficients of restitution

slender bars with external massive surfaces. The main results presented here can be summarized as follows:

1 The energy loss due to impact induced vibrations is represented as a simple scalar variable called the elastic energy percentile. This variable depends on the inclination angle and a nondimensional parameter (the *Hr* number), which represents the bar geometry, material, and it includes stiffness of the external impact surface.

2 The elastic energy percentiles of two bars with the same *Hr* number, dropping surface, and inclination angle are equal.

3 Using the elastic energy percentile, the classical definition of the coefficient of restitution is modified to include the energy loss due to the vibrational energy trapped in the bar.

4 The new coefficient of restitution is used in the traditional momentum based methods to predict the rebound velocities of slender bars with external surfaces. The new solution method is consistent with the classical approach and it becomes equivalent to it as the effect of impact induced vibrations diminish.

5 Predictions of the proposed concepts and methods are shown to be in agreement with experimentally observed outcomes.

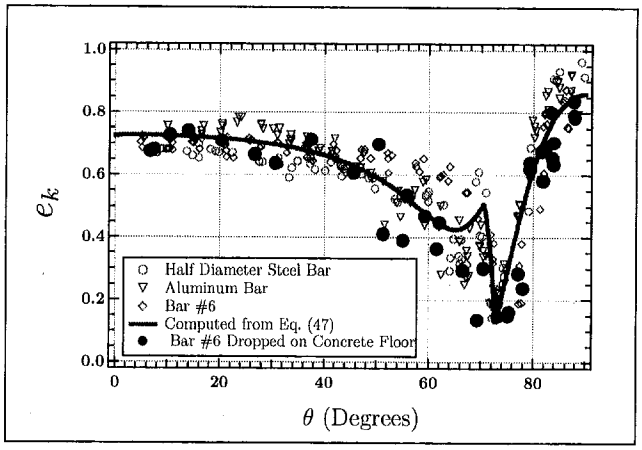


Fig. 13 Experimental verification of the *Hr* number

In closing we would like to emphasize that the present study addresses free-drop experiments conducted at relatively low-impact speeds. In addition we only consider relatively rigid bars with constant cross-sectional area and rounded ends. In addition, there were no extensive damage on the colliding objects resulting from these impacts (as we focus on low-velocity impacts). We believe that the proposed theory can be extended to collision problems of constrained mechanisms such as kinematic chains, a natural next phase of the present investigation. Different geometries can also be considered in the future. The most challenging next step, however, seems to be the extension of the analysis to three dimensions.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Acknowledgment

This material is based upon work supported by the National Science Foundation Grant No. MSS-9203097. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Brogliato, B., 1996, "Lecture Notes in Control: Nonsmooth Impact Mechanics; Models, Dynamics and Control," Springer, London, p. 107.

Chatterjee, A., and Ruina, A., 1997, discussion, "A Critical Study of the Concepts of Rigid Body Collision Theory," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 247–248.

Hurmuzlu, Y., and Marghitu, D., 1994, "Multi-contact Collisions of Kinematic Chains With External Surfaces," *International Journal of Robotic Research,* Vol. 13, No. 1, pp. 82–92.

Marghitu, D. B., and Hurmuzlu, Y., 1995, "Three-Dimensional Rigid-Body Collisions with Multiple Contact Points," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 725–732.

Marghitu, D. B., and Hurmuzlu, Y., 1996, "Nonlinear dynamics of an Elastic Rod with Frictional Impact," *Journal of Nonlinear Dynamics,* Vol. 10, pp. 187–201.

Maw, N., Barber, J. R., and Fawcett, J. N., 1976, "The Oblique Impact of Elastic Spheres," *Wear,* Vol. 38, pp. 101–114.

Newton, I., 1686, *Philosophiae Nuturalis Principia Mathematica,* Reg. Soc. Preases, London.

Routh, E. J., 1860, *Dynamics of a System of Rigid Bodies,* MacMillan, London.

Stronge, W. J., 1990, "Rigid Body Collisions with Friction," *Proceedings of Royal Society, London,* Vol. A431, pp. 169–181.

Stoianovici, S. P., and Hurmuzlu, Y., 1996, "A Critical Study of the Concepts of Rigid-Body Collision Theory," ASME JOURNAL OF APPLIED MECHANICS, Vol. 63, pp. 307–316.

Timoshenko, D., and Gere, J. M., 1961, "Theory of Elastic Stability," McGraw-Hill, New York.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

H. Canbolat

D. Dawson

C. Rahn[1]
Assoc. Mem. ASME

S. Nagarkatti

Center for Advanced Manufacturing,
Clemson University,
Clemson, SC 29634-0915

Adaptive Boundary Control of Out-of-Plane Cable Vibration

This paper develops active controllers for the out-of-plane vibration of a flexible cable using boundary actuators and sensors. An exact model knowledge controller exponentially stabilizes the cable displacement assuming known system parameters. An adaptive controller asymptotically stabilizes the cable displacement while compensating for parametric uncertainty in the actuator mass and cable tension. The performance of the controllers is experimentally demonstrated.

1 Introduction

Cables are used in many engineering applications due to their inherent low weight, flexibility, strength, and storability. The transverse stiffness of a cable depends on its tension and length, however, so long, sagged cables can vibrate excessively in response to relatively small disturbances. This vibration degrades the performance of the cable system and ultimately leads to failure.

To understand cable structure vibration, many researchers have developed sophisticated modeling and analysis techniques (see Irvine, 1981, for a review). Small sag approximation techniques (Soler, 1970), lumped parameter models (West et al., 1975), finite element techniques (Fried, 1982), and Galerkin methods (Perkins and Mote, 1987) represent several approaches used to predict the behavior of cables in a variety of applications. These analysis tools motivate the development of passive vibration control methods, including increase of the cable tension and/or addition of damping mechanisms. Increased tension induces high stress, however, reducing the cable life. Cable dampers can reduce the resonant forced response but may have little effect on the transient response.

Active vibration control has only recently been applied to cable systems. Cost and/or feasibility of this method often dictates that control actuators be located at the boundaries of the cable span. Fujino et al. (1993) developed axial boundary force control laws for a single-mode cable model and experimentally demonstrated a large reduction in peak resonant response. Fujino and Susumpow (1995) extended this approach to multiple cable modes. Baicu et al. (1996a) developed a linear feedback controller for the out-of-plane vibration of a distributed cable model. The proposed controller consisted of boundary position, velocity, and slope feedback. The cable position was proven to be stable under the proposed control. Simulation and experimental results demonstrated very good vibration damping.

Researchers have also applied boundary control to many other flexible systems. The motion of flexible gantry robots can be regulated with boundary control (Luo et al., 1995). Boundary controllers have been developed to stabilize the vibration associated with flexible beam-like structures (e.g., aircraft wings, robot links, or space structures) (Chen et al., 1987; Morgül, 1991; Rahn and Mote, 1993; Canbolat et al., 1997). Related boundary control work for string-like dynamic models can also be found in Morgül (1994) and Shahruz and Krishna (1996). Interestingly, the control structure proposed in this paper reduces to a form which is similar to the one given in Morgül (1994) if the cable tension is assumed to be constant.

In this paper, we develop new control strategies for the distributed cable model given in Baicu et al. (1996a). Specifically, we develop a model-based controller that exponentially stabilizes the displacement of the cable given exact knowledge of the mechanical system parameters and measurements of the slope, slope-rate, and velocity at the cable's actuated boundary. We then illustrate how the exact model knowledge controller can be redesigned as an adaptive controller which asymptotically stabilizes the cable displacement while compensating for parametric uncertainty. The control approach differs from the previous work in that (i) the stability analysis utilizes more intuitive mathematical tools to illustrate the exponential and asymptotic stability results, (ii) a new Lyapunov-like function is crafted to deal with the spatially varying tension, and (iii) adaptive nonlinear control techniques are applied.

2 Mathematical Model

Figure 1(*b*) shows a cable sagging under gravity loading. The cable vibrates perpendicular to the plane formed by the equilibrium cable configuration. Figure 1(*a*) shows the out-of-plane displacement $U(S, T)$ where S is the arc length coordinate and T is time. The left boundary is pinned and the right boundary is free to translate in the out-of-plane direction. A control force $F(T)$ applies in the out-of-plane direction at the left boundary. From Baicu et al. (1996a), the linearized out-of-plane equation of motion, valid for small amplitude motion, is

$$\rho A U_{TT} - [P(S)U_S]_S = 0, \tag{1}$$

where ρA is the cable mass/length and subscripts indicate partial differentiation. The equilibrium tension

$$P(S) = \sqrt{P_0^2 + \left[\rho A g\left(S - \frac{L}{2}\right)\right]^2}, \tag{2}$$

where P_0 is the midspan ($S = L/2$) or horizontal component of the cable tension, g is the acceleration due to gravity, and L is the cable length. The pinned boundary condition at $S = 0$ is

$$U(0, T) = 0, \tag{3}$$

and the controlled boundary condition at $S = L$ is

[1] To whom correspondence should be addressed. e-mail: cdrahn@eng.clemson.edu.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Aug. 27, 1997; final revision, May 11, 1998. Associate Technical Editor: A. A. Ferri.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

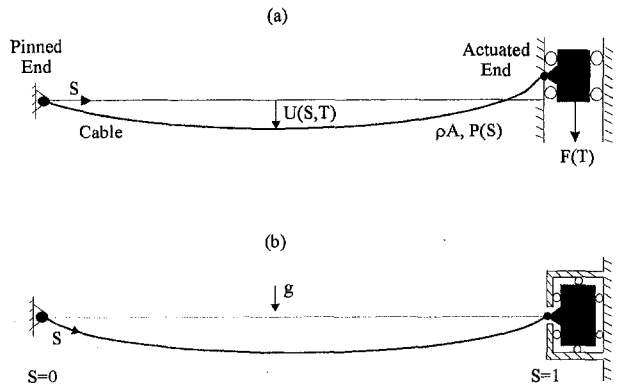


Fig. 1 Schematic diagram of the cable system: (*a*) top view, (*b*) side view

$$MU_{TT} + P(S)U_S - F_f = F, \tag{4}$$

where M is the actuator mass and $F_f(U_T(L, T))$ models friction and/or damping in the actuator.

Nondimensionalization simplifies the control development. Substitution of the nondimensional variables

$$s = \frac{S}{L}; \quad t = T\sqrt{\frac{g}{L}}; \quad u = \frac{U}{L}; \quad m = \frac{M}{\rho AL}; \quad f = \frac{F}{\rho AgL};$$

$$p(s) = \frac{P(S)}{\rho AgL} = \sqrt{p_0^2 + \left(s - \frac{1}{2}\right)^2}; \quad p_0 = \frac{P_0}{\rho AgL} \tag{5}$$

into Eq. (1) yields

$$u_{tt}(s, t) - [p(s)u_s(s, t)]_s = 0 \tag{6}$$

with boundary conditions[2]

$$u(0, t) = 0, \tag{7}$$

$$mu_{tt}(1, t) + p(1)u_s(1, t) + Y(u_t(1, t))\phi = f(t), \tag{8}$$

where $Y(u_t(1, t))\phi$ is the nondimensionalized F_f linearly parameterized as a known regression matrix $Y(u_t(1, t)) \in \Re^{1\times q}$ multiplied by a constant parameter vector $\phi \in \Re^{q\times 1}$. For example, an actuator with viscous damping b and Coulomb friction μ has

$$Y(u_t(1, t)) = [u_t(1, t) \quad \text{sgn}\,(u_t(1, t))], \tag{9}$$

$$\phi^T = [b \quad \mu], \tag{10}$$

where sgn $(\cdot)$ is the signum function. We will also assume that $Y(u_t(1, t))$ remains bounded $\forall t \in [0, \infty)$ if $u_t(1, t)$ is bounded $\forall t \in [0, \infty)$.

3 Control Formulation

In this section of the paper, we present two control strategies that stabilize the distributed cable vibration Eqs. (6)–(8) using the boundary control force $f(t)$. First, we define the structure and closed-loop boundary dynamics of a controller that requires exact knowledge of the system parameters. Using a Lyapunov-like stability analysis, we prove this exact model knowledge controller drives the cable displacement to zero exponentially fast. Dynamic compensation using an adaptive controller allows implementation without knowledge of the cable tension and actuator mass. This controller, however, provides only asymptotic regulation of the cable displacement.

[2] The pinned boundary conditions of (7) imply that $u_t(0, t) = 0$.

3.1 Exact Model Knowledge Control Law. The exact model knowledge control law is defined as follows:

$$f = -mu_{st}(1, t) + p(1)u_s(1, t) - k_s(u_t(1, t) + u_s(1, t)) + Y(u_t(1, t))\phi, \tag{11}$$

where k_s is a positive control gain. Note that the actuator mass m and tension $p(1)$ must be known to implement this controller.

THEOREM 1. *Given the field equation of (6) and the boundary conditions given by (7) and (8), the boundary controller given by (11) ensures that the cable displacement is regulated exponentially fast in the following sense,*

$$|u(s, t)| \le \sqrt{\frac{2\lambda_2}{\lambda_1 p}\kappa_o \exp\left(-\frac{\lambda_3}{\lambda_2}t\right)} \quad \forall s \in [0, 1], \tag{12}$$

where λ_1, λ_2, and λ_3 are positive bounding constants, the positive constant κ_o is given by

$$\kappa_o = \frac{1}{2}\int_0^1 u_t^2(\sigma, 0)d\sigma + \frac{1}{2}\int_0^1 p(\sigma)u_\sigma^2(\sigma, 0)d\sigma + (u_t(1, 0) + u_s(1, 0))^2, \tag{13}$$

if the controller gain k_s defined in (11) satisfies

$$k_s > \frac{p(1)}{2}. \tag{14}$$

Proof. To facilitate the stability proof, we first define an auxiliary tracking signal

$$\eta(t) = u_t(1, t) + u_s(1, t). \tag{15}$$

After differentiating (15) with respect to time, multiplying the resulting expression by m, and then substituting the right-hand side of (8) for $mu_{tt}(1, t)$, we have

$$m\dot{\eta} = mu_{st}(1, t) - p(1)u_s(1, t) - Y(u_t(1, t))\phi + f, \tag{16}$$

where $\dot{\eta} = \eta_t$. After substituting (11) into (16), we obtain the following closed-loop dynamics on the boundary:

$$m\dot{\eta} = -k_s\eta. \tag{17}$$

Remark 1. The following inequalities are used in the proof. First, the boundary condition Eq. (7) allows formulation of (see Hardy et al., 1959)

$$u^2(s, t) \le \int_0^1 u_\sigma^2(\sigma, t)d\sigma \quad \forall s \in [0, 1]. \tag{18}$$

Second, using the triangle inequality we have

$$-(z^2 + y^2) \le |zy| \quad z^2 + y^2 \ge zy \quad \forall z, y \in \Re. \tag{19}$$

To prove Eq. (12), we begin with the following function:

$$V(t) = E_s(t) + \tfrac{1}{2}m\eta^2(t) + E_c(t). \tag{20}$$

The cable vibration energy is

$$E_s(t) = \frac{1}{2}\int_0^1 u_t^2(\sigma, t)d\sigma + \frac{1}{2}\int_0^1 p(\sigma)u_\sigma^2(\sigma, t)d\sigma, \tag{21}$$

and the crossing term is

$$E_c(t) = 2\beta\int_0^1 \gamma(\sigma)u_t(\sigma, t)u_\sigma(\sigma, t)d\sigma, \tag{22}$$

where β is a positive design constant and the weighting function

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\gamma(s) = \frac{h(s)}{2p_0}\left[\arctan\left(\frac{2s-1}{2p_0}\right) + p_0 \ln\left(2s - 1 + h(s)\right) + \arctan\frac{1}{2p_0} + p_0 \ln\left(\sqrt{4p_0^2+1} - 1\right)\right] \quad (23)$$

with

$$h(s) = \sqrt{4p_0^2 + 4s^2 - 4s + 1}.$$

If the design constant β is selected according to

$$\beta < \frac{\min\{1, p_0\}}{4\gamma(1)}, \quad (24)$$

we can formulate the following upper and lower bound on $V(t)$ (see Lemma A.1 in the Appendix)

$$\lambda_1(E_s(t) + \eta^2(t)) \leq V(t) \leq \lambda_2(E_s(t) + \eta^2(t)), \quad (25)$$

where the positive constants λ_1 and λ_2 are

$$\lambda_1 = \min\left\{1 - \frac{4\beta\gamma(1)}{\min\{1, p_0\}}, \frac{m}{2}\right\},$$

$$\lambda_2 = \max\left\{1 + \frac{4\beta\gamma(1)}{\min\{1, p_0\}}, \frac{m}{2}\right\}. \quad (26)$$

After differentiating (20) with respect to time and substituting (6), (7), and (17), we obtain the following upper bound on the time derivative of $V(t)$ (See Lemma A.2 in the Appendix):

$$\dot{V}(t) \leq -\lambda_3(E_s(t) + \eta^2(t)), \quad (27)$$

where the positive constant[3] λ_3 is defined by

$$\lambda_3 = \min\left\{2\beta, k_s - \frac{p(1)}{2}\right\}. \quad (28)$$

Remark 2. The structure of the weighting function $\gamma(s)$ defined in (23) has been crafted to facilitate the construction of the inequalities given in (25) and (27).

From (25) and (27), we obtain the following upper bound for the time derivative of $V(t)$

$$\dot{V}(t) \leq -\frac{\lambda_3}{\lambda_2} V(t), \quad (29)$$

whose solution yields

$$V(t) \leq V(0)\exp\left(-\frac{\lambda_3}{\lambda_2}t\right) \leq \lambda_2\kappa_o \exp\left(-\frac{\lambda_3}{\lambda_2}t\right), \quad (30)$$

where Eq. (25) has been utilized to formulate the inequality on the right-hand side of Eq. (30) and the positive constant κ_o is defined in Eq. (13). In addition, we use Eqs. (18), (21), and (25) to formulate the following inequality:

$$\frac{p_0}{2}u^2(x, t) \leq \frac{1}{2}\int_0^1 p(\sigma)u_\sigma^2(\sigma, t)d\sigma \leq E_s(t) \leq \frac{1}{\lambda_1}V(t) \quad \forall x \in [0, 1]. \quad (31)$$

The inequality given in Eq. (12) now directly follows by combining Eqs. (30) and (31) and then using Eqs. (15) and (21). □

Remark 3. Since the proposed control strategies are relatively simple, smooth functions, we have assumed existence of solution for the dynamics given by (6) through (8) under the proposed control. To illustrate bounded-input, bounded-output stability for the closed-loop system, we invoke the following reasonable assumption: The distributed variable $u(s, t)$, and its time derivative $u_t(s, t)$, belong to a space of functions which has the following properties: (i) if $u(s, t)$ is bounded $\forall t \in [0, \infty)$ and $\forall s \in [0, 1]$, then $u_s(s, t)$ and $u_{ss}(s, t)$ are bounded $\forall t \in [0, \infty)$ and $\forall s \in [0, 1]$, and (ii) if the kinetic energy for the mechanical system of (6) through (8)

$$\frac{1}{2}\int_0^1 u_t^2(\sigma, t)d\sigma + \frac{1}{2}mu_t^2(1, t) \quad (32)$$

is bounded then $u_t(s, t)$ and $u_{st}(s, t)$ are bounded for $\forall t \in [0, \infty)$ and $\forall s \in [0, 1]$.

Remark 4. From (20) and (30), we can state that $E_s(t)$ and $\eta(t)$ are bounded $\forall t \in [0, \infty)$. Since $E_s(t)$ is bounded $\forall t \in [0, \infty)$, we can use (21), (18), and the properties discussed in Remark 3 to state that $u(s, t)$, $u_s(s, t)$, and $u_{ss}(s, t)$ are bounded $\forall t \in [0, \infty)$ and $\forall s \in [0, 1]$. Since $\eta(t)$ and $u_s(1, t)$ are bounded $\forall t \in [0, \infty)$, we can use (15) to state that $u_t(1, t)$ is bounded $\forall t \in [0, \infty)$; hence, the kinetic energy of the mechanical system defined in (32) is bounded $\forall t \in [0, \infty)$. Since the kinetic energy is bounded $\forall t \in [0, \infty)$, we can use the properties discussed in Remark 3 to state that $u_t(s, t)$ and $u_{st}(s, t)$ are bounded for $\forall t \in [0, \infty)$ and $\forall s \in [0, 1]$. From the above information, we can now state that all of the signals in the control of (11) and the mechanical system given by (6) through (8) remain bounded $\forall t \in [0, \infty)$ during closed-loop operation.

3.2 Adaptive Control Law. In this subsection, we redesign the exact model knowledge controller (11) to compensate for *constant* parametric uncertainty while asymptotically stabilizing the mechanical model given by (6) through (8). The adaptive control law is

$$f = -W\hat{\theta} - k_s\eta, \quad (33)$$

where $W(t) \in \Re^{1\times(q+2)}$ is a known regression matrix and $\hat{\theta}(t) \in \Re^{(q+2)\times 1}$ is a dynamic parameter estimate vector:

$$W = [u_{st}(1, t), -u_s(1, t), -Y(u_t(1, t))],$$

$$\hat{\theta} = [\hat{m}, \widehat{p(1)}, \hat{\phi}^T]^T. \quad (34)$$

The parameter estimate vector introduced in (33) is updated online according to

$$\dot{\hat{\theta}} = \Gamma W^T\eta, \quad (35)$$

where $\Gamma \in \Re^{(q+2)\times(q+2)}$ is a constant, diagonal gain matrix with elements $\Gamma_i > 0$.

THEOREM 2. *Given the field equation of (6) and the boundary conditions given by (7) and (8), the boundary controller given by (33) and (35) ensures that the cable displacement is regulated asymptotically fast in the following sense:*

$$\lim_{t\to\infty} |u(s, t)| = 0 \quad \forall s \in [0, 1], \quad (36)$$

where the controller gain k_s defined in (33) must be selected to satisfy (14).

Proof. After differentiating (15) with respect to time, multiplying the resulting expression by m, and then substituting the right-hand side of (8) for $mu_{tt}(L, t)$, we have

$$m\dot{\eta} = W\theta + f, \quad (37)$$

where $W(t)$ was defined in (34) and $\theta \in R^{(q+2)\times 1}$ is the unknown constant parameter vector defined as

[3] The controller gain k_s must satisfy (14) to ensure that λ_3 is positive.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

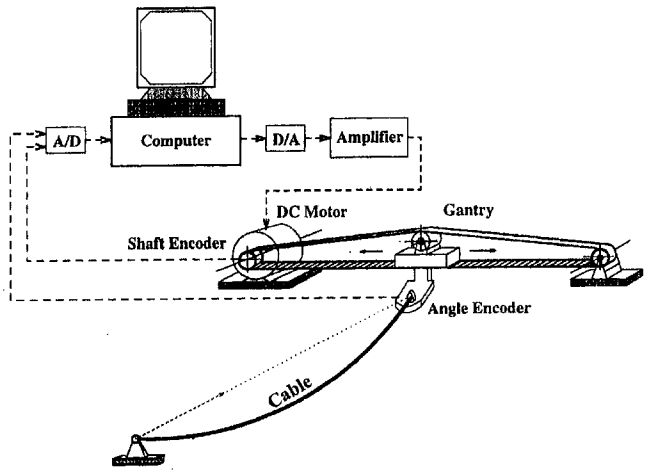


Fig. 2 Schematic diagram of the experimental setup

$$\theta = [m \quad p(1) \quad \phi^T]^T. \tag{38}$$

After substituting (33) into (37), we obtain the following closed-loop dynamics on the boundary:

$$m\dot{\eta} = -k_s\eta + W\tilde{\theta},$$
$$\dot{\tilde{\theta}} = -\Gamma W^T\eta, \tag{39}$$

where $\tilde{\theta}(t) = \theta - \hat{\theta}(t) \in \Re^{(q+2)\times 1}$ is the parameter estimate error vector, and (35) has been used to obtain the parameter estimation error dynamics.

To prove Eq. (36), we begin with the following function:

$$V_a(t) = V(t) + \tfrac{1}{2}\tilde{\theta}^T(t)\Gamma^{-1}\tilde{\theta}(t), \tag{40}$$

where $V(t)$ was defined in (20). If the design constant β of (22) is selected to be sufficiently small, Lemma A.1 can be used to show

$$\lambda_{1a}[E_s(t) + \eta^2(t) + \|\tilde{\theta}(t)\|^2]$$
$$\leq V_a(t) \leq \lambda_{2a}[E_s(t) + \eta^2(t) + \|\tilde{\theta}(t)\|^2], \tag{41}$$

where the positive constants λ_{1a} and λ_{2a} are defined by

$$\lambda_{1a} = \min\left\{1 - \frac{4\beta\gamma(1)}{\min\{1,\rho\}}, \frac{m}{2}, \frac{1}{2}\min\{1/\Gamma_i\}\right\},$$

$$\lambda_{2a} = \max\left\{1 + \frac{4\beta\gamma(1)}{\max\{1,\rho\}}, \frac{m}{2}, \frac{1}{2}\max\{1/\Gamma_i\}\right\}.$$

After differentiating (40) with respect to time, substituting (6), (7), and (39), and using Lemma A.2, we obtain the following upper bound for the time derivative of $V_a(t)$:

$$\dot{V}_a(t) \leq -\lambda_3(E_s(t) + \eta^2(t)) \triangleq g_a(t), \tag{42}$$

where λ_3 was defined in (28). We can now use (41), (42), and a similar argument as that outlined in Remark 3 to state that all of the signals in the control of (33) and (35) are bounded and that all of the signals in the mechanical system given by (6) through (8) remain bounded.

After differentiating the right-hand side of (42) with respect to time, we have

$$\dot{g}_a(t) = -\lambda_3(\dot{E}_s(t) + 2\eta(t)\dot{\eta}(t)). \tag{43}$$

Since we illustrated how all of the system signals remain bounded $\forall t \in [0, \infty)$, we can use (53) and (39) to show that $\dot{E}_s(t)$ and $\dot{\eta}(t)$ are bounded $\forall t \in [0, \infty)$; hence, we can see from (43) that the time derivative of the right-hand side of (42) is also bounded $\forall t \in [0, \infty)$. We can now invoke Lemma A.3 in the Appendix to the right-hand side of (42) to show that

$$\lim_{t\to\infty} E_s(t), \eta(t) = 0. \tag{44}$$

We can now use the result given by (44) and the inequality-type bound developed in (31) to state the result given by (36). □

Remark 5. It should be noted that several extensions to the basic controllers given in the previous sections can be devel-

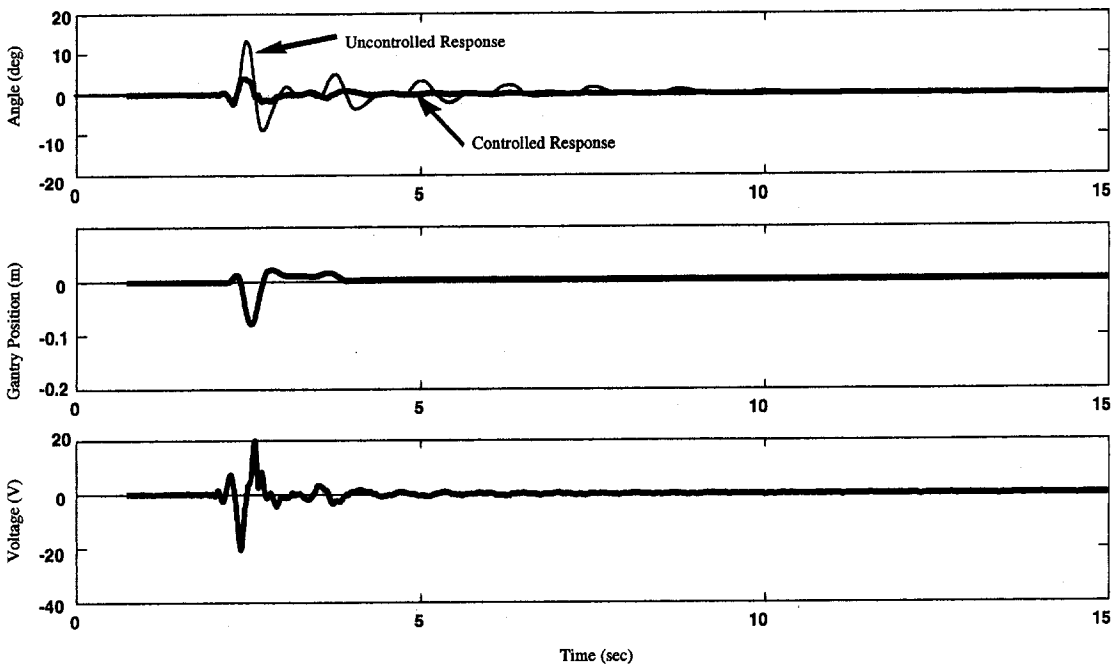


Fig. 3 Experimental results—exact model knowledge controller

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

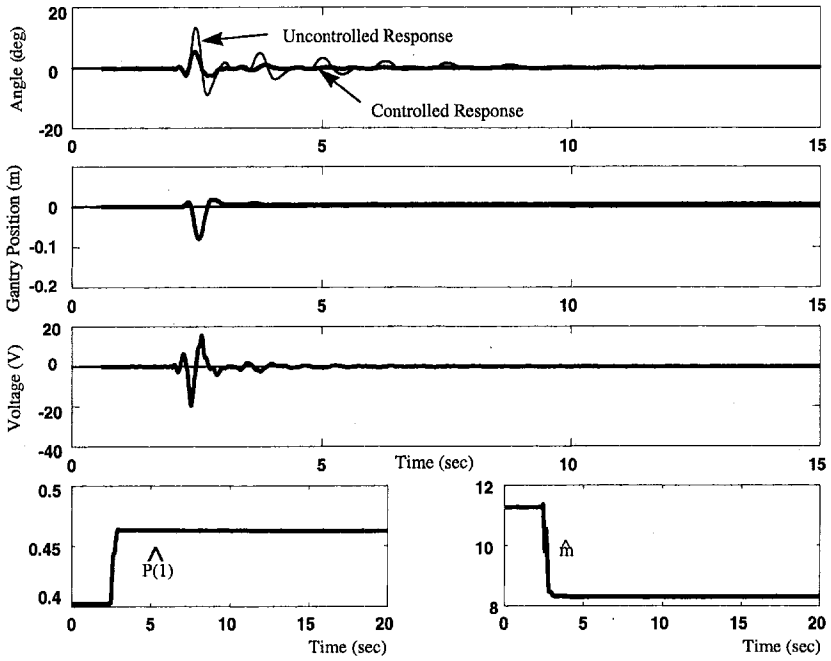


Fig. 4 Experimental results—adaptive controller

oped. For example, an additional gain constant in the definition of the tracking variable $\eta(t)$ defined in (15) can be added. Specifically, we can redefine the exact model knowledge input control force originally defined in (11) as follows:

$$f = -m\alpha u_{st}(1, t) + p(1)u_s(1, t) - k_s\eta + Y(u_t(1, t))\phi, \quad (45)$$

where α is a positive controller gain, and the variable $\eta(t)$ originally defined in (15) is now defined as follows:

$$\eta = u_t(1, t) + \alpha u_s(1, t). \quad (46)$$

As illustrated in Canbolat et al. (1997) for a similar mechanical system, we can slightly modify the arguments outlined in Remark 3 and the proof of Theorem 1 to show that all of the signals in the control and the mechanical system remain bounded and to state a similar exponential stability result as that given by (12). By following the adaptive control technique given in the previous section, we can also redesign the controller given by (45) as an adaptive controller which compensates for parametric uncertainty and achieves asymptotic displacement regulation.

4 Experimental Results

The proposed controller was implemented on a cable control system designed and built in-house (see Fig. 2). A braided polyester rope, pinned at one end, and connected to a horizontally translating gantry at the other end was used for the experiments. A brushed DC motor (Baldor model 3300) was coupled to the gantry using a timing belt. Two 1000-count rotary encoders (Hohner) were used to measure the gantry position and the cable departure angle. A mounting bracket attached to the gantry aligned the departure angle encoder axis with the normal (in-plane) direction. Thus, the measured encoder rotation caused by cable motion corresponded to out-of-plane slope at the controlled boundary $u_s(1, t)$. A 486 ISA-based PC hosting a Texas Instruments TMS320C30 digital signal processing board served as the computational engine. An encoder interface card (Integrated Motions, Inc., Model DS-2) allowed for quadrature extrapolation of the encoder signals. The DS2 board also supported two channels of 16-bit ADCs and DACs. The linear and angular velocities were obtained by applying a backwards difference algorithm to the position and angle signals, respectively. To eliminate quantization noise, the velocity signals were filtered using a second-order digital filter. To test the response of the proposed controller, the cable was perturbed using a repeatable displacement input disturbance near the pinned end. As defined in (5), the parameter values for the mechanical system were determined via standard test procedures to be as follows:

$$M = 3.229 \text{ kg} \quad L = 2.69 \text{ m} \quad \rho A = 0.085 \text{ kg/m}$$
$$g = 9.81 \text{ m/s}^2 \quad P_0 = 0.127 \text{ N}$$

$$m = 14.1 \quad p_0 = 0.056 \quad p(1) = 0.503$$

Two experiments were conducted to test the performance of the proposed controller. First, the exact model knowledge controller given by (45) was implemented using the following settings:

$$k_s = 5.0 \quad \alpha = 4.0 \quad Y(u_t(1, t))\phi = 0.$$

The results of these experiments appear in Fig. 3. As is clear from the figure, the cable system exhibits excellent transient response under the proposed controller. The uncontrolled swinging of the cable to the same input is shown for comparison purposes. The gantry position and the voltage signals are also shown. It should also be noted that the $Y(u_t(1, t))\phi$ term could have been used to model static friction on the rail of the actuator; however, the feedback portion of the controller seemed to adequately compensate for friction in our experiment.

The adaptive version of (45) was implemented with the update law given by (35) and the parameter estimates initialized to 80 percent of their nominal values. The best performance was achieved using the following settings:

$$k_s = 5.0 \quad \alpha = 4.0 \quad Y(u_t(1, t))\phi = 0 \quad \Gamma = \text{diag}\{5, 2\}.$$

The results of these experiments appear in Fig. 4. In the first subplot of Fig. 4, the transient performance of the adaptive controller to the disturbance is compared to the uncontrolled swinging of the cable for the same input. Subsequent subplots show the gantry position and the voltage signals. As is clear

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

from the figure, the estimates for the parameters m and $p(1)$ remain bounded during closed-loop operation.

Remark 6. The input control force at the actuator is actually applied by rescaling this force as a desired motor torque. The desired motor torque is then achieved by using a high-gain current feedback loop (recall for the brushed DC motor that $\tau_{\text{motor}} \propto I$ where I is the motor current). The current feedback loop and the torque-force conversion are incorporated into the control software and power is supplied to the motor by the single-channel linear power amplifier which is capable of outputting up to 1000 W at 100 V with a power bandwidth of 0 to 40 KHz. It should be noted that the backstepping paradigm (Kokotovic, 1992) allows electrical dynamics to be incorporated in the overall control solution for this class of hybrid partial/ordinary differential equation problems as illustrated in Baicu et al. (1996b).

Remark 7. Dimensional quantities are used in the experiment so the control force is scaled according to (5) (i.e., $F = f\rho AgL$). This implies that the weight of the cable must be known in order to implement the controllers.

5 Conclusion

In this paper, we develop new boundary control strategies for a flexible cable with actuator dynamics. The control structure differs substantially from the controller in Baicu et al. (1996a) because different signals are sensed.[4] While the controller given in Baicu et al. (1996a) only provides for bounded-input, bounded-output stability, the proposed exact model knowledge controller exponentially stabilizes the position of the cable and the adaptive controller asymptotically stabilizes the position of the cable and compensates for parametric uncertainty in the cable tension and actuator mass. The experimental results are similar to those of Baicu et al. (1996a), demonstrating fast transient decay under the proposed control.

Acknowledgment

This work is supported in part by the U.S. National Science Foundation Grants DMI-9457967, DDM-931133269, DMI-9622220, the Office of Naval Research URI-3139-YIP01, the Square D Corporation, the Union Camp Corporation, the AT& T Foundation, and Mersin University, Mersin, Turkey.

References

Baicu, C. F., Rahn, C. D., and Nibali, B. D., 1996a, "Active Boundary Control of Elastic Cables: Theory and Experiment," *Journal of Sound and Vibration,* Vol. 198, No. 1, pp. 17–26.

Baicu, C. F., Rahn, C. D., and Dawson, D., 1996b, "Boundary Control of a Flexible Link Electrically Driven Gantry Robot," *Active Control of Vibration and Noise,* ASME, New York, pp. 109–115.

Canbolat, H., Dawson, D., and Rahn, C., 1997, "Boundary Control of a Cantilevered Flexible Beam with Point-Mass Dynamics at the Free End," *Proc. of the 38th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and the AIAA/ASME/AHS Adaptive Structures Forum,* Part 2, Kissimmee, FL, Apr., pp. 1589–1598.

Chen, G., Delfour, M., Krall, A., and Payre, G., 1987, "Modeling, Stabilization, and Control of Serially Connected Beams," *SIAM Journal of Control and Optimization,* Vol. 25, No. 3, pp. 526–546.

Fried, I., 1982, "Large Deformation Static and Dynamic Finite Element Analysis of Extensible Cables," *Computers and Structures,* Vol. 15, pp. 315–319.

Fujino, Y., Warnitchai, P., and Pacheco, B., 1993, "Active Stiffness Control of Cable Vibration," ASME JOURNAL OF APPLIED MECHANICS, Vol. 60, pp. 948–953.

Fujino, Y., and Susumpow, T., 1995, "Active Control of Multimodal Cable Vibrations by Axial Support Motion," *Smart Materials and Structures,* Vol. 4, pp. A41–A51.

Hardy, G., Littlewood, J., and Polya, G., 1959, *Inequalities,* Cambridge University Press, Cambridge, UK.

Irvine, H., 1981, *Cable Structures,* MIT Press, Cambridge, MA.

Kokotovic, P., 1992, "The Joy of Feedback: Nonlinear and Adaptive," *IEEE Control Systems Magazine,* Vol. 12, June, pp. 7–17.

Luo, Z., Kitamura, N., and Guo, B., 1995, "Shear Force Feedback Control of Flexible Robot Arms," *IEEE Transactions on Robotics and Automation,* Vol. 11, No. 5, pp. 760–765.

Morgül, Ö., 1991, "Orientation and Stabilization of a Flexible Beam Attached to a Rigid Body: Planar Motion," *IEEE Transactions on Automatic Control,* Vol. 36, No. 8, pp. 953–962.

Morgül, Ö., Rao, B., and Conrad, F., 1994, "On the Stabilization of a Cable with a Tip Mass," *IEEE Transactions on Automatic Control,* Vol. 39, No. 10, pp. 2140–2145.

Perkins, N., and Mote, C., 1987, "Three-dimensional Vibration of Travelling Elastic Cables," *Journal of Sound and Vibration,* Vol. 114, No. 2, pp. 325–340.

Rahn, C., and Mote, C., 1993, "Axial Force Stabilization of Transverse Beam Vibration," *Vibration and Control of Mechanical Systems,* ASME, New York, pp. 29–34.

Shahruz, S. M., and Krishna, L. G., 1996, "Boundary Control of Nonlinear String," *Proc. of the ASME Dynamics Systems and Control Division,* ASME, New York, pp. 831–835.

Slotine, J., and Li, W., 1991, *Applied Nonlinear Control,* Prentice-Hall, Englewood Cliffs, NJ.

Soler, A., 1970, "Dynamic Response of Single Cable with Initial Sag," *Journal of the Franklin Institute,* Vol. 190, pp. 377–387.

West, H., Geschwinder, L., and Suchoski, J., 1975, "Natural Vibrations of Suspension Cables," *Journal of the Structural Division, ASCE,* Vol. 101, pp. 2277–2291.

APPENDIX

Stability Lemmas

LEMMA A.1. *The non-negative function given by (20) can be upper and lower bounded as given by (25).*

Proof. First, note that (19) can be used to bound $E_c(t)$ of (22) as follows:

$$E_c = 2\beta \int_0^1 \gamma(\sigma) u_t u_\sigma d\sigma \leq 2\beta\gamma(1) \int_0^1 (u_t^2 + u_\sigma^2) d\sigma,$$

$$\leq \frac{4\beta\gamma(1)}{\min(1, p_0)} \left[\frac{1}{2} \int_0^1 (u_t^2 + p(\sigma) u_\sigma^2) d\sigma \right], \tag{47}$$

where we have used the facts that $p(s)$ is minimum at $s = \frac{1}{2}$ for $s \in [0, 1]$ and that $\gamma(s)$ of (23) is maximum at $s = 1$ for $s \in [0, 1]$. We can now use (21) and (47) to establish the following inequality:

$$-\frac{4\beta\gamma(1)}{\min(1, p_0)} E_s \leq E_c \leq \frac{4\beta\gamma(1)}{\min(1, p_0)} E_s; \tag{48}$$

hence, if β is selected according to (24), we have

$$0 \leq \left(1 - \frac{4\beta\gamma(1)}{\min\{1, p_0\}}\right) E_s \leq E_s + E_c$$

$$\leq \left(1 + \frac{4\beta\gamma(1)}{\min\{1, p_0\}}\right) E_s. \tag{49}$$

Given the structure of $V(t)$ defined in (20) and the inequality given by (49), the inequality given by (25) is obvious.

LEMMA A.2. *The time derivative of the nonnegative function given by (20) can be upper bounded as given by (27).*

Proof. After differentiating (20) with respect to time, we have

$$\dot{V} = \dot{E}_s + \dot{E}_c - k_s \eta^2, \tag{50}$$

where (17) has been used. To determine $\dot{E}_s(t)$ in (50), we differentiate (21) with respect to time to obtain

$$\dot{E}_s = \int_0^1 u_t [p(\sigma) u_\sigma]_\sigma d\sigma + \int_0^1 p(\sigma) u_\sigma u_{\sigma t} d\sigma, \tag{51}$$

where (6) has been used. If we integrate the first term on the right-hand side of (51) by parts, we obtain

$$\dot{E}_s = p(1) u_t(1, t) u_s(1, t) - p(0) u_t(0, t) u_s(0, t). \tag{52}$$

After applying the boundary conditions given in (7) to (52), we have

$$\dot{E}_s = p(1) u_t(1, t) u_s(1, t), \tag{53}$$

[4] The controller given in Baicu et al. (1996a) requires measurement of boundary position, velocity, and the slope while the proposed controller requires measurement of the boundary velocity, slope, and slope rate.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

which can be written as

$$\dot{E}_s = -\frac{p(1)}{2}(u_t^2(1,t) + u_s^2(1,t)) + \frac{p(1)}{2}\eta^2 \quad (54)$$

upon application of (15).

To determine $\dot{E}_c(t)$ in (50), we differentiate (22) with respect to time as follows:

$$\dot{E}_c = A_1 + A_2, \quad (55)$$

where

$$A_1 = 2\beta \int_0^1 \gamma(\sigma) u_t u_{\sigma t} d\sigma,$$

$$A_2 = 2\beta \int_0^1 \gamma(\sigma) u_\sigma [p(\sigma) u_\sigma]_\sigma d\sigma, \quad (56)$$

after using (6) in the expression for A_2. After integrating the expression for A_1 given by (56) by parts, we obtain

$$A_1 = 2\beta \left(\gamma(1) u_t^2(1,t) - \int_0^1 \gamma_\sigma(\sigma) u_t^2 \, d\sigma \right) - 2\beta \int_0^1 \gamma(\sigma) u_{\sigma t} u_t d\sigma, \quad (57)$$

where we have used the fact that $\gamma(0) = 0$. After noting that the last term in (57) is equal to A_1, we can write (57) as follows:

$$A_1 = \beta \left(\gamma(1) u_t^2(1,t) - \int_0^1 \gamma_\sigma(\sigma) u_t^2 \, d\sigma \right), \quad (58)$$

where $\gamma_s(s)$ is explicitly given by

$$\gamma_s(s) = \frac{8s-4}{4p_0 h(s)} \left[\arctan\left(\frac{2s-1}{2p_0}\right) + p_0 \ln(2s - 1 + h(s)) - \arctan\frac{1}{2p_0} + p_0 \ln(\sqrt{4p_0^2 + 1} - 1) \right] + \frac{h(s)}{2p_0} \left(\frac{1}{p_0\left(1 + \frac{(2s-1)^2}{4p_0^2}\right)} + \frac{p_0\left(2 + \frac{8s-4}{2h(s)}\right)}{2s - 1 + h(s)} \right).$$

It can be shown that $\gamma_s(s) \geq 1$ for $s \in [0, 1]$; hence, we can utilize (58) to construct the following upper bound for A_1:

$$A_1 \leq \beta \left(\gamma(1) u_t^2(1,t) - \int_0^1 u_t^2 \, d\sigma \right). \quad (59)$$

After integrating the expression for A_2 given by (56) by parts, we obtain

$$A_2 = 2\beta \left(\gamma(1) p(1) u_s^2(1,t) - \int_0^1 \gamma_\sigma(\sigma) p(\sigma) u_\sigma^2 \, d\sigma - \int_0^1 \gamma(\sigma) p(\sigma) u_\sigma u_{\sigma\sigma} d\sigma \right), \quad (60)$$

where we have used the fact that $\gamma(0) = 0$. After noting that the expression for A_2 given by (56) can be expanded into the following form,

$$A_2 = 2\beta \left(\int_0^1 \gamma(\sigma) u_\sigma^2 p_\sigma(\sigma) d\sigma + \int_0^1 \gamma(\sigma) p(\sigma) u_\sigma u_{\sigma\sigma} d\sigma \right), \quad (61)$$

we can combine (60) and (61) to eliminate the last term in (60) as follows:

$$A_2 = \beta \left(\gamma(1) p(1) u_s^2(1,t) - \int_0^1 [\gamma_\sigma(\sigma) p(\sigma) - \gamma(\sigma) p_\sigma(\sigma)] u_\sigma^2 d\sigma \right). \quad (62)$$

From the structure of $\gamma(s)$ and $p(s)$, it is straightforward to show that

$$\gamma_s(s) p(s) - \gamma(s) p_s(s) = 1 + p(s); \quad (63)$$

hence, we can use (62) and (63) to construct the following upper bound for A_2:

$$A_2 \leq \beta \left(\gamma(1) p(1) u_s^2(1,t) - \int_0^1 p(\sigma) u_\sigma^2 \, d\sigma \right). \quad (64)$$

After substituting (59) and (64) into (55) and then substituting the resulting expression along with (54) into (50), we have

$$\dot{V} \leq -\left(\frac{p(1)}{2} - \beta\gamma(1) \right) u_t^2(1,t) - \left(\frac{p(1)}{2} - \beta\gamma(1) p(1) \right) u_s^2(1,t)) - \left(k_s - \frac{p(1)}{2} \right) \eta^2 - 2\beta E_s, \quad (65)$$

where (21) has been utilized. From (65) and (24), it is clear that if the controller gain k_s is selected according to (14) and the design constant β is selected according to

$$\beta < \min\left\{ \frac{p(1)}{2\gamma(1)}, \frac{1}{2\gamma(1)}, \frac{\min\{1, p\}}{4\gamma(1)} \right\}, \quad (66)$$

then $\dot{V}(t)$ can be upper bounded by the nonpositive scalar function given in (27).

LEMMA A.3. *If (i) $V_a(t)$ is a non-negative, scalar function which is lower bounded by zero, (ii) $\dot{V}_a(t) \leq -f(t)$ where $f(t)$ is a scalar, non-negative function, and (iii) $\dot{f}(t)$ is bounded then*

$$\lim_{t\to\infty} f(t) = 0. \quad (67)$$

Proof. First, we define the following function:

$$V_n(t) = V_a(t) - \int_0^t (\dot{V}_a(\tau) + f(\tau)) d\tau, \quad (68)$$

which is lower bounded by zero (since we have assumed that $V_a(t) \geq 0$ and that $\dot{V}_a(t) \leq -f(t)$, we know that $V_n(t) \geq 0$). If we differentiate (68) with respect to time, we obtain

$$\dot{V}_n(t) = -f(t). \quad (69)$$

We now apply a lemma from Slotine and Li (1991) (see page 127) which states that if: (i) $V_n(t)$ is a non-negative, scalar function which is lower bounded by zero, (ii) $\dot{V}_n(t) = -f(t)$ where $f(t)$ is a scalar, non-negative function, and (iii) $\dot{f}(t)$ is bounded then

$$\lim_{t\to\infty} f(t) = 0. \quad (70)$$

Application of the above lemma to (68) and (69) yields the result given by (67).

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Anisotropic Constraint Effects on Interactive Damage Progression of General Lay-up Composites

J. Fan[1]

J. Zhang

Department of Engineering Mechanics,
Chongqing University,
Chongqing, P.R. China

In this paper, a five-layer model is proposed to analyze the anisotropic constraint effects on interactive damage progression. Incorporating this model with the general theory of overall moduli of a cracked body, the effective damage factors proposed by Fan and Zhang in 1993 have been extended as quantitative measures for characterizing the anisotropic constraint effects on damage-induced stiffness reduction of general lay-up composites. By applying the first-order shear deformation plate theory to each sublayer these factors are determined and applied for investigating progressive matrix cracking-delamination interaction. Results show that by this rather unique approach, the intractable problem of anisotropic constraint effects of general lay-up composites on interactive damage progression can be handled in a systematic manner.

1 Introduction

Polymeric matrix composites develop matrix cracking parallel to the fiber directions of nonaligned plies under fairly low levels of applied loads. The density of the cracks increases with load or with the number of fatigue cycles until reaching a saturation level denoted as "characteristic damage state" (CDS). Subsequently, the laminates develop interlaminar delaminations and fiber breaks of aligned plies, which lead to ultimate failure. Current models of progressive failure in polymeric matrix composites have predicted well up to, but not beyond, the CDS level. In most cases, analyses were confined to cross-ply composites. However, since delaminations are much more crucial to the integrity of composite laminates than the transverse matrix cracks considered in the CDS analysis, the prediction of their initiation and growth is much more relevant for reliable designs. On the other hand, since cross-ply laminates are of limited use in engineering applications, damage models for laminates of general lay-ups are technically important.

The predominant characteristics of damage progression in laminated composites are the important anisotropic constraint effects. By definition, these effects are internal constraint effects which exist between different layers and, in general, they are not equal along different directions. Parvizi, Garrett, and Bailey (1978) reported that when the thickness of the 90-deg ply in a cross-ply laminate is reduced to a certain value, laminate failure occurred prior to the development of any transverse matrix cracks in the 90-deg plies. These experimental results imply that the relatively stronger anisotropic constraint effects on the thinner 90-deg ply group, when compared with a thicker one, control the damage pattern and failure mode of the laminate. Similar effects were noted by Crossman and Wang (1982) and Wang (1984) for T300/934 graphite/epoxy $[\pm 25/90_n]_s$ laminates, again highlighting the significance of constraint effects. The latter authors noted that additional matrix cracks developed in the 90-deg plies in conjunction with the growth of edge delaminations. This interaction between matrix cracks and edge delaminations was found to be strongly dependent on the ply number, n, of the 90-deg ply group.

Bailey, Curtis, and Parvizi (1979), and Highsmith and Reifsnider (1979) used shear-lag models to investigate stiffness reductions due to transverse matrix cracking of cross-ply laminates. The constraint effects on damage progression of 90-deg plies were quantified under the assumption that the applied stress is transferred back into the 90-deg plies from the constraining layers through shear deformation. Flaggs and Kural (1982) reported that the highest in-situ transverse strength of a lamina in laminates $[\pm\theta_m/90°_n]_s$ of their test is 2.48 times the corresponding value when measured by unidirectional laminate specimens. Flaggs (1985) then developed a two-dimensional shear-lag model for the constraint effects on the first-ply failure strain of 90-deg plies and obtained reasonable agreement with the experiment. Progressive matrix cracking was, however, not considered. From the results of Flaggs and Kural, we can see that lamina strength is not strictly an intrinsic material property. Instead, the intrinsic strength of a lamina not only depends on the material property itself, but also the defect population as well as its constraint conditions. The latter is closely related to the stacking sequence. Also, in a unidirectional laminate the first failure causes sample failure, but in a multidirectional laminate with alternative load paths, first failure does not necessarily cause the sample to break.

Some further improvement of shear-lag analysis was done by incorporating statistical analysis (Wang, et al., 1984; Laws, et al., 1988; Fukunaga, et al., 1984) and by extending to biaxial loading (Lee, et al., 1989; Tsai, et al., 1990). There are other damage models which implicitly account for the anisotropic constraint effects, such as a variational approach (Hashin, 1985; Nairn and Hu, 1992), a self-consistent scheme (Dvorak et al., 1985) and an approximate elasticity approach (Nuismer and Tan, 1988, Lee and Daniel, 1990). However, due to mathematical difficulties most of them are limited to cross-ply laminates. The internal variable formulations proposed by Talreja (1985), and Allan et al. (1987) need additional tests to determine various parameters employed therein. This requirement and the inherent difficulty in formulating damage evolution laws limit their applications. Gudmundson and Zhang (1993) used known analytical solutions, such as the solution for a row of cracks in an infinite isotropic medium, for damage analysis, but these analytical solutions are of limited applicability to composite laminates.

[1] Present address: The Georgia W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405. Mem. ASME.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Apr. 22, 1996; final revision, Aug. 14, 1997. Associate Technical Editor: I. M. Daniel.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

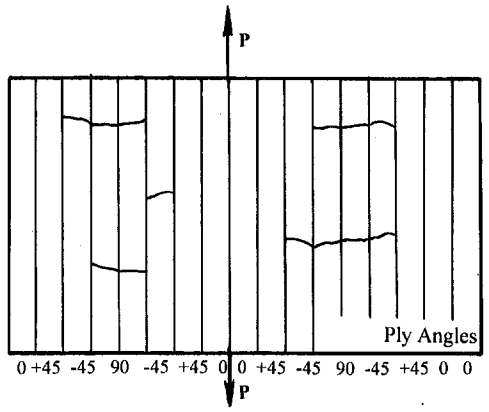


Fig. 1 Schematic description of damage pattern for a $[0/45/-45/90]_{3s}$ laminate of AS4/3501-6 Gr/Ep composites (referred from Timmer and Hahn, 1993)

An experimental report given by Timmer and Hahn (1993) has stimulated more interests in investigating the anisotropic constraint effects for general lay-up composites. Their results are schematically shown in Fig. 1. The tests were conducted on a quasi-isotropic $[0° \pm 45°/90°]_{s3}$ laminate of AS4/3501-6 graphite/epoxy composites under tensile loading along the longitudinal direction. Matrix cracking in the 90-deg layers occurred first, followed by matrix cracking in the −45-deg layers, but there were no matrix cracks at all in the 45-deg layers. To find the reason for this phenomenon, the constraint conditions of 45-deg layers and −45-deg layers were checked. From Fig. 1, we can see that the 45-deg layer and −45-deg layer were directly constrained by the strongest 0-deg layers and the weakest 90-deg layers, respectively. The much stronger anisotropic constraint effects on the 45-deg layers offer much larger resistance to damage initiation than that on the −45-deg-layers, and thus prevented damage in the 45-deg layers.

Choi, Downs, and Chang (1991a, b) conducted low-velocity impact experiments for investigating damage initiation in beams of T300/976 graphite/epoxy composites. The damage pattern in the composite specimens consists of matrix cracking and delaminations. Their experiments confirmed that the threshold level of impact energy for the initiation of matrix cracking as well as the amount of energy necessary to cause delaminations depended very strongly on the laminate's stacking sequence. Specific results taken from their test curves (Fig. 13 of Choi et al., 1991a) are listed in Table 1. This is a significant experimental program since it shows that the impact resistance of composite beams can be much improved by changing their lay-ups. However, their model using an empirical relation overpredicts the test results for some cases, while for other cases it underpredicts (see Table 2 of Choi et al., 1991b). There was no definite trend.

The above-mentioned situation indicates that there are existing needs in developing effective damage models which can account for the anisotropic constraint effects on damage initiation and progression. In a series of papers (1992, 1993), Fan and Zhang and their co-workers proposed a three-layer model to consider the anisotropic constraint effects on damage progression. In that model, all plies above and below the 90-deg layers are lumped into two equivalent constraint layers, an upper one and a lower one (Fan and Zhang, 1993). A two-dimensional shear-lag analysis with constant shear-lag coefficients is then applied for damage analysis of cross ply and $[\pm\theta_m/90_n]_s$ lay-ups using the concept of effective damage functions (or factors, see later). This original work was improved later by the work of Zhang, Fan and Soutis (1992a)[1] and Zhang, Soutis and Fan (1994a) where the out-of-plane shear stresses are assumed to vary linearly across the thickness of the cracked ply group and partially linear in the constraint layers. Applications and verifications of the three-layer model include the in-plane stiffness properties, development of transverse ply cracks, effects of matrix cracking and hygrothermal stress on the strain energy release rate for edge delamination, as well as energy release rate associated with local delamination (see also Zhang, Fan, and Soutis, 1992b; Zhang, Soutis, and Fan, 1994b). The three-layer model was further improved by Thomas (1995) for the cases of unsymmetric lay-ups in which quadratic displacement field is assumed for each of the arbitrary three layers. Numerical results for several lay-ups show that Thomas's approach has reasonable accuracy. On the other hand, as mentioned by Thomas the solution of the governing differential equation limits its application so that "it is not applicable to laminates whose configuration is such that the undamaged layers are balanced with respect to the damaged layer." While the three-layer model is significant in investigating the constraint effects this model has difficulties explaining the experimental phenomena shown in Fig. 1 and Table 1. More specifically, that model cannot distinguish the essential contributions of adjacent layers from those contributed by other layers which are further away from the layer under consideration. This shortcoming of the three-layer model limits its applications for general lay-up composites.

In this work, a five-layer model is proposed for a generic damaged ply, k, of general lay-up composites. Incorporating this model with the general theory of overall moduli of a cracked body, the effective damage factors proposed by Fan and Zhang (1993) have been further extended to any arbitrary ply as quantitative measures for the anisotropic constraint effects on interactive damage progression of general lay-up composites. This new methodology is exemplified through investigating progressive interaction of matrix cracking and local delamination under tensile loading perpendicular to the fiber direction of a non-aligned ply.

2 A Five-Layer Model for a Generic Ply

A laminate Cartesian coordinate system, xyz, is shown in Fig. 2(a). The lay-up angle, $\theta^{(k)}$, which defines the direction of fibers in a nonaligned ply group, k, is measured counterclockwise from the x-axis. A ply coordinate system, $x_1x_2x_3$, is introduced for each ply such that x_1 and x_2-axes are parallel and perpendicular, respectively, to the fibers in the ply at hand.

The central problem of modeling general lay-up composites is to develop a method for damage analysis of a generic ply group, k. In this regard, two things are important to simplify this analysis. One is to develop an equivalent constraint model to replace the complicated laminate structure for investigating the anisotropic constraint effects. The other is to emphasize the important role of the ply coordinates. Actually, this generic ply

Table 1 Lay-up effects on damage initiation energy $E^{\text{initiation}}$ (data obtained from test curves of Fig. 13 of Choi, et al., 1991a)

Lay-up type	Material type	Lay-ups	$E^{\text{initiation}}$ (J/m)*
A	T300/976 Gr/Ep	$[0_3^\circ/90_3^\circ/0_3^\circ/90_{1.5}^\circ]_s$	415
B	T300/976 Gr/Ep	$[0_4^\circ/\pm45_4^\circ/90_2^\circ]_s$	190
C	T300/976 Gr/Ep	$[0_6^\circ/\pm45_4^\circ/90_3^\circ]_s$	50
D	T300/976 Gr/Ep	$[90_3^\circ/0_3^\circ/90_3^\circ]$	75

* Line loading is assumed to calculate the impact energy

[1] This work was submitted latter and published earlier than the original work by Fan and Zhang (1993).

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

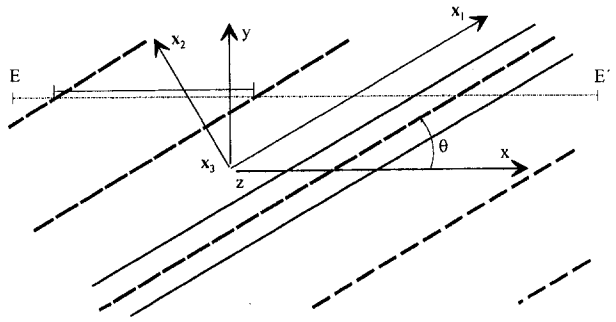


Fig. 2(*a*) A front view in the negative *z*-direction

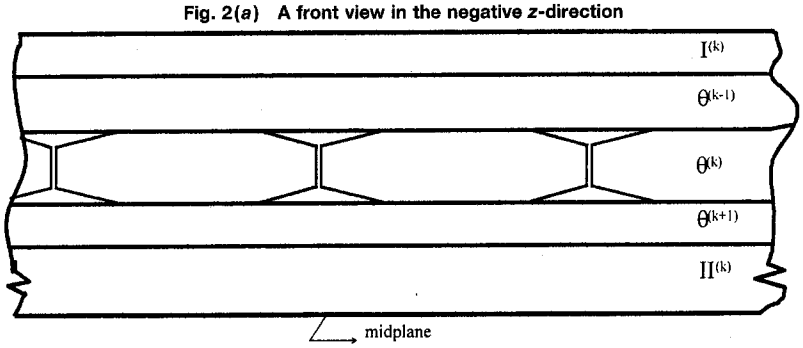


Fig. 2(*b*) An edge view at the section *E-E'*

Fig. 2 A five-layer model for the *k*th ply group within a general lay-up laminate

group, k, as any other ply can be viewed as "a local 90-deg layer" with respect to its ply coordinate system, $x_1x_2x_3$.

A scrutiny of the experimental results in Fig. 1 and Table 1 suggests the importance of the adjacent layer for the anisotropic constraint effects on damage in the ply under consideration. In addition, the anisotropic constraint effects decrease with increasing thickness of the weakest 90-deg ply group. Based on these results, the mechanism of anisotropic constraint effects may be proposed as follows: The neighboring layers of the generic layer, k, provide constraint forces through the contact surfaces between the adjacent layers and the layer at hand. These constraint forces change the stress state in the generic ply, k, and increase resistance to damage progression. Obviously, the constraint forces applied by the constraining layers increase with increase of their stiffness along the deformation direction, and decrease as the distance between the generic layer and constraining layers increases.

From the above experimental results and the proposed mechanism, it may be suitable to account for the anisotropic constraint effects on the generic damaged ply, k, in a way which can distinguish between the effects of adjacent layers and the effects of remaining layers. To emphasize the essential contributions of the adjacent layers, all their geometric and material parameters such as the thicknesses ($h^{(k-1)}$, $h^{(k+1)}$), ply angles ($\theta^{(k-1)}$, $\theta^{(k+1)}$), and in-plane stiffnesses ($\bar{Q}_{ij}^{(k-1)}$, $\bar{Q}_{ij}^{(k+1)}$) are explicitly considered in the five-layer model. However, the remaining layers will be handled in an approximate manner. Specifically, all remaining layers above and below the adjacent layers of ply groups $(k-1)$ and $(k+1)$, respectively, will be lumped into two equivalent homogeneous anisotropic layers, an upper group I_k and a lower group II_k. In general, this consideration results in a five-layer model for a generic damaged layer k. The five layers are: the kth ply group itself, two adjacent layers $(k-1)$ and $(k+1)$, and the remote equivalent layers I_k and II_k (see Fig. 2(b)). It may be noted that the five-layer model can be degenerated into a four-layer or a three-layer model when the generic kth layer approaches or become the top (or bottom) layer. Also, since the analysis for the kth ply group is performed as the analysis for a "local 90-deg ply," it is advantageous to express all the stiffnesses of the five layers in the ply coordinates of the ply group, k.

3 Quantitative Measures for the Anisotropic Constraint Effects on Damage Interaction

In this section, quantitative measures for the anisotropic constraint effects on stiffness reductions of a generic ply group will be introduced. This will be done within the framework of a general theory of overall moduli of a cracked body, contributed by Hill (1963), Horii and Nemat-Nasser (1983) and Kachanov (1992), to name a few. This is an elegant theory since it can be applied to a body with finite volume under general loading. However, the premise of this theory is that the medium must be homogeneous. This premise makes it difficult to directly apply this theory to laminates because laminates are highly inhomogeneous media with layers in different directions. Due to this fundamental difficulty, to the authors' knowledge, there has been no relationship between this elegant theory and damage analysis of laminated composites until the work given by the present author and his co-workers for cross ply and $[\pm\theta_m/90_n]_n$ lay-ups (Fan and Zhang, 1993).

In this paper, the approach therein will be extended to matrix cracking-delamination interaction for an arbitrary layer, k, of general lay-ups. The fact that the above-mentioned theory cannot be directly applied to the whole laminate does not mean that the theory can not be effectively applied to each individual ply. In this work, each individual ply is considered as an effective homogeneous anisotropic medium with distributed cracks. Consequently, the theory can be suitably applied to any individual ply separately. To conduct the above-mentioned analysis, the important thing is to determine the in-situ plane stiffness matrix $\bar{Q}_{ij}^{(k)}$ for the generic ply, k. By definition, this stiffness matrix correlates the average stress and average strain in the kth ply. Employing the contracted notation related to the local coordinate system, one expresses this correlation as

$$\bar{\sigma}_i^{(k)} = \bar{Q}_{ij}^{(k)}\bar{\epsilon}_j^{(k)}. \tag{1}$$

Applying the formulation of Horii and Nemat-Nasser (1983)

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

to the generic ply group, k, the plane stiffness matrix $\bar{Q}_{ij}^{(k)}$ can be obtained. For the case of transverse matrix crack-delamination interaction, $\bar{Q}_{16}^{(k)}$ and $\bar{Q}_{26}^{(k)}$ vanish, and Eq. (3) can be written in the following form (Appendix A):

$$\begin{bmatrix} Q_{11}^{(k)} & Q_{12}^{(k)} & 0 \\ Q_{12}^{(k)} & Q_{22}^{(k)} & 0 \\ 0 & 0 & Q_{66}^{(k)} \end{bmatrix} = \begin{bmatrix} Q_{11}^{0} & Q_{12}^{0} & 0 \\ Q_{12}^{0} & Q_{22}^{0} & 0 \\ 0 & 0 & Q_{66}^{0} \end{bmatrix} - \begin{bmatrix} \dfrac{(Q_{12}^{0})^2}{Q_{22}^{0}} \Lambda_{22}^{(k)} & Q_{12}^{0}\Lambda_{22}^{(k)} & 0 \\ Q_{12}^{0}\Lambda_{22}^{(k)} & Q_{22}^{0}\Lambda_{22}^{(k)} & 0 \\ 0 & 0 & Q_{66}^{0}\Lambda_{66}^{(k)} \end{bmatrix} \quad (2)$$

where Q_{ij}^{0} are the stiffnesses of an intact ply. The second matrix of the right-hand side of Eq. (2) denotes the stiffness reductions caused by damage. From the expressions of their components, we see that the reduced stiffness are expressible by means of just two factors, $\Lambda_{22}^{(k)}$ and $\Lambda_{66}^{(k)}$. Since the delamination is parallel to the $x_1 - x_2$-plane under consideration, the delamination does not change the form of the relationships between in-plane stresses and strains. Therefore, the form of Eqs. (1) and (2) will remain valid for the case of transverse matrix cracking only (Fan and Zhang, 1993).

It is important to notice that these two factors depend not only on the damage state, but also on the anisotropic constraint effects. This is because these factors correlate the average stress and average strain, and the constraint forces applied at the contact surfaces will change the average stress and average strain in the kth ply. Specifically, from Eqs. (1) and (2) these factors are directly related to the average stresses and strains in the ply according to

$$\Lambda_{22}^{(k)} = 1 - \frac{\bar{\sigma}_2^{(k)}}{Q_{22}^{0}\bar{\epsilon}_2^{(k)} + Q_{12}^{0}\bar{\epsilon}_1^{(k)}} \quad (3a)$$

$$\Lambda_{66}^{(k)} = 1 - \frac{\bar{\sigma}_6^{(k)}}{Q_{66}^{0}\bar{\epsilon}_6} . \quad (3b)$$

These factors are functions of the in-situ anisotropic constraint conditions, and therefore they will be taken as quantitative measures for the anisotropic constraint effects. The challenge we are facing is to develop an efficient method to express the parameters $\Lambda_{22}^{(k)}$ and $\Lambda_{66}^{(k)}$ for a damaged ply within a general multidirectional laminates. Obviously, these factors cannot be determined by an analysis of an isolated damaged ply alone. Rather, it should be determined through the analysis of the five-layer model.

4 Determination of the Effective Damage Factors

While the concept of the effective damage factors originates from the theory of overall stiffness of a cracked body, its determination will be quite different from the approach therein. In fact, the method proposed by Horri and Nemat-Nasser (1983) for the damage-induced term, $\bar{H}_{ijkl}$, (see Appendix A) is to express it by Green function or by an integration of displacement along the crack surface (see Eqn. (12)–(13) of their paper). This kind of approach is very complicated to apply for the damage problem herein and a new approach is developed in this work.

4.1 Stress Analysis of a Five-Layer Model. To exemplify the proposed methodology and avoid lengthy mathematical description, a problem of transverse matrix cracking-local delamination interaction of a symmetric five-layer model under tensile loading will be analyzed. The unsymmetric analysis with general in-plane loading will be published elsewhere. Here, local delamination denotes the interfacial cracks originate from transverse matrix cracks in the "local 90-deg ply." The assumption of plane strain with respect to the x_1-direction (fiber direction, see Fig. 2(a)) is adopted. This assumption makes the analysis to be a two-dimensional problem (x_2, x_3) in the principal directions of the damaged ply group. This assumption is reasonable since our main interest domain in this work is the interior area (not the edge) of laminates. Under these conditions, the constraint stress offered by the fiber along the x_1-direction is large enough to make the displacement along the lateral direction being much smaller than that along the x_2-direction. It turns out that these lateral constraints enforce an approximate plane-strain condition.

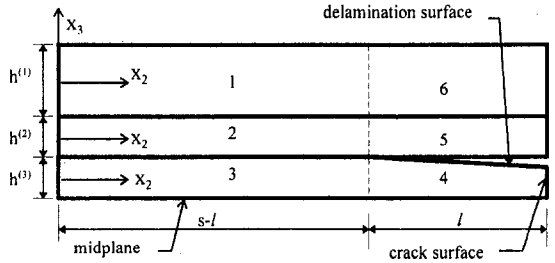


Fig. 3 Subregions of a quarter of the repeated interval of a symmetric five-layer model

In the symmetric case, only one quarter of the repeated interval of the five-layer model is needed for the analysis. This quarter which is subjected to a tensile loading along the x_2-direction can be considered as a unit cell for our analysis. It can be divided into six subregions in which subregions 3, 4 denote the 90-deg layers, 2, 5—adjacent layers, and 1, 6—the homogeneous remote layers (Fig. 3). As we can see from this unit cell, the half-delamination length, l, and the half-spacing, s, between two matrix cracks take dominate roles in the stress and strain distribution as well as the interaction between these two damage modes. This point of view can be described more clearly through a discussion about initiation and propagation of delamination. Actually, the initiation of delamination is triggered by the stress concentration at the tip of transverse matrix cracks. Its propagation depends, to a large extent, on the unmatched material property between the adjacent layers as well as the stress field in the matrix between two cracks. That is why delamination appears when there are matrix cracks in two adjacent layers which are not at the same angle. If the spacing, s, between two matrix cracks reduces, the stress in the matrix reduces and the driving forces (or the energy release rate) for delamination also reduce. This can cause the propagation of delamination to be stopped. Obviously, this interaction can be analyzed through an analysis of the proposed unit cell and quantitatively accounted for by the introduced parameters, s and l, as well as their normalized counterparts (see Eqs. (26), (28a, b), and (29a, b) for more detail).

The analysis for each subregion will be conducted by the first-order shear deformation plate theory. In this theory, the displacement field is assumed to be a linear function along the thickness direction. This stress analysis was used previously for a two-layer unit cell by Armanios et al. (1991). For a three-layer model, the application of this plate theory shows a good accuracy when it compares with the variational mechanics approach and the shear lag-analysis of a parabolic displacement field (see later). Here, by using this theory for each subregions of the five-layer model a sufficient accuracy for the analysis should be expected.

The extension and shear stress resultants, as well as the bending moment resultant, are denoted by N, Q, and M, respectively. The peel and interlaminar shear stresses are denoted by P and T, with t and b subscripts for the top and bottom surfaces, respectively. The number of the subregions will be denoted by subscripts. The basic equations are listed as follows:

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Displacement Field: Let v and w denote displacements relative to the x_2 and x_3-axes, V and W represents the axial midplane stretching and transverse displacement, respectively, and β is the rotation. The displacement field is described as

$$\nu(x_2, x_3) = V(x_2) + x_3\beta(x_2) \quad w(x_2, x_3) = W(x_2). \quad (4a, b)$$

The transverse normal strain, ϵ_3, is neglected and W is zero along center horizontal line of the unit cell. It turns out that the transverse displacement, w, is zero in subregion 1, 2, 3, and 4.

Equilibrium Equations: The x_3-axis is located at the central vertical line between two transverse cracks. The equilibrium equations of each subregion take the form

$$N_{,2} + T_t - T_b = 0 \quad (5a)$$

$$M_{,2} - Q + \frac{h}{2}(T_t + T_b) = 0 \quad (5b)$$

$$Q_{,2} + P_t - P_b = 0 \quad (5c)$$

where the comma indicates derivative; h is the thickness of a subregion.

Constitutive Equations in Terms of Forces and Moments:

$$N = A_{22}U_{,2} + B_{22}\beta_{,2} \quad (6a)$$

$$M = B_{22}U_{,2} + D_{22}\beta_{,2} \quad (6b)$$

$$Q = A_{44}(\beta + W_{,2}) \quad (6c)$$

where A_{ij}, B_{ij}, and D_{ij} are the stretching, coupling and bending stiffness from the classical laminate theory, respectively. In the following, analysis will be conducted through the following three groups of subregions: group I includes subregions 1, 2 and 3; group II—subregions 5, and 6; group III—subregion 4. The common procedure for formulating governing equations is to get relationships between interfacial forces (T_t, T_b) and displacements (β, V, W) through combining Eqs. ($5a, b$) with ($6a, b, c$). Then continuity conditions of displacements and reciprocal relations of forces at the interfaces between the subregions are used to obtain governing differential equations for $\beta^{(i)}$ ($i = 1, 2, \ldots, 5$). For group III, there is only one layer; its governing equation for $M^{(4)}$ can be easily obtained. The main results are as follows.

(A) Group I: Subregions 1, 2, and 3. The governing equations are

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{Bmatrix} \beta^{(1)}_{,22} \\ \beta^{(2)}_{,22} \\ \beta^{(3)}_{,22} \end{Bmatrix} - \begin{bmatrix} A^{(1)}_{44} & 0 & 0 \\ 0 & A^{(2)}_{44} & 0 \\ 0 & 0 & A^{(3)}_{44} \end{bmatrix} \begin{Bmatrix} \beta^{(1)} \\ \beta^{(2)} \\ \beta^{(3)} \end{Bmatrix} = 0 \quad (7)$$

where a_{ij} are symmetric. They are functions of the elastic constants and of the lay-up parameters (Appendix B). The characteristic equation corresponding to the Eq. (7) reads as follows:

$$A\lambda^{*3} - B\lambda^{*2} + C\lambda^{*} - D = 0 \quad (8a)$$

$$\text{where } \lambda^{*} = \lambda^2 \quad (8b)$$

$$A = a_{11}a_{22}a_{33} + 2a_{12}a_{13}a_{23} - a_{33}a_{12}^2 - a_{11}a_{23}^2 - a_{22}a_{13}^2 \quad (9a)$$

$$B = a_{11}a_{22}A^{(3)}_{44} + a_{33}a_{22}A^{(1)}_{44} + a_{33}a_{11}A^{(2)}_{44} - a_{12}^2A^{(3)}_{44} - a_{23}^2A^{(1)}_{44} - a_{13}^2A^{(2)}_{44} \quad (9b)$$

$$C = a_{11}A^{(2)}_{44}A^{(3)}_{44} + a_{22}A^{(1)}_{44}A^{(3)}_{44} + a_{33}A^{(2)}_{44}A^{(1)}_{44} \quad (9c)$$

$$D = A^{(1)}_{44}A^{(2)}_{44}A^{(3)}_{44}. \quad (9d)$$

The three roots λ_j^* ($j = 1, 2, 3$) are positive real numbers for all laminates examined here. An eigenvector $\{p_j\} = (p_j^{(1)}\ p_j^{(2)}\ 1)^T$ corresponding to the eigenvalue λ_j^* is given, where

$$\begin{Bmatrix} p_j^{(1)} \\ p_j^{(2)} \end{Bmatrix} = \frac{\lambda_j^*}{(a_{11}\lambda_j^* - A^{(1)}_{44})(a_{22}\lambda_j^* - A^{(2)}_{44}) - a_{12}^2\lambda_j^{*2}} \times \begin{Bmatrix} a_{13}A^{(2)}_{44} - (a_{22}a_{13} - a_{12}a_{23})\lambda_j^* \\ a_{23}A^{(1)}_{44} - (a_{11}a_{23} - a_{12}a_{13})\lambda_j^* \end{Bmatrix}$$

$$\text{for } j = 1, 2, 3. \quad (10)$$

Since $w_{,x} = 0$ and Q is antisymmetric about the x_3-axis, from Eq. ($6c$) β is also antisymmetric, i.e., $\beta(x_2) = -\beta(-x_2)$. Using this condition for the sublaminates 1, 2 and 3 the solutions $\beta^{(i)}$ in the governing Eqs. (7) read

$$\beta^{(i)} = \sum_{j=1}^{3} \alpha_j p_j^{(i)} \sinh(\lambda_j x_2) \quad \text{for } i = 1, 2, 3 \quad (11)$$

the parameters α_j are undetermined constants. Combining Eq. (11) with Eqs. ($5a, b, c$) and Eqs. ($6a, b, c$), interfacial forces, moments, and force resultants can be expressed as follows:

$$N^{(i)} = \sum_{j=1}^{3} \alpha_j \eta_j^{(i)} \cosh(\lambda_j x_2) + A^{(i)}_{22}\alpha_{3+i} \quad \text{for } i = 1, 2, 3 \quad (12)$$

$$M^{(i)} = \sum_{j=1}^{3} \alpha_j \xi_j^{(i)} \cosh(\lambda_j x_2) + B^{(i)}_{22}\alpha_{3+i} \quad \text{for } i = 1, 2, 3 \quad (13)$$

$$T_b^{(1)} = T_t^{(2)} = \sum_{j=1}^{3} \alpha_j \lambda_j^2 (A^{(1)}_{22}\gamma_j^{(1)} + B^{(1)}_{22}p_j^{(1)}) \sinh(\lambda_j x_2) \quad (14)$$

$$T_b^{(2)} = T_t^{(3)} = -A^{(3)}_{22} \sum_{j=1}^{3} \alpha_j \gamma_j^{(3)} \lambda_j^2 \sinh(\lambda_j x_2) \quad (15a)$$

$$V^{(i)} = \sum_{j=1}^{3} \alpha_j \gamma_j^{(i)} \sinh(\lambda_j x_2) + \alpha_{3+i} x_2 \quad (15b)$$

where $\gamma_j^{(i)}$, $\eta_j^{(i)}$, and $\xi_j^{(i)}$ are constants (Appendix B).

(B) Group II: Subregions 5–6. The governing equations are

$$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{Bmatrix} \beta^{(5)}_{,22} \\ \beta^{(6)}_{,22} \end{Bmatrix} + \frac{A^{(1)}_{44}A^{(2)}_{44}}{A^{(1)}_{44} + A^{(2)}_{44}} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{Bmatrix} \beta^{(5)} \\ \beta^{(6)} \end{Bmatrix} = 0 \quad (16)$$

where b_{ij} are symmetric and functions of the elastic constants and the lay-up parameters (Appendix B). The general solutions of Eq. (16) are written as

$$\begin{Bmatrix} \beta^{(5)} \\ \beta^{(6)} \end{Bmatrix} = \theta_1 e^{\omega x_2} \begin{Bmatrix} q \\ 1 \end{Bmatrix} + \theta_2 e^{-\omega x_2} \begin{Bmatrix} q \\ 1 \end{Bmatrix} + \theta_3 x_2 \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} + \theta_4 \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \quad (17)$$

where ω is the nonzero eigenvalue of Eq. (16) and is given by

$$\omega = \sqrt{\frac{A^{(1)}_{44}A^{(2)}_{44}(b_{11} + b_{22} + 2b_{12})}{(A^{(1)}_{44} + A^{(2)}_{44})(b_{11}b_{22} - b_{12}^2)}} \quad (18)$$

and

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$q = -\frac{b_{22} + b_{12}}{b_{11} + b_{12}}. \tag{19}$$

Combining Eq. (17) with Eqs. (5a, b, c) and Eq. (6a, b, c), the moment and force resultants can be determined as follows:

$$N^{(5)} = -k_1 A_{22}^{(2)} \omega(\theta_1 e^{\omega x_2} - \theta_2 e^{-\omega x_2}) + A_{22}^{(2)}\theta_5 \tag{20a}$$

$$N^{(6)} = k_1 A_{22}^{(2)} \omega(\theta_1 e^{\omega x_2} - \theta_2 e^{-\omega x_2}) + A_{22}^{(1)}\theta_6 + B_{22}^{(1)}\theta_3 \tag{20b}$$

$$M^{(5)} = k_2(\theta_1 e^{\omega x_2} - \theta_2 e^{-\omega x_2}) + D_{22}^{(2)}\theta_3 \tag{20c}$$

$$M^{(6)} = \omega(D_{22}^{(1)} + B_{22}^{(1)}k_3)(\theta_1 e^{\omega x_2} - \theta_2 e^{-\omega x_2}) + D_{22}^{(1)}\theta_3 + B_{22}^{(1)}\theta_6 \tag{20d}$$

$$V^{(5)} = -k_1(\theta_1 e^{\omega y} + \theta_2 e^{-\omega y}) + \theta_5 y + \theta_7 \tag{20e}$$

$$V^{(6)} = k_3(\theta_1 e^{\omega y} + \theta_2 e^{-\omega y}) + \theta_6 y + \theta_8 \tag{20f}$$

where the laminate constant k_i are given in Appendix B; the unknown coefficients θ_i ($i = 1, 2, \ldots, 8$) are to be determined by boundary conditions.

(C) Group III: Subregion 4. For subregion 4, we have

$$N^{(4)} = 0 \tag{21a}$$

$$M_{,22}^{(4)} - \frac{A_{44}^{(4)}}{D_{22}^{(4)}} M_2^{(4)} = 0. \tag{21b}$$

The solution can be expressed as

$$M^{(4)} = D_{22}^{(3)}(\psi_1 e^{\omega_1 x_2} - \psi_2 e^{-\omega_1 x_2}) \tag{21c}$$

$$\text{where } \omega_1 = \sqrt{\frac{A_{44}^{(4)}}{D_{11}^{(4)}}}. \tag{21d}$$

In order to determine the 16 constants of α_i ($i = 1, 2, \ldots, 6$), θ_j ($j = 1, 2, \ldots, 8$) and ψ_k ($k = 1, 2$), the same number of independent boundary and continuity conditions have to be prescribed. Assuming the laminate subjected to a tensile force in the x_2-direction, N, the boundary conditions at the ends of the unit cell are given as follows:

$$\beta^{(5)}(s) = 0 \tag{22a}$$

$$\beta^{(6)}(s) = 0 \tag{22b}$$

$$U^{(5)}(s) = U^{(6)}(s) \tag{22c}$$

$$N^{(5)}(s) + N^{(6)}(s) = \tfrac{1}{2}N \tag{22d}$$

$$N^{(1)}(0) + N^{(2)}(0) + N^{(3)}(0) = \tfrac{1}{2}N \tag{22e}$$

$$U^{(1)}(s-l) - \frac{h^{(1)}}{2}\beta^{(1)}(s-l) = U^{(2)}(s-l) + \frac{h^{(2)}}{2}\beta^{(2)}(s-l) \tag{22f}$$

$$U^{(2)}(s-l) - \frac{h^{(2)}}{2}\beta^{(2)}(s-l) = U^{(3)}(s-l) + \frac{h^{(3)}}{2}\beta^{(3)}(s-l) \tag{22g}$$

$$U^{(1)}(s-l) = U^{(6)}(s-l) \tag{22h}$$

$$U^{(2)}(s-l) = U^{(5)}(s-l) \tag{22i}$$

$$\beta^{(1)}(s-l) = \beta^{(6)}(s-l) \tag{22j}$$

$$\beta^{(2)}(s-l) = \beta^{(5)}(s-l) \tag{22k}$$

$$\beta^{(3)}(s-l) = \beta^{(4)}(s-l) \tag{22l}$$

$$N^{(2)}(s-l) = N^{(5)}(s-l) \tag{22m}$$

$$N^{(3)}(s-l) = N^{(4)}(s-l) \tag{22n}$$

$$M^{(2)}(s-l) = M^{(5)}(s-l) \tag{22o}$$

$$M^{(3)}(s-l) = M^{(4)}(s-l). \tag{22p}$$

The continuity conditions of the force and moment resultants, N and M, between the sublaminates 1 and 6 are not needed to be specified since they are derivable.

4.2 A Closed-Form Expressions for the Parameter Λ_{22}. For the case of matrix cracking-local delamination, the assumed plane-strain condition implies $\bar{\epsilon}_1 = 0$ and $\bar{\epsilon}_{21} = 0$. In this case, only Λ_{22} is needed. When using Eq. (3a) to calculate the effective damage factors, the compatibility conditions of average strain between different layers are used, i.e.,

$$\bar{\epsilon}_2^{(k)} = \bar{\epsilon}_2^{(k+1)} = \bar{\epsilon}_2^{(l_k)} = \frac{\bar{N}_2^{(l_k)}}{A_{22}^{(l_k)}}. \tag{23}$$

This result allows to replace $\bar{\epsilon}_2^{(k)}$ in Eq. (3a) by $\bar{N}_2^{(l_k)}/A_{22}^{(l_k)}$, and leads to

$$\Lambda_{22}^{(k)} = 1 - \frac{A_{22}^{(l_k)}\bar{N}_2^{(k)}}{A_{22}^{(k)}\bar{N}_2^{(l_k)}} \tag{24}$$

where

$$\bar{N}_2^{(k)} = \frac{1}{s}\left(\int_0^{s-l} N_2^{(3)} dx_2 + \int_{s-l}^{s} N_2^{(4)} dx_2\right) \tag{25a}$$

$$\bar{N}_2^{(l_k)} = \frac{1}{s}\left(\int_0^{s-l} N_2^{(1)} dx_2 + \int_{s-l}^{s} N_2^{(6)} dx_2\right). \tag{25b}$$

Substituting Eqs. (12), (20a, b), and (21a) into Eqs. (25a, b), then into Eq. (24), we obtain

$$\Lambda_{22} = \frac{(1+\chi)\left[\frac{l}{s} + \sum_{j=1}^{3} \bar{\alpha}_j(s, l)\gamma_j^{(3)} \frac{\tanh(\lambda_j(s-l))}{s}\right]}{1 + \chi\left[\frac{l}{s} + \sum_{j=1}^{3} \bar{\alpha}_j(s, l)\gamma_j^{(3)} \frac{\tanh(\lambda_j(s-l))}{s}\right]} \tag{26}$$

where χ, $\bar{\alpha}_j(s, l)$ ($j = 1, 2, 3$) are given in Appendix B.

5 Damage Evolution Law

While the proposed methodology can readily handle the hygrothermal effects on the energy release rate (Fan and Zhang, 1993), these effects will not be discussed here for simplicity. In general, if ϕ denotes the total potential energy of a material body containing m cracks of areas A_i ($i = 1, 2, \ldots m$), then the available energy release rate, $G^{(n)}$, associated with the growth of the ith crack is given by

$$G^{(i)} = -\frac{\partial \phi}{\partial A_i}. \tag{27}$$

By introducing the following definitions of the dimensionless delamination length and the matrix crack density for the rth lamina,

$$D^{(r)^{mc}} = h^{(r)}/2s^{(r)} \tag{28a}$$

$$D^{(r)^{dl}} = l^{(r)}/s^{(r)}, \tag{28b}$$

where $h^{(r)}$ is the thickness of the rth lamina, and is replaced with $2h^{(r)}$ for the central ply group, the expressions for the

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

energy release rates associated with matrix cracking and delamination are obtained, respectively (Fan, Zhang, 1993),

$$G^{(k)^{mc}} = -\sum_r h^{(r)} \frac{\partial \bar{Q}_{ij}^{(r)}}{\partial D^{(r)^{mc}}} \bar{\epsilon}_i^{(r)} \bar{\epsilon}_j^{(r)} \quad (29a)$$

$$G^{(k)^{dl}} = -\frac{1}{2} \sum_r h^{(r)} \frac{\partial \bar{Q}_{ij}^{(r)}}{\partial D^{(r)^{dl}}} \bar{\epsilon}_i^{(r)} \bar{\epsilon}_j^{(r)} \quad (29b)$$

where repeated indices, i and j, imply summation over their ranges ($i, j = 1, 2, 6$). If damage only occurs in the weakest 90-deg layer or the interactive effects between damage in the generic kth ply and stiffness reductions in the remaining damaged layers can be neglected, the calculation of the energy release rate becomes much simpler. In this case, the only contribution to $G^{(k)dl}$ and $G^{(k)mc}$ comes from the reduced sums, i.e.,

$$G^{(k)^{mc}} = -h^{(k)} \frac{\partial \bar{Q}_{ij}^{(k)}}{\partial D^{(k)^{mc}}} \bar{\epsilon}_i^{(k)} \bar{\epsilon}_j^{(k)} \quad \text{(no summation for } k) \quad (30a)$$

$$G^{(k)^{dl}} = -\frac{h^{(k)}}{2} \frac{\partial \bar{Q}_{ij}^{(k)}}{\partial D^{(k)^{dl}}} \bar{\epsilon}_i^{(k)} \bar{\epsilon}_j^{(k)} \quad \text{(no summation for } k). \quad (30b)$$

By using Eqs. (2), (26), and (30a, b), the energy release rates, $G^{(k)dl}$ and $G^{(k)mc}$, can be obtained from the derivatives of Λ_{22} with respect to $D^{(k)dl}$ and $D^{(k)mc}$, respectively. Based on these calculations, the level of available fracture energies under the current anisotropic constraint conditions is evaluated and arranged in hierarchical order for delamination versus matrix cracking modes, with individual accounting for each damaged ply. The Griffith energy principle is used in developing criteria for the damage initiation and progression of matrix cracking and delamination. These criteria are:

$$\text{matrix cracking: } G^{(k)mc} = G_c^{mc} \quad (31a)$$

$$\text{delamination: } G^{(k)dl} = G_c^{dl}. \quad (31b)$$

The quantities in the left-hand side of Eqs. (31a, b) denote energy release rates (or available fracture energies), and those on the right-hand side denote critical energy release rate (or required fracture energy for cracking). The damage scenario is generated by gradually incrementing the load amplitudes, tracking the comparative ratios between levels of available and required fracture energies, and switching to new damage configurations for which the criterion of damage evolution has been satisfied.

6 Numerical Results and Verication

The approach proposed in this work is based on the concept of effective damage factors which was previously named as

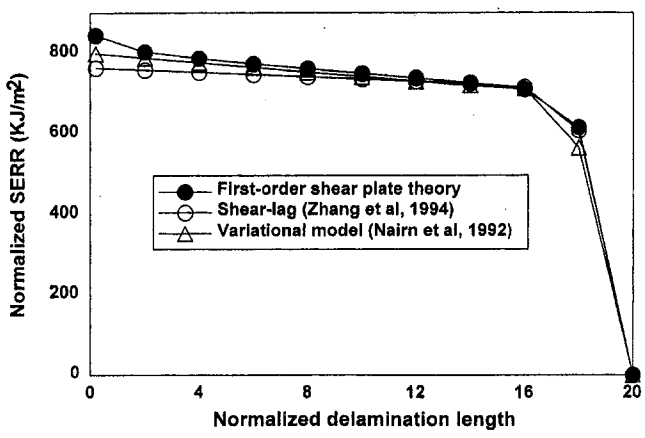


Fig. 4 The normalized strain energy release rate, G^{dl}/ϵ_x^2, versus the normalized delamination length, l/h, for the $[0/90]_s$ laminate

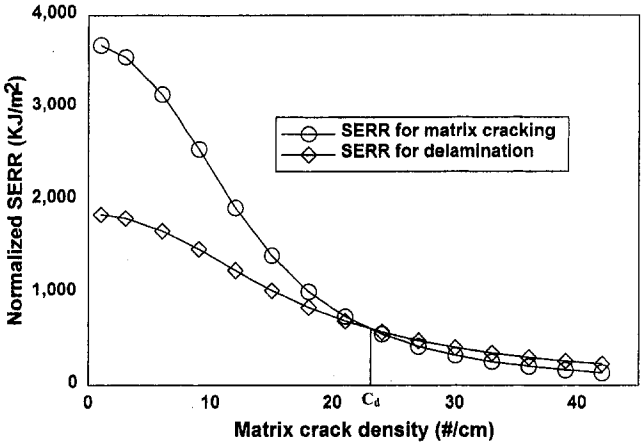


Fig. 5 The normalized strain energy release rates for delamination initiation at the −45 deg/90 deg interface and for continued 90-deg ply cracking as functions of matrix crack density for the $[0_4/\pm 45_4/90_n]_s$ of T300/976 Gr/Ep

IDEF, short for the in-situ damage effective functions (IDEF). This concept was applied to cross-ply and $[\pm\theta_m/90_n^\circ]_s$ laminates using a shear lag analysis for a three-layer model. The present work has been developed to apply this concept to general lay-up composites with two differences from the previous approach. One is the use of a five-layer model instead of the three-layer one to describe constraint effects. The other is the use of the first-order shear deformation plate theory instead of shear lag analysis for stress analysis. While the verification for the previous approach was extensive, the verification for the new formulation is difficult. This is because the fact that there are very limited experimental results of general lay-up composites as well as for interactive damage progression. The consideration here is to conduct the verification in a hierarchical and indirect fashion. The material constants of T300/976 graphite/epoxy used in the numerical analysis are (Choi et al., 1991a)

$$E_{11} = 144.8 \text{ GPa}, \quad E_{22} = E_{33} = 11.38 \text{ GPa},$$

$$G_{12} = G_{13} = 6.48 \text{ GPa}, \quad G_{23} = 3.45 \text{ GPa},$$

$$\nu_{12} = \nu_{13} = 0.3, \text{ ply thickness } t = 0.132 \text{ mm}.$$

First of all, results obtained by the present approach are compared with those of the Narin-Hu variational mechanics approach (1992) and our previous three-layer shear lag analysis. The latter assumes a linear stress distribution or parabolic variation of the in-plane displacement for each layer. The detail of the calculation is described elsewhere (Zhang, Fan, and Herrmann, 1998). The comparison is for cross-ply, $[0/90]_s$ which is due to the difficulties of using the latter two approaches to general lay-up composites. This comparison is shown in Fig. 4, where the normalized strain energy release rates (SEER) at the 0 deg/90 deg interface are plotted against the normalized delamination length. The normalization for energy release rate is with respect to the square of the uniaxial laminate strain and delamination is to the ply thickness. The comparison between the three models suggests that the present approach using the first-order shear deformation plate theory provide a reasonable agreement with the variational method and the modified shear-lag analysis.

Next, general lay-up composites are used for numerical analysis to take the advantage of the present approach. Figure 5 shows the normalized strain energy release rates for delamination initiation at the −45 deg/90 deg interface and for continued 90-deg ply cracking of the $[0_4/\pm 45_4/90_2]_s$ laminate. These energy release rats are described as functions of matrix cracking density. At the first stage the available fracture energy for the transverse matrix cracking is higher than that for initiation of the delamination. Consequently, matrix cracking is the first failure

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

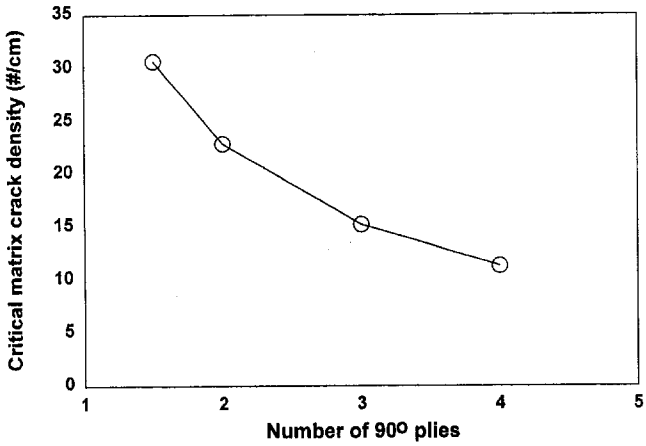


Fig. 6 Critical matrix crack density versus the number of 90-deg plies for the $[0_4/\pm 45_4/90_n]_s$

mode (Remark: The difference of critical energy release rates between delamination and matrix cracking are small, see Nairn and Hu, 1992). When matrix crack density increases both energy release rates are reduced, with the reduction for matrix cracking being much faster than the one for delamination. Eventually, at some critical crack density, C_d, the strain energy release rate for delamination exceeds the one for matrix cracking. In this case, a more applied stress is required for further matrix cracking than for initiation of delamination and, in turn, the transition from matrix cracking to delamination occurs at the critical point.

Figure 6 shows how the critical matrix cracking density varies with the number of the 90-deg plies. Since the thickness of each ply is the same, this figure describes the effects of the thickness of the 90-deg layer on the transition from matrix cracking to delamination. It can be seen that as the thickness of the 90-deg layer increases the critical crack density decreases for the laminate $[0_4/\pm 45_4/90_n]_s$. This result means that when the 90-deg layer becomes very thin, the transition from matrix cracking to delamination becomes very difficult.

7 Conclusions

1 In this work, a new method of damage analysis has been developed. The constraint effects are accounted for by proposing a five-layer model for a generic damaged ply. Incorporating this model with the general theory of overall moduli of a cracked body and considering any generic damaged ply as a "local 90-deg layer," the effective damage factors proposed by Fan and Zhang in 1993 for characterizing the reduction of in-plane stiffnesses have been extended to interactive damage progression of general lay-up composites.

2 The effective damage factors of a damaged ply can be determined by conducting stress analysis of a relevant five-layer model. As an example the effective damage factor Λ_{22} of the central 90-deg layer within symmetric laminates is derived as a function of matrix crack density and delamination length. The strain energy release rates associated with matrix cracking and delamination are directly related to the partial derivatives of the effective damage factors with respect to corresponding damage parameters.

3 By the proposed approach, the initiation and progression of each damage mode, as well as the transition from matrix cracking to delamination can be predicted. This methodology was exemplified by investigating the progressive matrix cracking-local delamination interaction.

Acknowledgments

The first author would like to thank Prof. Y. Jack Weitsman for his valuable suggestions. This work was financially supported by the National Science Foundation of China through contract number 59172097.

References

Allen, D. H., and Harris, C. E., 1987, "A Thermomechanical Constitutive Theory for Elastic Composites with Distributed Damage," *Int. J. Solids & Structures,* Vol. 23, pp. 1301–1338.

Armanios, E. A., Sriram, P., and Badir, A. M., 1991, "Fracture Analysis of Transverse Crack-tip and Free-edge Delamination in Laminated Composites," *Comp. Mater.: Fatigue and Fracture,* ASME STP 1110, pp. 269–286.

Bailey, J. E., Curtis, P. T., and Parvizi, A., 1979, "On the Transverse Cracking and Longitudinal Splitting Behavior of Glass and Carbon Fiber-Reinforced Epoxy Cross-ply Laminates and The Effect of Poisson and thermal Generated Strain," *Proc. R. Soc. Lond.,* Vol. A366, pp. 599–633.

Choi, H. Y., Downs, R. J., and Chang, F. K., 1991a, "A New Approach Toward Understanding Damage Mechanisms and Mechanics of Laminated Composites Due to Low-Velocity Impact. Part 1—Experiments," *J. Comp. Mater.,* Vol. 25, pp. 992–1011.

Choi, H. Y., Downs, R. J., and Chang, F. K., 1991b, "A New Approach Toward Understanding Damage Mechanisms and Mechanics of Laminated Composites Due to Low-Velocity Impact. Part 2—Analysis," *J. Comp. Mater.,* Vol. 25, pp. 1011–1038.

Crossman, F. W., and Wang, A. S. D., 1982, "The Dependence of Transverse Cracking and Delamination on Ply Thickness in Graphite/Epoxy Laminates," *Damage in Comp. Mater.,* ASTM STP 775, pp. 118–139.

Dvorak, G. J., Laws, N., and Hejazi, M., 1985, "Analysis of progressive matrix cracking in composite laminates," *J. Comp. Mater.,* Vol. 19, pp. 216–234.

Fan, J., and Zhang, J., 1993, "In-site Damage Evolution and Micro/Macro Transition for Laminated Composites," *Comp. Sci. Tech.,* Vol. 47, pp. 107–118.

Flaggs, D. L., and Kural, M. H., 1982, "Experimental Determination of the In-situ Transverse Lamina Strength in Graphite/Epoxy Laminates," *J. Comp. Mater.,* Vol. 16, pp. 103–116.

Flaggs, D. L., 1985, "Prediction of Tensile Matrix Failure in Composite Laminates," *J. Comp. Mater.,* Vol. 19, pp. 29–50.

Fukunaga, H., Chou, T., Peters, P. W., and Schulte, K., 1984, "Probabilistic Failure Strength Analyses of Graphite/Epoxy Cross-Ply Laminates," *J. Comp. Mater,* Vol. 18, pp. 339–356.

Gudmundson, P., and Zhang, W., 1993, "An Analytical Model for Thermoelastic Properties of Composite Laminates containing Transverse Matrix Cracks," *Int. J. Solids and Structures,* Vol. 30, pp. 3211–3231.

Han, Y. M., and Hahn, H. T., 1989, "Ply Cracking and Property Degradation of Symmetric Balanced Laminates Under General In-Plane Loading," *Comp. Sci. Tech,* Vol. 35, pp. 377–397.

Hashin, Z., 1985, "Analysis of Cracked Laminates: A Variational Approach," *Mech. Mater,* Vol. 4, pp. 121–136.

Highsmith, A. L., and Reifsnider, K. L., 1982, "Stiffness Reduction Mechanisms in Composite Laminates," *Damage in Comp. Mater.,* ASTM STP 775, pp. 103–117.

Hill, R., 1963, "Elastic Properties of Reinforced Solids: Some Theoretical Principles," *J. Mech. Phys. Solids,* Vol. 11, pp. 357–372.

Horii, H., and Nemat-Nasser, S., 1980, "Overall Moduli of Solids with Microcracks: Load Induced Anisotropy," *J. Mech. Phys. Solids,* Vol. 31, pp. 155–177.

Kachanov, M., 1992, "Effective elastic properties of cracked solids: critical review of some basic concepts," ASME *Appl. Mech. Rev.* Vol. 45, pp. 304–335.

Laws, N., and Dvorak, G. J., 1988, "Progressive Transverse cracking in Composite Laminates," *J. Comp. Mater.,* Vol. 22, pp. 900–915.

Lee, J.-W., and Allen, D. H., 1989, "Internal State Variable Approach for Predicting Stiffness Reductions in Fibrous Laminated Composites with Matrix Cracks," *J Comp Mater,* Vol. 23, pp. 1273–1291.

Lee, J.-W., and Daniel, I. M., 1990, "Progressive transverse cracking of cross-ply composite laminates," *Composites Science and Technology,* Vol. 24, pp. 1225–1243.

Nairn, J. A., and Hu, S., 1992, "The Initiation and Growth of Delaminations Induced by Matrix Microcracks in Laminated Composites," *Int. J. Fracture,* Vol. 57, pp. 1–24.

Nuismer, R. J., and Tan, S. C., 1988, "A Theory for Progressive Matrix Cracking in Composite Laminates," *J. Comp. Mater,* Vol. 22, pp. 307–321.

Parvizi, A., Garrett, K. W., and Bailey, J. E., 1978, "Constrained Cracking in Glass Fiber-Reinforced Epoxy Glass-Ply Laminates," *J. Mater. Sci.,* Vol. 13, pp. 195–201.

Talreja, R., 1985, "A Continuum Mechanics Characterization of Damage in Composite Materials," *Proc. R. Soc. Lond.,* Vol. A399, pp. 195–216.

Thomas, D. J., 1995, "Analysis of the Progressive Failure of Brittle Matrix Composites," NASA Contract Report 195415.

Timmer, J. L., and Hahn H. T., 1993, "The Effect of Preloading on Fatigue Damage in Composites," *Mech. Comp. Review,* pp. 144–153.

Tsai, C. L., and Daniel, I. M., 1990, "Progressive Transverse Cracking of Cross-ply Composite Laminates," *Microcracking induced Damage in Comp.,* AMD-Vol. 111, MD-Vol. 22, pp. 9–18.

Wang, A. S. D., Chou, P. C., and Lei, S. C., 1984, "A Stochastic Model for the Growth of Matrix in Composite Laminates," *J. Comp. Mater.,* Vol. 18, pp. 239–254.

Zhang, J., Fan, J., and Soutis, C., 1992a, "Progressive matrix cracking in composites loaded in tension," *Advanced Composites Letters,* Vol. 1, pp. 12–15.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Zhang, J., Fan, J., and Soutis, C., 1992b, "Analysis of Multiple Matrix Cracking in $[\pm\theta_m/90_n]_s$ Composite Laminates, Part I: In-plane Stiffness Properties," *Composites,* Vol. 23, pp. 291–298.

Zhang, J., Fan, J., and Soutis, C., 1992c, "Analysis of Multiple Matrix Cracking in $[\pm\theta_m/90_n]_s$ Composite Laminates, Part II: Development of Transverse Matrix Cracks," *Composites,* Vol. 23, pp. 299–304.

Zhang, J., Soutis, C., and Fan, J., 1994a, "Effects of Matrix Cracking and Hygrothermal Stresses on the Strain Energy Release Rate for Edge Delamination Composite Laminates," *Composites,* Vol. 25, 27–35.

Zhang, J., Soutis, C., and Fan, J., 1994b, "Strain energy release rate associated with local delamination in cracked composite laminates," *Composites,* Vol. 25, pp. 851–862.

Zhang, J., Fan, J., and Herrmann, K. P., 1998 "Delaminations induced by constrained transverse cracking in symmetric composite laminates," *Int. J. Solids & Structures,* submitted for publication.

APPENDIX A

Employing the contracted notation for the lamina coordinate system, the compliances of a damaged lamina are expressed as (Horii and Nemat-Nasser, 1983)

$$S_{ij} = S^0_{ij} + H_{ij}. \tag{A1}$$

From Eq. (12) of Horii and Nemat-Nasser (1983), H_{12} and H_{11} can be proved to be zero for the case of transverse matrix cracking-local delamination with respect to the ply coordinate system. In fact from that equation one can write

$$H_{12}\bar{\sigma}_1 = \frac{1}{V}\left(\int_{s^{mc}} u_2 n_2 ds^{mc} + \int_{s^{dl}} u_2 n_2 ds^{dl}\right) \tag{A2}$$

$$H_{11}\bar{\sigma}_1 = \frac{1}{V}\left(\int_{s^{mc}} u_1 n_1 ds^{mc} + \int_{s^{dl}} u_1 n_1 ds^{dl}\right) \tag{A3}$$

where n_i ($i = 1, 2, 3$) are components of the unit normal vector of crack surfaces.

For matrix cracking surface s^{mc}:

$$n_1 = \pm 1,\ n_2 = 0,\ n_3 = 0 \tag{A4}$$

For delamination surface s^{dl}: $n_1 = 0,\ n_2 = 0,\ n_3 = \pm 1$ (A5)

Due to Eqs. (A4) and (A5), the right-hand side of Eq. (A3) and the second member of the Eq. (A2) are zero. It turns out $H_{11} = 0$ and

$$H_{12}\bar{\sigma}_1 = \frac{1}{V}\int_{s^{mc+}} [u_2^+ - u_2^-] ds^{mc+} \tag{A6}$$

where superscript "+" denotes the related quantity; u_2^+ belongs to the positive surface of transverse matrix crack with $n_1 = 1$, and "−" denotes the negative crack surface with $n_1 = -1$. Because applied stress $\bar{\sigma}_1$ cannot cause the displacement discontinuous of u_2 across the transverse matrix crack, we have $[u_2^+ - u_2^-] = 0$, and in turn, the right-hand side of Eq. (A6) is zero. It turns out $H_{12} = 0$. Furthermore, since the stiffness and compliance matrix are mutually inverse, after some algebraic work Eq. (2) can be obtained with

$$\Lambda^{(k)}_{ii} = \frac{Q^0_{ii} H^{(k)}_{ii}}{1 + Q^0_{ii} H^{(k)}_{ii}} \quad \text{for } ii = 22, 66. \tag{A7}$$

APPENDIX B

$$A_{22} = 2(A^{(1)}_{22} + A^{(2)}_{22} + A^{(3)}_{22}) \tag{B1a}$$

$$\chi = \frac{A^{(3)}_{22}}{A^{(1)}_{22} + A^{(2)}_{22}} \tag{B1b}$$

$$a_{11} = D^{(1)}_{22} + \frac{h^{(1)}(4B^{(1)}_{22} + h^{(1)}A^{(1)}_{22})(A^{(2)}_{22} + A^{(3)}_{22}) - 4B^{(1)2}_{22}}{2A_{22}} \tag{B2a}$$

$$a_{12} = \frac{h^{(2)}(B^{(1)}_{22} + \frac{1}{2}h^{(1)}A^{(1)}_{22})(A^{(2)}_{22} + 2A^{(3)}_{22})}{A_{22}} \tag{B2b}$$

$$a_{13} = \frac{h^{(3)}(B^{(1)}_{22} + \frac{1}{2}h^{(1)}A^{(1)}_{22})A^{(3)}_{22}}{A_{22}} \tag{B2c}$$

$$a_{22} = D^{(2)}_{22} + \frac{h^{(2)2}(4A^{(1)}_{22}A^{(3)}_{22} + A^{(2)}_{22}A^{(3)}_{22} + A^{(1)}_{22}A^{(2)}_{22})}{2A_{22}} \tag{B2d}$$

$$a_{23} = \frac{h^{(2)}h^{(3)}A^{(3)}_{22}(A^{(2)}_{22} + 2A^{(1)}_{22})}{2A_{22}} \tag{B2e}$$

$$a_{33} = D^{(3)}_{22} + \frac{h^{(3)2}A^{(3)}_{22}(A^{(2)}_{22} + A^{(1)}_{22})}{2A_{22}} \tag{B2f}$$

$$\eta^{(1)}_j = (B^{(1)}_{22}p^{(1)}_j + A^{(1)}_{22}\gamma^{(1)}_j)\lambda_j \tag{B3a}$$

$$\eta^{(2)}_j = A^{(2)}_{22}\gamma^{(2)}_j\lambda_j \tag{B3b}$$

$$\eta^{(3)}_j = A^{(3)}_{22}\gamma^{(3)}_j\lambda_j = -(\eta^{(1)}_j + \eta^{(2)}_j) \tag{B3c}$$

$$\xi^{(1)}_j = (D^{(1)}_{22}p^{(1)}_j + B^{(1)}_{22}\gamma^{(1)}_j)\lambda_j \tag{B3d}$$

$$\xi^{(2)}_j = D^{(2)}_{22}p^{(2)}_j\lambda_j \tag{B3e}$$

$$\xi^{(3)}_j = D^{(3)}_{22}\lambda_j \tag{B3f}$$

$$\gamma^{(1)}_j = \frac{p^{(1)}_j[h^{(1)}(A^{(2)}_{22} + A^{(3)}_{22}) - 2B^{(1)}_{22}] + p^{(2)}_j h^{(2)}(A^{(2)}_{22} + 2A^{(3)}_{22}) + h^{(3)}A^{(3)}_{22}}{A_{22}} \tag{B3g}$$

$$\gamma^{(2)}_j = -\frac{p^{(1)}_j(h^{(1)}A^{(1)}_{22} + 2B^{(1)}_{22}) + p^{(2)}_j h^{(2)}(A^{(1)}_{22} - A^{(3)}_{22}) - h^{(3)}A^{(3)}_{22}}{A_{22}} \tag{B3h}$$

$$\gamma^{(3)}_j = -\frac{p^{(1)}_j(h^{(1)}A^{(1)}_{22} + 2B^{(1)}_{22}) + p^{(2)}_j h^{(2)}(2A^{(1)}_{22} + A^{(2)}_{22}) + h^{(3)}(A^{(1)}_{22} + A^{(2)}_{22})}{A_{22}} \quad \text{for } j = 1, 2, 3 \tag{B3i}$$

$$b_{11} = D^{(2)}_{22} + \frac{h^{(2)2}A^{(1)}_{22}A^{(2)}_{22}}{4(A^{(1)}_{22} + A^{(2)}_{22})} \tag{B4a}$$

$$b_{21} = b_{12} = \frac{h^{(2)}A^{(2)}_{22}(2B^{(1)}_{22} + h^{(1)}A^{(1)}_{22})}{4(A^{(1)}_{22} + A^{(2)}_{22})} \tag{B4b}$$

$$b_{22} = D^{(1)}_{22} + \frac{h^{(1)}A^{(2)}_{22}(4B^{(1)}_{22} + A^{(1)}_{22}h^{(1)}) - 4B^{(1)2}_{22}}{4(A^{(1)}_{22} + A^{(2)}_{22})} \tag{B4c}$$

$$k_1 = \frac{A^{(1)}_{22}(h^{(2)}q + h^{(1)}) + 2B^{(1)}_{22}}{2(A^{(1)}_{22} + A^{(2)}_{22})} \tag{B5a}$$

$$k_2 = \omega q D^{(2)}_{22} \tag{B5b}$$

$$k_3 = \frac{A^{(2)}_{22}(h^{(2)}q + h^{(1)}) - 2B^{(1)}_{22}}{2(A^{(1)}_{22} + A^{(2)}_{22})} \tag{B5c}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$k_4 = \frac{(h^{(1)} + h^{(2)})A_{22}^{(2)} - 2B_{22}^{(1)}}{2(A_{22}^{(1)} + A_{22}^{(2)})} \tag{B5d}$$

$$k_5 = \frac{(h^{(1)} + h^{(2)})A_{22}^{(1)} + 2B_{22}^{(1)}}{2(A_{22}^{(1)} + A_{22}^{(2)})} \tag{B5c}$$

$$k_6 = \frac{\chi q}{k_5 q - k_1} \tag{B5f}$$

$$k_7 = -\frac{\chi}{2\omega(k_5 q - k_1)} \tag{B5g}$$

$$\delta_j = \frac{\lambda_j(q\gamma_j^{(2)} + k_1 p_j^{(2)})}{k_1 - k_5 q} \tag{B6a}$$

$$\rho_j = -\frac{\lambda_j(\gamma_j^{(2)} + k_5 p_j^{(2)})}{2\omega(k_1 - k_5 q)} \tag{B6b}$$

for $j = 1, 2, 3$.

$$\begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{Bmatrix} \bar{\alpha}_1 \\ \bar{\alpha}_2 \\ \bar{\alpha}_3 \end{Bmatrix} = \begin{Bmatrix} R_1 \\ R_2 \\ R_3 \end{Bmatrix} \tag{B7a}$$

$$\alpha_j = -\bar{\alpha}_j \frac{N}{A_{22} \cosh(\lambda_j(s - l))} \tag{B7b}$$

$$F_{1j}(l, s - l) = 2\rho_j(1 - q) \tanh(\omega l) + (p_j^{(1)} - p_j^{(2)}) \tanh(\lambda_j(s - l)) \tag{B8a}$$

$$F_{2j}(l, s - l) = (1 - q)\delta_j l - (p_j^{(1)} q - p_j^{(2)}) \tanh(\lambda_j(s - l)) \tag{B8b}$$

$$F_{3j}(l, s - l) = \eta_j^{(3)} \tag{B8c}$$

for $j = 1, 2, 3$.

$$\alpha_4 = \alpha_5 = \alpha_6 = \frac{N}{A_{22}} \tag{B9}$$

$$R_1 = 2(1 - q) \tanh(\omega l) k_7 \tag{B10a}$$

$$R_2 = (1 - q) k_6 l \tag{B10b}$$

$$R_3 = A_{22}^{(3)}. \tag{B10c}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

H. T. Banks
Center for Research in Scientific Computation
North Carolina State University,
Raleigh, NC 27695-8205

Zheng-Hua Luo
Department of Mechanical Engineering,
Nagaoka University of Technology,
Nagoka, Niigata 94021,
Japan

L. A. Bergman
Department of Aeronautical and Astronautical Engineering
University of Illinois,
Urbana, IL 61801

D. J. Inman
Department of Engineering Science and Mechanics,
Virginia Polytechnic Institute & State University
Blacksburg, VA 24061-0219

On the Existence of Normal Modes of Damped Discrete-Continuous Systems

In this paper we investigate a class of combined discrete-continuous mechanical systems consisting of a continuous elastic structure and a finite number of concentrated masses, elastic supports, and linear oscillators of arbitrary dimension. After the motion equations for such combined systems are derived, they are formulated as an abstract evolution equation on an appropriately defined Hilbert space. Our main objective is to ascertain conditions under which the combined systems have classical normal modes. Using the sesquilinear form approach, we show that unless some matching conditions are satisfied, the combined systems cannot have normal modes even if Kelvin-Voigt damping is considered.

1 Introduction

The vibration of distributed parameter structures is a subject of widespread and lasting interest in engineering, and there is a large literature on the subject contributed by both mathematically inclined theorists and practicing structural engineers. As is well known, the dynamical models for vibrating structures, given by partial differential equations together with appropriate boundary and initial conditions, are essentially infinite dimensional systems. For theoretical convenience and analysis, it is often assumed that the system does not possess damping. The concept of undamped systems not only serves a useful purpose for investigating the intrinsic properties of infinite dimensional systems but can also be loosely justified as an approximation in certain circumstances involving light damping and short time observation of vibrations. But in fact all physical systems do possess damping, no matter how small. Indeed, if a system (even in a vacuum) is set in motion and allowed to vibrate freely, the vibration will eventually die out; the rate of decay depends on the amount of damping. There are several mathematical models representing physical damping. The most often encountered types of damping in vibration studies are linear viscous damping and Kelvin-Voigt damping which are special cases of proportional damping. Viscous damping usually models external friction forces such as air resistance acting on the vibrating structures and is thus called "external damping," while Kelvin-Voigt damping originates from the internal friction of the material of the vibrating structures and is thus called "internal damping" or "material damping."

In system analysis, identification, and control, it is convenient and highly desirable that the original partial differential equations together with the boundary conditions be decomposable into an infinite number of *uncoupled* ordinary differential equations. When possible, this is usually done by using an eigenfunction expansion of the system stiffness operator. Indeed, this is the basic assumption underlying most engineering analyses and testing of complex physical systems, as well as a popular assumption in feedback control design. If a system has this decoupling property, we say that the system has "Normal Modes." For undamped systems, it is widely recognized that a continuous system usually has normal modes. For damped systems, however, the situation is much more complicated. For lumped systems consisting of mass, spring, and dampers, it has been shown that the commutability of the mass weighted stiffness operator and damping operator is a necessary and sufficient condition for the existence of normal modes (Caughey, 1960). These results are extended to continuous systems in Caughey and O'Kelley (1965) and to continuous systems with special asymmetries in Inman and Olsen (1988). In Meirovitch (1967), it is shown that a damped continuous mechanical system has normal modes if the damping operator is a linear combination of its mass operator and stiffness operator; this is physically a very special case. Mathematically, the condition for the existence of "normal modes" is that one be able to simultaneously diagonalize (through a coordinate transformation) the mass, stiffness, and damping operators of the system. This is the standard definition of normal modes as used in both vibration analysis and in vibration testing (Ewins, 1984).

In this paper we investigate a more general class of mechanical systems, consisting of a continuous structure and a finite number of concentrated masses, elastic supports, and oscillators. Such systems are often called "combined dynamical systems" in engineering literature. These include systems arising in the emerging field of smart material structures as well as oft-studied articulated structures (such as those arising in large space structures and robotics applications). Interest in such systems is likewise of long standing, but it has only been in the last several years that methods for the systematic and practical analysis of their vibration, both free and forced, have emerged (see, for example, Bergman and Nicholson, 1985a, b; Nicholson and

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Feb. 17, 1997; final revision, Aug. 13, 1997. Associate Technical Editor: W. K. Liu.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Bergman, 1985, 1986a, b; Bergman and McFarland, 1988a, b). While the parallels between the analyses of simple continuous structures and of these combined systems have been recognized for some time (Miles, 1956; Bishop and Johnson, 1960; Lin, 1962; Chen, 1963a, b; Pan, 1965; Gupta, 1970; Jacquot, 1970; Mead, 1971; Davies and Roger, 1979; Dowell, 1979; Kasprzyk, 1984; Banks, Gates, Rosen, and Wang, 1988), these have not, heretofore, been fully investigated. Since one quickly enters the realm of models with unbounded operators, the analysis is often quite delicate and difficult to carry out with rigor.

We shall be working with the weak form of system dynamic equations by using sesquilinear forms. We believe these are the most natural tools to use for the discussions here because the stiffness and damping operators are essentially differential operators containing delta functions and/or distributional derivatives. Earlier ad hoc approaches (Bergman and McFarland, 1988, 1990) to examination of such systems attempted to extend the commutativity of operators conditions (see also Inman and Olsen, 1988) directly to combined systems. Such an approach leads immediately to conceptual difficulties with domains of operators and interpretations of derivatives. To avoid such difficulties we employ a sesquilinear form approach in the context of weak or variational solutions. Using this approach, we shall derive some conditions under which such systems have normal modes. Before we begin work on the general model, we shall look at several special cases. One is a cantilever flexible beam system with piezoceramic patches (the standard elementary model in smart material studies). The other is a cantilever beam with an oscillator attached at some point. The first model arises from control of articulated multidimensional flexible structures using sensor/actuator configurations involving piezoceramic patches. In this case, the coefficients of the damping operator as well as the stiffness operator (and even the mass density) contain characteristic functions and thus are not continuous. We shall show that the system generally does not have normal modes even for Kelvin-Voigt damping. For the second system, we shall also show that the system does not possess normal modes except in very special cases where the beam and oscillator are "matched." Based on these observations, we next present a general model for a combined discrete and continuous system and discuss the conditions under which the overall system has normal modes. When Kelvin-Voigt damping is considered, it is found, much like the second case above, that the overall system does not possess normal modes unless some matching conditions are satisfied. These results are interesting because they reveal that a combined discrete-continuous system rarely possesses normal modes even if the continuous system itself does. Thus the usual engineering analyses that begin with the assumption of normal modes are simply not valid and hence not useful in analyzing these very important systems. Despite this, when the system is fully descretized in the usual way via the finite element or assumed modes method, the propensity is, indeed, to invoke the assumption of classical damping, accepting the likelihood of modeling error without a full understanding of its origin or effect.

2 Motivating Examples

To motivate the general cases, we first consider three special cases.

Example 1. As the most common example used in the "smart structures" literature, we consider a flat beam subject to transverse vibrations with a pair of piezoceramic patches attached for sensing and actuating. The beam, depicted in Fig. 1, is assumed to have length l, width b, thickness h, and cantilever end conditions with the fixed end at $x = 0$ and free end at $x = l$. It is assumed that the beam is homogeneous and constructed from a material which essentially satisfies the Euler-Bernoulli hypothesis for displacement and Kelvin-Voigt damping hypothesis (this means that damping is proportional to strain rate). To control the vibration of the beam, a pair of identical piezoceramic patches are bonded to opposite sides of the beam over the region $x_1 < x < x_2$. The Young's moduli, linear mass densities, and damping coefficients for the beam and patch are denoted by E, E_p, ρ, ρ_p and C_d, C_{dp}, respectively, with the subscript p denoting patch properties. The patch thickness is denoted by h_p.

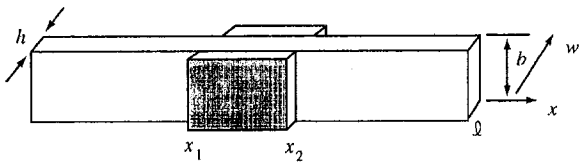


Fig. 1 Cantilevered beam with piezoceramic patches

Let $w(t, x)$ denote the transverse displacement of the beam at time t and at position x. From fundamental mechanics (force and moment balance for small disturbances and initial conditions), we know that the dynamic equation governing the free vibrations (no control action) of the beam is given by (Banks, Smith, and Wang, 1996)

$$\tilde{\rho}\frac{\partial^2 w}{\partial t^2} + \frac{\partial^3}{\partial t \partial x^2}\left(\widetilde{C_d I}(x)\frac{\partial^2 w}{\partial x^2}\right) + \frac{\partial^2}{\partial x^2}\left(\widetilde{EI}(x)\frac{\partial^2 w}{\partial x^2}\right) = 0$$

$$w(t, 0) = \frac{\partial w}{\partial x}(t, 0) = \frac{\partial^2 w(t, l)}{\partial x^2} = \frac{\partial^3 w(t, l)}{\partial x^3} = 0 \quad (1)$$

where the derivatives are interpreted in the sense of distributions (e.g., see Banks, Smith, and Wang (1996) for details as well as experimental validation of the appropriateness of this model for our example) and

$$\tilde{\rho} = \rho h b + 2b\rho_p h_p \chi_p(x)$$

$$\widetilde{C_d I}(x) = C_d I + \frac{2b}{3} C_{dp} a \chi_p(x)$$

$$\widetilde{EI}(x) = EI + \frac{2b}{3} E_p a \chi_p(x)$$

with $I = (b^3 b/12)$, $a = (h/2 + h_p)^3 - h^3/8$, and $\chi_p(x)$ is the characteristic function for $[x_1, x_2]$.

It is clear that $\widetilde{EI}(x)$ and $\widetilde{C_d I}(x)$ are discontinuous as a function of x. Define $V = H_L^2(0, l) = \{\phi \in H^2(0, l) | \phi(0) = \phi'(0) = 0\}$, $H = L^2(0, l)$ and assume the Gelfand triple construction $V \subset H \subset V^*$ (see Wloka, 1992). Here $H^2(0, l)$ is the usual complex Sobolov space $W^{2,2}(0, l)$. Then the stiffness operator $A: V \to V^*$ and damping operator $B: V \to V^*$ are given by

$$A = \frac{\partial^2}{\partial x^2}\left(\widetilde{EI}(x)\frac{\partial^2}{\partial x^2}\right), \quad B = \frac{\partial^2}{\partial x^2}\left(\widetilde{C_d I}(x)\frac{\partial^2}{\partial x^2}\right),$$

where the derivatives are interpreted in the distributional sense. Or, more precisely $(A\phi)(\psi) = \langle \widetilde{EI}\phi'', \psi'' \rangle$, $(B\phi)(\psi) = \langle \widetilde{C_d I}\phi'', \psi'' \rangle$ for $\phi, \psi \in V$, $\langle \cdot, \cdot \rangle$ is the complex inner product in $L^2(0, l)$ and prime denotes spatial differentiation.

The system (1) can be written thus as (again see Banks, Smith, and Wang (1996))

$$M\ddot{y}(t) + B\dot{y}(t) + Ay(t) = 0 \text{ in } V^* \quad (2)$$

where M is $\tilde{\rho} Id$ with Id the identity on V (or H or V^*). For normal modes one must have a complete orthonormal family $\{\phi_n\}$ satisfying

$$\langle M\phi_n, \phi_m \rangle = \delta_{nm}$$

$$(A\phi_n)(\phi_m) = \lambda_n \delta_{nm}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$(B\phi_n)(\phi_m) = \tilde{\lambda}_n \delta_{nm}$$

where δ_{nm} is the usual Kronecker delta. In particular, we must have for all n, m

$$\rho hb\langle \phi_n, \phi_m \rangle + 2\rho_p h_p b \int_{x_1}^{x_2} \phi_n \bar{\phi}_m = \delta_{nm}$$

$$EI\langle \phi_n'', \phi_m'' \rangle + \frac{2b}{3} E_p a \int_{x_1}^{x_2} \phi_n'' \bar{\phi}_m'' = \lambda_n \delta_{nm}$$

$$C_d I\langle \phi_n'', \phi_m'' \rangle + \frac{2b}{3} C_{dp} a \int_{x_1}^{x_2} \phi_n'' \bar{\phi}_m'' = \tilde{\lambda}_n \delta_{nm}.$$

The first equation just states that $\{\phi_n\}$ is an orthonormal set in $L^2(0, l)$ with the weighted inner product $\langle \tilde{\rho} \cdot, \cdot \rangle$. Defining $\hat{E} = EI$, $\hat{C}_d = C_d I$, $\hat{E}_p = (2b/3)E_p a$, $\hat{C}_{dp} = (2b/3)C_{dp} a$, we may write the second and third equations as

$$\mathcal{K}\begin{pmatrix} \langle \phi_n'', \phi_m'' \rangle \\ \int_{x_1}^{x_2} \phi_n'' \phi_m'' \end{pmatrix} = \begin{pmatrix} \lambda_n \delta_{nm} \\ \lambda_n \delta_{nm} \end{pmatrix} \tag{3}$$

where $\mathcal{K} = \begin{bmatrix} \hat{E} & \hat{E}_p \\ \hat{C}_d & \hat{C}_{dp} \end{bmatrix}$. Thus, if $\mathcal{K}$ is invertible, that is, if we do not have existence of a $k \neq 0$ such that $(\hat{E}, \hat{E}_p) = k(\hat{C}_d, \hat{C}_{dp})$, we find that

$$\begin{pmatrix} \langle \phi_n'', \phi_m'' \rangle \\ \int_{x_1}^{x_2} \phi_n'' \phi_m'' \end{pmatrix} = \mathcal{K}^{-1}\begin{pmatrix} \lambda_n \delta_{nm} \\ \tilde{\lambda}_n \delta_{nm} \end{pmatrix} = \begin{pmatrix} \gamma_n \delta_{nm} \\ \tilde{\gamma}_n \delta_{nm} \end{pmatrix} \tag{4}$$

for all n, m. In particular, this implies that $\langle \tilde{\rho}\phi_n, \phi_m \rangle = \delta_{nm}$, $\langle \phi_n'', \phi_m'' \rangle = \gamma_n \delta_{nm}$ and the $\{\phi_n\}$ satisfy the boundary conditions for an undamped beam with mass density $\tilde{\rho}$ and stiffness $EI = 1$ (such a family is completely determined by the beam parameters and boundary conditions in the usual manner). However, the second equation arising from (4) yields that this set $\{\phi_n\}$ must also satisfy $\int_{x_1}^{x_2} \phi_n'' \bar{\phi}_m'' = 0$ for $n \neq m$, a condition that does not in general hold for an arbitrary patch on $[x_1, x_2] \subset (0, l)$. Hence the assumption of invertibility of $\mathcal{K}$ leads to a contradiction.

Thus if we are to have a set of normal modes for the cantilevered beam with patches, we must have that $(\hat{E}, \hat{E}_p) = k(\hat{C}_d, \hat{C}_{dp})$ for some $k \neq 0$. This condition (while an extremely rare and physically unreasonable condition in general) is clearly also sufficient for normal modes. Thus we have argued the following.

Theorem 1. *A necessary and sufficient condition that the system (1) have normal modes is that there exist $k > 0$ such that $(\hat{E}, \hat{E}_p) = k(\hat{C}_d, \hat{C}_{dp})$ where $\hat{E} = EI$, $\hat{E}_p = (2b/3)E_p a$, $\hat{C}_d = C_d I$, $\hat{C}_{dp} = (2b/3)C_{dp} a$.*

It is interesting to observe that while the standard "smart structure," i.e., system (1) with piezoceramic patches does not in general possess normal modes, the system without the patches and with homogeneous material does have normal modes when Kelvin-Voigt damping is present.

Example 2. Consider a cantilever beam with an oscillator attached to the beam at point α, as illustrated in Fig. 2. Again, consider Kelvin-Voigt damping. Let the notations for the beam be the same as in Example 1, except that now $\tilde{EI} = EI$ and $\tilde{C_d I} = C_d I$ are constants.

Let the mass, damping, and spring coefficients of the oscillator be denoted by m_H, c_H and k_H, respectively. Let $w(t, x)$ denote the displacement of the beam and let $z(t)$ denote the displacement of the mass of the oscillator. It is assumed that

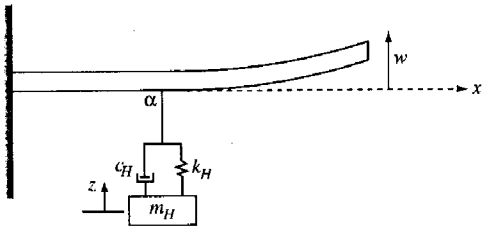


Fig. 2 Cantilevered beam with oscillator attached

both are measured from the equilibrium state. Then the dynamical system of the composite structure is given by

$$\rho \ddot{w}(t, x) + C_d I \dot{w}''''(t, x) + EIw''''(t, x) = [c_H(\dot{z}(t) - \dot{w}(t, x)) + k_H(z(t) - w(t, x))]\delta(x - \alpha), \quad 0 < x < l,$$

$$m_H \ddot{z}(t) + c_H(\dot{z}(t) - \dot{w}(t, \alpha)) + k_H(z(t) - w(t, \alpha)) = 0,$$

$$w(0, x) = w_0(x), \dot{w}(0, x) = w_1(x), z(0) = z_0, \dot{z}(0) = z_1, \tag{5}$$

where $\delta(x - \alpha)$ is the Dirac Delta distribution, and hence $\phi(\alpha) = \int_0^l \phi(x)\delta(x - \alpha)dx$ for $\phi \in H^1(0, l)$, and w_0, w_1, z_0, z_1 are given initial data.

To proceed further, we define a complex Hilbert space $H = L_2(0, l) \times \mathbb{C}$ with inner product

$$\langle y_1, y_2 \rangle = \int_0^l w_1(x)\bar{w}_2(x)dx + z_1\bar{z}_2,$$

$$\text{for} \quad y_1 = (w_1, z_1), y_2 = (w_2, z_2) \in H.$$

Define a Hilbert space $V = \{(w, z) \in H^2(0, l) \times \mathbb{C} | w(0) = w'(0) = 0\}$ with inner product induced norm

$$\|(w, z)\|_V^2 = \int_0^l w''(x)\bar{w}''(x)dx + (z - w(\alpha))(\bar{z} - \bar{w}(\alpha)).$$

Let $V \hookrightarrow H \hookrightarrow V^*$ again be the usual Gelfand triple formulation.

Define a stiffness operator $A: V \to V^*$ by

$$A\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} EIw''''(x) + k_H(w(x) - z)\delta(x - \alpha) \\ k_H(z - w(\alpha)) \end{pmatrix}$$

and a damping operator $B: V \to V^*$ by

$$B\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} C_d Iw''''(x) + c_H(w(x) - z)\delta(x - \alpha) \\ c_H(z - w(\alpha)) \end{pmatrix},$$

where again the derivatives and Delta operator are interpreted in the distributional sense. Then Eq. (5) can be written as the abstract second-order initial value problem

$$M\ddot{y}(t) + B\dot{y}(t) + Ay(t) = 0 \text{ in } V^*, t > 0,$$

$$y(0) = y_0, \dot{y}(0) = y_1 \tag{6}$$

where $y = (w, z)^T$ and $My = (\rho w, m_H z)^T$, $y_0 = (w_0, z_0)$, $y_1 = (w_1, z_1)$.

Let $\sigma_A: V \times V \to \mathbb{C}$ be the symmetric sesquilinear form given by

$$\sigma_A(y_1, y_2) = EI \int_0^l w_1''(x)\bar{w}_2''(x)dx + k_H(w_1(\alpha) - z_1)(\bar{w}_2(\alpha) - \bar{z}_2) \tag{7}$$

for $y_1 = (w_1, z_1)$, $y_2 = (w_2, z_2) \in V$.

Similarly, let σ_B be the sesquilinear form which has exactly the same form as σ_A except the constants EI and k_H are replaced, respectively, by $C_d I$ and c_H. At this point, it is easy to show that

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\sigma_A(y_1, y_2) = \langle Ay_1, y_2\rangle_{V^*,V} \quad \text{and} \quad \sigma_B(y_1, y_2) = \langle By_1, y_2\rangle_{V^*,V}$$

where $\langle \cdot, \cdot \rangle_{V^*,V}$ denotes the usual duality pairing (Banks, Smith, and Wang, 1996; Wloka, 1992). Furthermore, we have the following lemma.

Lemma 1. For any $y_1, y_2 \in V$, there exist positive constants γ and c such that

(*i*) $|\sigma_A(y_1, y_2)| \leq c\|y_1\|_V\|y_2\|_V$, i.e., σ_A is V-continuous.
(*ii*) $\sigma_A(y_1, y_1) \geq \gamma\|y_1\|_V^2$, i.e., σ_A is V-elliptic.

The proof of this lemma is obvious and is thus omitted.

Similarly, we can show that σ_B has the same properties as stated in Lemma 1.

Using these notations, we are able to write Eq. (5) as

$$\langle M\ddot{y}(t), \Phi\rangle + \sigma_A(y(t), \Phi) + \sigma_B(y(t), \Phi) = 0, \quad \text{for all} \quad \Phi \in V. \quad (8)$$

This is the precise meaning of the equation in (6) which is called the weak form or variational form of system (5). We can rewrite Eq. (8) as an equivalent first order vector system, or leave it in second-order form. In either case, there are basic abstract results (see Chapter 4 of Banks, Smith, and Wang (1996)) for such systems that guarantee the existence of a unique solution in an appropriate sense. We are more interested here in the conditions under which there exist normal modes.

Theorem 2. There exist normal modes for system (6) if and only if there exists a constant $k > 0$ such that

$$\begin{pmatrix} EI \\ k_H \end{pmatrix} = k\begin{pmatrix} C_dI \\ c_H \end{pmatrix}. \quad (9)$$

Proof: Condition (9) is obviously sufficient. We need only to show the necessity. Let Φ_n be the generalized orthonormal eigenvectors of the operator A such that

$$\langle M\Phi_n, \Phi_m\rangle = \delta_{nm}, \quad \text{for all} \quad n, m.$$

That is, the $\{\Phi_n\}$ simultaneously diagonalize M and A. It is shown in the Appendix how to *uniquely* solve this eigenvalue problem. If there exist normal modes for (6), then it is necessary that these $\{\Phi_n\}$ also diagonalize B, that is,

$$\sigma_A(\Phi_n, \Phi_m) = 0, \ \sigma_B(\Phi_n, \Phi_m) = 0, \quad \text{for all} \quad \Phi_n = (w_n, z_n), \quad \Phi_m = (w_m, z_m), \quad \text{for all} \quad n \neq m.$$

This means that

$$EI\int_0^l w_n''(x)\bar{w}_m''(x)dx + k_H(z_n - w_n(\alpha))(\bar{z}_m - \bar{w}_m(\alpha)) = 0,$$

$$C_dI\int_0^l w_n''(x)\bar{w}_m''(x)dx + c_H(z_n - w_n(\alpha))(\bar{z}_m - \bar{w}_m(\alpha)) = 0,$$

for all $n \neq m$. Suppose condition (9) does not hold. Then from these two equations, we have

$$\int_0^l w_n''(x)\bar{w}_m''(x)dx = 0$$

$$(z_n - w_n(\alpha))(\bar{z}_m - \bar{w}_m(\alpha)) = 0,$$

for all $n \neq m$, which implies $z_n = w_n(\alpha)$ or $z_m = w_m(\alpha)$, for all $n \neq m$. Therefore, there exists at most one integer n_1 such that $z_{n_1} \neq w_{n_1}(\alpha)$ and $z_n = w_n(\alpha)$ for all n except $n = n_1$. Since $\Phi_n(x) = \begin{pmatrix} w_n(x) \\ z_n \end{pmatrix}$ are the eigenfunctions of A which also diagonalize M, the equation

$$A\Phi_n(x) = \lambda_n M\Phi_n(x)$$

must be satisfied for some scalar λ_n. This is equivalent to

$$EIw_n''''(x) + k_H(w_n(\alpha) - z_n)\delta(x - \alpha) = \rho\lambda_n w_n(x)$$

$$k_H(z_n - w_n(\alpha)) = m_H\lambda_n z_n$$

$$w_n(0) = w_n'(0) = w_n''(l) = w_n'''(l) = 0. \quad (10)$$

Since $z_n = w_n(\alpha)$ for $n \neq n_1$, the second equation above gives $z_n = 0$ for all $n \neq n_1$ because $\lambda_n \neq 0$. Hence $w_n(\alpha) = 0$ for all $n \neq n_1$. Thus, Eq. (10) reduces to

$$EIw_n''''(x) = \rho\lambda_n w_n(x)$$

$$w_n(x) = w_n'(0) = w_n''(l) = w_n'''(l) = 0, \ n \neq n_1,$$

which is exactly the eigenvalue problem for a cantilever beam with density ρ, stiffness EI without any attached oscillators. Such a set $\{w_n\}$ is determined uniquely (independent of α) and does not satisfy $w_n(\alpha) = 0$ for all $n \neq n_1$.

Thus condition (9) is necessary for the beam with oscillator system to have normal modes. When condition (9) is satisfied, we say that the beam and the oscillator are "matched." Since they are rarely matched for arbitrarily given oscillators, the overall system rarely has normal modes.

Remark: Consider the special case where the oscillator is attached at the tip end, i.e., $\alpha = l$. In this case, the dynamic equations can be written as follows (using force balance notions on the states $y(t) = (w(t, x), w(t, l), z(t))$):

$$\rho\ddot{w}(t, x) + C_dI\dot{w}''''(t, x) + EIw''''(t, x) = 0, \quad 0 < x < l,$$

$$-C_dI\dot{w}'''(t, l) - EIw'''(t, l) = c_H(\dot{z}(t) - \dot{w}(t, l)) + k_H(z(t) - w(t, l))$$

$$m_H\ddot{z}(t) + c_H(\dot{z}(t) - \dot{w}(t, l)) + k_H(z(t) - w(t, l)) = 0. \quad (11)$$

The second equation above can be formally obtained by integrating the first line of Eq. (5) from l^- to l^+, noting that $w'''(t, l^+) = 0$, $\int_{l^-}^{l^+} \ddot{w}(t, x)dx = 0$, and then setting $l^- = l$. This model is extensively studied in existing literature (see, for example, Banks, Gates, Rosen, and Wang (1988)). It should be noted that there exists an essential difference between Eq. (5) and Eq. (11). In Eq. (11), $w(t, l)$ is an independent variable. That is to say, $(w(t, x), w(t, l), z(t))$ is chosen as state variable if we wish to write Eq. (11) in the abstract form as before. This is in sharp contrast with Eq. (5) where $w(t, \alpha)$ is not an independent variable. In that case only $y(t) = (w(t, x), z(t))$ was chosen as the state variable. Since the mathematical formulation is thus somewhat different for discrete elements attached to the free end of a beam, we shall assume that the discrete elements such as oscillators are attached only at the interior points of continuous structures in the next section.

In Inman (1989, p. 281) it is shown formally (not rigorously) that the modal analysis of the system of Fig. 2 "effectively couples the design problem so that passive control cannot be performed on a per mode basis." It is also noted in Inman (1989) from an analysis of Bergman and Nicholson (1985b) that proportionality results only if $\lambda_n\lambda_m$ = constant for all m, n. This, of course, is not possible with the exception of perhaps one mode. In Bellos (1989) it is shown experimentally that in fact only one mode of the system of Fig. 2 exhibits a normal mode behavior.

Example 3. Let us now consider the situation where an elastic support is attached to the beam at $x = \beta$. Let k_e and c_e denote, respectively, the spring and damping coefficients of the elastic support. Using the notations defined in Example 2, we

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

are able to write the dynamical equations governing the free vibrations of the beam as follows:

$$\rho\ddot{w}(t, x) + C_dI\dot{w}''''(t, x) + EIw''''(t, x) + k_ew(t, x)\delta(x - \beta) + c_e\dot{w}(t, x)\delta(x - \beta) = 0, \quad 0 < x < l.$$

Let $V = \{w \in H^2(0, l) | w(0) = w'(0) = 0\}$. Define $A: V \to V^*$ and $B: V \to V^*$ by

$$Aw(x) = EIw''''(x) + k_ew(x)\delta(x - \beta),$$

$$Bw(x) = C_dIw''''(x) + c_ew(x)\delta(x - \beta).$$

Let σ_A and σ_B be the sesquilinear forms associated with A and B. We show that for the system to have normal modes, it is necessary and sufficient that there exists a constant k such that

$$\begin{pmatrix} EI \\ k_e \end{pmatrix} = k\begin{pmatrix} C_dI \\ c_e \end{pmatrix}.$$

The proof of this claim is similar to but simplier than that in Example 2. So we only provide an outline. Let $\{w_n\}$ be the generalized orthogonal eigenfunctions of A. To have normal modes, it is necessary that

$$\sigma_A(w_n, w_m) = 0,$$

$$\sigma_B(w_n, w_m) = 0, \quad \text{for all} \quad n \neq m,$$

which means

$$EI\int_0^l w_n''(x)\bar{w}_m''(x)dx + k_ew_n(\beta)\bar{w}_m(\beta) = 0,$$

$$C_dI\int_0^l w_n''(x)\bar{w}_m''(x)dx + c_ew_n(\beta)\bar{w}_m(\beta) = 0,$$

$$\text{for all} \quad n \neq m.$$

Therefore, if $(EI, k_e) \neq k(C_dI, c_e)$, then

$$\int_0^l w_n''(x)\bar{w}_m''(x)dx = 0,$$

$$w_n(\beta)\bar{w}_m(\beta) = 0,$$

which implies $w_n(\beta) = 0$ for all n, except possibly one integer n_1. But then the eigenvalue problem $Aw_n(x) = \lambda_n w_n(x)$ is reduced to

$$EIw_n''''(x) = \lambda_n w_n(x),$$

$$w_n(0) = w_n'(0) = w_n''(l) = w_n'''(l) = 0, \quad n \neq n_1,$$

which is the eigenvalue problem for a free beam without any attachments. For the latter, the eigenfunctions are already determined and they cannot satisfy the constraint $w_n(\beta) = 0$ for all $n \neq n_1$ and arbitrary β. This contradiction shows that $(EI, k_e) = k(C_dI, c_e)$ is necessary for the system to have normal modes.

3 Normal Modes of Combined Discrete-Continuous Systems

In this section we generalize the findings of the specific examples of the previous section to treat combined systems that include a multidimensional distributed parameter structural member, such as a beam, plate or shell, subject to transverse vibrations, to which a finite number of discrete systems have been attached. Other vibrating structures, such as shafts in torsion, are readily treated by the ideas to be presented here.

We consider a typical two-dimensional planar member, occupying a region denoted by Ω with a boundary of $\partial\Omega$ arbitrary shape. This structure, which has area mass density ρ, is assumed to undergo transverse displacements $w(t, x_1, x_2)$ that can be described by the usual linear theory of infinitesimal elasticity. We further assume the following.

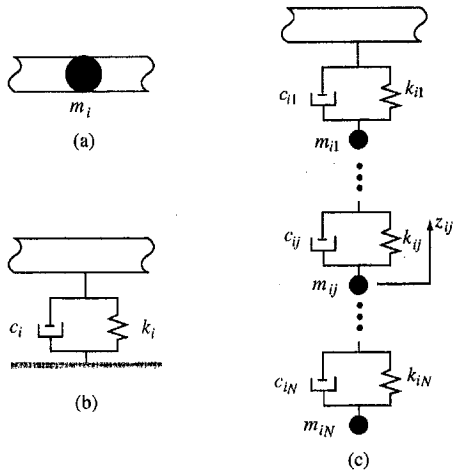


Fig. 3 Examples of attached discrete systems

Assumption (H1): Some part of the boundary is clamped and the natural boundary conditions on the remaining part are such that the stiffness operator is self-adjoint in some appropriate Hilbert spaces to be defined below.

This is a reasonable assumption for a significant number of realistical physical systems. To treat the problem in a unified form, we further assume in the following that there are no attachments of the discrete subsystems on the boundary. However, as noted in the remark in the last section, the case where attachments do exist on the boundary can also be treated.

A total of $\mathcal{N}$ discrete subsystems is attached to the continuous member at the points ξ_i, $i = 1, \ldots, \mathcal{N}$ in Ω. Specifically, let $\mathcal{N}_e$ be the number of elastic supports (linear springs), $\mathcal{N}_m$ the number of concentrated masses, and $\mathcal{N}_o$ the number of lumped parameter linear oscillators in the combined system, so that $\mathcal{N} = \mathcal{N}_e + \mathcal{N}_m + \mathcal{N}_o$. Examples of such discrete systems are depicted in Fig. 3. The ith elastic support has stiffness k_i and includes a dashpot with viscous damping coefficient c_i; the ith concentrated mass is of mass m_i; and the ith discrete oscillator has $N = N(i)$ degrees-of-freedom, with stiffness k_{ij}, damping coefficients c_{ij}, and masses m_{ij}, $j = 1, \ldots, N(i)$. The displacement of the jth mass of the ith oscillator is denoted by $z_{ij}(t)$, and the displacements of all the masses of this oscillator can be written as the vector

$$z_i(t) = (z_{i1}(t), \ldots, z_{iN(i)}(t))^T, \quad i = 1, \ldots, \mathcal{N}_o.$$

Since in our discussions we are only concerned with the existence of normal modes of the basic system, we do not assume that any specific external forces are applied to the system.

Let $w(t, x)$ denote the displacement of the continuous structure at time t and position x. Let L_K and L_C denote, respectively, the stiffness and damping operators of the continuous structure. We assume the following.

Assumption (H2): L_C is proportional to L_K.

We assume this because we already know the continuous system alone has normal modes under this condition and our objective here is to see what happens for the combined discrete-continuous systems under this same condition. It is clear that if we consider Kelvin-Voigt damping in a homogeneous continuous structure, then the above assumption is satisfied automatically. Kelvin-Voigt damping is a proportional strong damping that is most likely to allow the system to have normal modes. In other words, if a system does not have normal modes when

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Kelvin-Voigt damping is considered, then it is most unlikely to have normal modes for other kinds of damping.

Using these notations, the free vibration of the continuous member is governed by the equation

$$\rho\ddot{w}(t, x) + L_C\dot{w}(t, x) + L_K w(t, x) = \sum_{i=1}^{\mathcal{N}} F_i(t)\delta(x - \xi_i), \quad (12)$$

where $F_i(t)$ represent reaction forces exerted on the continuous structure by the discrete subsystems. For an elastic support, they are given by

$$F_i(t) = -k_i w(t, \xi_i) - c_i\dot{w}(t, \xi_i); \quad (13)$$

for a concentrated mass,

$$F_i(t) = -m_i\ddot{w}(t, \xi_i) \quad (14)$$

and for an oscillator,

$$F_i(t) = k_{i1}[z_{i1}(t) - w(t, \xi_i)] + c_{i1}[\dot{z}_{i1}(t) - \dot{w}(t, \xi_i)] = -\sum_{j=1}^{N(i)} m_{ij}\ddot{z}_{ij}(t). \quad (15)$$

The dynamical equations of the ith oscillator subsystems are given by

$$\begin{aligned}
&m_{i1}\ddot{z}_{i1}(t) + c_{i1}[\dot{z}_{i1}(t) - \dot{w}(t, \xi_i)] + c_{i2}[\dot{z}_{i1}(t) - \dot{z}_{i2}(t)] \\
&\quad + k_{i1}[z_{i1}(t) - w(t, \xi_i)] + k_{i2}[z_{i1}(t) - z_{i2}(t)] = 0 \\
&m_{ij}\ddot{z}_{ij}(t) + c_{ij}[\dot{z}_{ij}(t) - \dot{z}_{i(j-1)}(t)] + c_{i(j+1)}[\dot{z}_{ij}(t) - \dot{z}_{i(j+1)}(t)] \\
&\quad + k_{ij}[z_{ij}(t) - z_{i(j-1)}] + k_{i(j+1)}[z_{ij}(t) - z_{i(j+1)}(t)] = 0, \\
&\qquad \vdots \\
&m_{iN(i)}\ddot{z}_{iN(i)}(t) + c_{iN(i)}[\dot{z}_{iN(i)}(t) - \dot{z}_{i(N(i)-1)}(t)] \\
&\quad + k_{iN(i)}[z_{iN(i)}(t) - z_{i(N(i)-1)}(t)] = 0. \quad (16)
\end{aligned}$$

Motivated by Example 2, we wish to write the coupled Eqs. (12) and (16) in a compact form similar to (6). For this purpose, let us define a Hilbert space

$$H = L^2(\Omega) \times \mathbb{C}^{N(1)} \times \cdots \times \mathbb{C}^{N(\mathcal{N}_o)}$$

with inner products

$$\langle U, Y\rangle = \int_\Omega u(x)\bar{y}(x)dx + u_1\bar{y}_1 + \cdots + u_{\mathcal{N}_o}\bar{y}_{\mathcal{N}_o}$$

for $U(x) = (u(x), u_1, \ldots, u_{\mathcal{N}_o})$, $Y(x) = (y(x), y_1, \ldots, y_{\mathcal{N}_o}) \in H$. We also define operators $\hat{\rho}$ and $\hat{L}_K$ by

$$\hat{\rho}U(x) = \begin{pmatrix} [\rho + \sum_{i=1}^{\mathcal{N}_m} m_i\delta(x - \xi_i)]u(x) \\ M_1u_1 \\ \vdots \\ M_{\mathcal{N}_o}u_{\mathcal{N}_o} \end{pmatrix},$$

and

$$\hat{L}_K U(x) = \begin{pmatrix} L_K u(x) + \sum_{i=1}^{\mathcal{N}_o} k_{i1}(u(\xi_i) - u_{i1})\delta(x - \xi_i) + \sum_{i=1}^{\mathcal{N}_e} k_i u(\xi_i)\delta(x - \xi_i) \\ K_1u_1 \\ \vdots \\ K_{\mathcal{N}_o}u_{\mathcal{N}_o} \end{pmatrix},$$

where

$$M_i = \text{diag}\,(m_{ij}), \quad j = 1, \ldots, N(i), \quad i = 1, \ldots, \mathcal{N}_o$$

and

$$K_iu_i = \begin{pmatrix} k_{i1}(u_{i1} - u(\xi_i)) + k_{i2}(u_{i1} - u_{i2}) \\ \vdots \\ k_{iN(i)}(u_{iN(i)} - u_{i(N(i)-1)}) \end{pmatrix}, \quad i = 1, \ldots, \mathcal{N}_o.$$

The operator $\hat{L}_C$ can also be defined just in the same way as $\hat{L}_K$ except that L_K and k_{ij} in $\hat{L}_K$ should be replaced by L_C and c_{ij}, respectively.

With these definitions, we can formally write Eqs. (12)–(16) in an abstract form on the Hilbert space H by

$$\hat{\rho}\ddot{W}(t) + \hat{L}_C\dot{W}(t) + \hat{L}_K W(t) = 0, \quad (17)$$

for $W(t) = (w(t, \cdot), z_1(t), \ldots, z_{\mathcal{N}_o}(t))$.

Next, let us define a Hilbert space

$$V = H_L^2(\Omega) \times \mathbb{C}^{N(1)} \times \cdots \times \mathbb{C}^{N(\mathcal{N}_o)}$$

where $H_L^2(\Omega)$ is a space of all functions which are twice differentiable in $L^2(\Omega)$ and satisfy the clamped boundary conditions on $\partial\Omega$. The inner product in V can be defined as a summation of the inner product in $H_L^2(\Omega)$ and the inner products in $\mathbb{C}^{N(i)}$, $i = 1, \ldots, \mathcal{N}_o$.

For $U(x) = (u(x), u_1, \ldots, u_{\mathcal{N}_o})$, $Y(x) = (y(x), y_1, \ldots, y_{\mathcal{N}_o})$, we define a sesquilinear form by

$$\begin{aligned}
\hat{\sigma}_K(U, Y) = \langle\hat{L}_K U, Y\rangle &= \sigma_K(u, y) + \sum_{i=1}^{\mathcal{N}_e} k_i u(\xi_i)\bar{y}(\xi_i) \\
&\quad + \sum_{i=1}^{\mathcal{N}_o} k_{i1}(u(\xi_i) - u_{i1})\bar{y}(\xi_i) \\
&\quad + (K_1u_1)\bar{y}_1 + \cdots + (K_{\mathcal{N}_o}u_{\mathcal{N}_o})\bar{y}_{\mathcal{N}_o} \\
&= \sigma_K(u, y) + \sum_{i=1}^{\mathcal{N}_e} k_i u(\xi_i)\bar{y}(\xi_i) \\
&\quad + \sum_{i=1}^{\mathcal{N}_o} k_{i1}(u_{i1} - u(\xi_i))(\bar{y}_{i1} - \bar{y}(\xi_i)) \\
&\quad + \sum_{i=1}^{\mathcal{N}_o} k_{i2}(u_{i2} - u_{i1})(\bar{y}_{i2} - \bar{y}_{i1}) \\
&\quad \vdots \\
&\quad + \sum_{i=1}^{\mathcal{N}_o} k_{iN(i)}(u_{iN(i)} - u_{i(N(i)-1)})(\bar{y}_{iN(i)} - \bar{y}_{i(N(i)-1)})
\end{aligned}$$

where $\sigma_K(\cdot, \cdot)$ is the sesquilinear form associated with L_K.

Similarly, we can define an sesquilinear form associated with the damping operator $\hat{L}_C$ as follows:

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\hat{\sigma}_C(U, Y) = \langle \hat{L}_C U, Y\rangle = \sigma_C(u, y) + \sum_{i=1}^{\mathcal{N}_e} c_i u(\xi_i)\bar{y}(\xi_i) + \sum_{i=1}^{\mathcal{N}_o} c_{i1}(u_{i1} - u(\xi_i))(\bar{y}_{i1} - \bar{y}(\xi_i)) + \sum_{i=1}^{\mathcal{N}_o} c_{i2}(u_{i2} - u_{i1})(\bar{y}_{i2} - \bar{y}_{i1}) \vdots + \sum_{i=1}^{\mathcal{N}_o} c_{iN(i)}(u_{iN(i)} - u_{i(N(i)-1)})(\bar{y}_{iN(i)} - \bar{y}_{i(N(i)-1)})$$

where $\sigma_C(\cdot, \cdot)$ is the sesquilinear form associated with the operator L_C.

With these definitions, we are able to write Eq. (17) in the usual weak form:

$$\langle \hat{\rho}\ddot{W}(t), \Phi\rangle + \hat{\sigma}_C(\dot{W}(t), \Phi) + \hat{\sigma}_K(W(t), \Phi) = 0 \quad \text{for all} \quad \Phi \in V. \tag{18}$$

Since the boundary conditions for the continuous structure are assumed to be in such a form that the operators L_K and L_C are self-adjoint, it is easy to see that σ_K and σ_C are symmetric. Therefore, from the above definitions, it is easily verified that both $\hat{\sigma}_K$ and $\hat{\sigma}_C$ are symmetric, V-continuous and V-elliptic, from which the existence of a unique solution of Eq. (18) taken with appropriate initial conditions easily follows (again, see Chapter 4 of Banks, Smith, and Wang (1996)).

Our primary concern is to use these sesquilinear forms to derive conditions under which there exist normal modes for system (17) (or (18)).

Theorem 3. *Under assumptions (H1) and (H2), there exist normal modes for the discrete-continuous system (17) if and only if there exists a constant $k > 0$ such that*

$$p_K = kp_C \tag{19}$$

where

$$p_K = (1, k_1, \ldots, k_{\mathcal{N}_e}, k_{11}, \ldots, k_{1N(1)}, \ldots, k_{\mathcal{N}_o N(\mathcal{N}_o)})$$

and

$$p_C = (\gamma, c_1, \ldots, c_{\mathcal{N}_e}, c_{11}, \ldots, c_{1N(1)}, \ldots, c_{\mathcal{N}_o N(\mathcal{N}_o)}).$$

Proof: The sufficiency is obvious. We need only to show the necessity. Let $\Phi_n(x) = (w^n(x), z_1^n, \ldots, z_{\mathcal{N}_o}^n)^T$ be the generalized eigenfunctions of the operator $\hat{L}_K$ such that

$$\langle \hat{\rho}\Phi_n, \Phi_m\rangle = \delta_{nm}, \quad \text{for all} \quad n, m.$$

These eigenfunctions can be obtained using the same procedures as given in the Appendix. In order to have normal modes, it is necessary that

$$\hat{\sigma}_K(\Phi_n, \Phi_m) = 0$$
$$\hat{\sigma}_C(\Phi_n, \Phi_m) = 0 \tag{20}$$

holds for any $n \neq m$. Assumption (H2) implies that

$$\sigma_C(w^n, w^m) = \gamma\sigma_K(w^n, w^m)$$

for a constant γ. Thus Eq. (20) is equivalent to the following equations:

$$x_0^{nm} + \sum_{i=1}^{\mathcal{N}_e} k_i x_i^{nm} + \sum_{i=1}^{\mathcal{N}_o} k_{i1} y_{i1}^{nm} + \sum_{i=1}^{\mathcal{N}_o} k_{i2} y_{i2}^{nm} + \cdots + \sum_{i=1}^{\mathcal{N}_o} k_{iN(i)} y_{iN(i)}^{nm} = 0 \tag{21}$$

$$\gamma x_0^{nm} + \sum_{i=1}^{\mathcal{N}_e} c_i x_i^{nm} + \sum_{i=1}^{\mathcal{N}_o} c_{i1} y_{i1}^{nm} + \sum_{i=1}^{\mathcal{N}_o} c_{i2} y_{i2}^{nm} + \cdots + \sum_{i=1}^{\mathcal{N}_o} c_{iN(i)} y_{iN(i)}^{nm} = 0 \tag{22}$$

where

$$x_0^{nm} = \sigma_K(w^n, w^m)$$
$$x_i^{nm} = w^n(\xi_i)\bar{w}^m(\xi_i), \quad i = 1, \ldots, \mathcal{N}_e$$
$$y_{i1}^{nm} = (z_{i1}^n - w^n(\xi_i))(\bar{z}_{i1}^m - \bar{w}^m(\xi_i)), \quad i = 1, \ldots, \mathcal{N}_o$$
$$\vdots$$
$$y_{iN(i)}^{nm} = (z_{iN(i)}^n - z_{i(N(i)-1)}^n)(\bar{z}_{iN(i)}^m - \bar{z}_{i(N(i)-1)}^m), \quad i = 1, \ldots, \mathcal{N}_o.$$

From Eq. (21), we know that for each fixed n and m

$$v^{nm} = (x_0^{nm}, x_1^{nm}, \ldots, x_{\mathcal{N}_e}^{nm}, y_{11}^{nm}, \ldots, y_{\mathcal{N}_o 1}^{nm}, \ldots, y_{\mathcal{N}_o N(\mathcal{N}_o)}^{nm})^T$$

is orthogonal to p_K when both of them are viewed as $1 + \mathcal{N}_e + \sum_{i=1}^{\mathcal{N}_o} N(i)$ dimensional vectors.

We prove $p_K = kp_C$ by induction.

We have seen in Example 2 that when there is only one oscillator attached to the continuous member, $p_K = kp_C$ is necessary and sufficient in order to have normal modes. Example 3 shows also that $p_K = kp_C$ is necessary and sufficient for a system with one elastic support attached to the continuous member.

Consider then a system with $\mathcal{N}_m$ number of concentrated masses, $\mathcal{N}_e$ number of elastic supports, and $\sum_{i=1}^{\mathcal{N}_o} N(i)$ number of oscillators attached to the continuous member. Let $N = 1 + \mathcal{N}_e + \sum_{i=1}^{\mathcal{N}_o} N(i)$. Suppose that $p_K^N = kp_C^N$ is necessary for this system to have normal modes, where p_K^N, p_C^N are the vectors p_K, p_C for the N-dimensional system.

We are going to augment the above N-dimensional system into an $N + 1$-dimensional augmented system. There are three ways to augment the system by one dimension.

(1) Case 1: Attach one additional oscillator to an existing oscillator group, say, the jth oscillator group, where $1 \leq j \leq \mathcal{N}_o$.

(2) Case 2: Attach one additional oscillator as a new oscillator group with only one oscillator, say, the $(\mathcal{N}_o + 1)$th oscillator group.

(3) Case 3: Attach one additional elastic support, say, the $(\mathcal{N}_e + 1)$th elastic support to the continuous member.

We have to show that

$$p_K^{N+1} = kp_C^{N+1}$$

is necessary for this augmented system to have normal modes for each of the three cases above, where

$$p_K^{N+1} = (p_K^N, p_1), p_C^{N+1} = (p_C^N, p_2),$$

and $p_1 = k_{j(N(j)+1)}$, $p_2 = c_{j(N(j)+1)}$ for case 1; $p_1 = k_{(\mathcal{N}_o+1)1}$, $p_2 = c_{(\mathcal{N}_o+1)1}$ for case 2; and $p_1 = k_{(\mathcal{N}_e+1)}$, $p_2 = c_{(\mathcal{N}_e+1)}$ for case 3.

We prove the necessity only for case 1 as an illustration; the other two cases can be similarly argued. Since we have that normal modes for the $N + 1$-dimensional system implies normal modes for the N-dimensional system and hence $p_K^N = kp_C^N$, we need only to show $k_{j(N(j)+1)} = kc_{j(N(j)+1)}$ is also necessary. Suppose $k_{j(N(j)+1)} \neq kc_{j(N(j)+1)}$. Then it follows from Eqs. (21)–(22) that

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$(p_K^N, k_{j(N(j)+1)})\cdot(v^{nm}, y^{nm}_{j(N(j)+1)}) = 0,$$
$$(p_C^N, c_{j(N(j)+1)})\cdot(v^{nm}, y^{nm}_{j(N(j)+1)}) = 0,$$

where

$$y^{nm}_{j(N(j)+1)} = (z^n_{j(N(j)+1)} - z^n_{jN(j)})(\bar{z}^m_{j(N(j)+1)} - \bar{z}^m_{jN(j)}).$$

Since $p_K^N = kp_C^N$, multiplying both sides of the second equation by k, and subtracting from the first we obtain

$$(k_{j(N(j)+1)} - kc_{j(N(j)+1)})y^{nm}_{j(N(j)+1)} = 0.$$

If $k_{j(N(j)+1)} - kc_{j(N(j)+1)} \neq 0$, we have $y^{nm}_{j(N(j)+1)} = 0$ for any $n \neq m$, which means

$$z^n_{j(N(j)+1)} - z^n_{jN(j)} = 0$$

holds for all n, except possibly one n_1. Substituting this equation into the eigenvalue equation of $\hat{L}_K$, i.e.,

$$\hat{L}_K\Phi_n = \lambda_n\Phi_n,$$

we see that

$$z^n_{j(N(j)+1)} = z^n_{jN(j)} = \cdots = z^n_{j1} = w^n(\xi_j) = 0$$

for all $n \neq n_1$, which is impossible since the eigenfunctions $\{w^n\}$ are determined a priori, once the physical system's configuration is determined. They cannot satisfy the constraint $w^n(\xi_j) = 0$ for all $n \neq n_1$ and arbitrary ξ_j. This contradiction shows that $k_{j(N(j)+1)} = kc_{j(N(j)+1)}$ is necessary for the augmented system to have normal modes. We have thus shown that (19) is a necessary and sufficient condition for the discrete-continuous system (17) to have normal modes.

When the relation in Eq. (19) holds, we say that the discrete-continuous systems are "matched." Since condition (19) is rarely satisfied, the two combined systems are most unlikely to be matched. Therefore, generically, the combined discrete-continuous systems do not have normal modes even if the continuous system alone does have normal modes.

It should also be remarked that the necessary and sufficient condition in Theorem 3 is a generalization of the well-known commutability condition. To see this, let us consider a situation where there is only one group of oscillators in a series connection. Let n be the number of the oscillators, and let the spring coefficients and damping coefficients be, respectively, numbered as $k_1, k_2, \ldots, k_n$ and $c_1, c_2, \ldots, c_n$. Then from the definition of $\hat{L}_K$, we see that the stiffness operator in this case can be expressed as

$$K^n z = \begin{pmatrix} k_1z_1 + k_2(z_1 - z_2) \\ k_2(z_2 - z_1) + k_3(z_2 - z_3) \\ \vdots \\ k_n(z_n - z_{n-1}) \end{pmatrix} = \begin{pmatrix} k_1 + k_2 & -k_2 & & \\ -k_2 & k_2 + k_3 & -k_3 & \\ & -k_3 & \ddots & \\ & & & k_n \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

and

$$C^n z = \begin{pmatrix} c_1z_1 + c_2(z_1 - z_2) \\ c_2(z_2 - z_1) + c_3(z_2 - z_3) \\ \vdots \\ c_n(z_n - z_{n-1}) \end{pmatrix} = \begin{pmatrix} c_1 + c_2 & -c_2 & & \\ -c_2 & c_2 + c_3 & -c_3 & \\ & -c_3 & \ddots & \\ & & & c_n \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}.$$

It is well known that a necessary and sufficient condition for this system (we assume without loss of generality that diag $(m_i) = M = I$, i.e., these are the mass weighted stiffness and damping operators) to have normal modes is that K^n and C^n commute (Caughey, 1960), i.e.,

$$K^nC^n = C^nK^n.$$

On the other hand, according to Theorem 3, we know that a necessary and sufficient condition is that there exists a scalar constant k such that

$$p_K^n = kp_C^n,$$

where

$$p_K^n = (k_1, k_2, \ldots, k_n),$$
$$p_C^n = (c_1, c_2, \ldots, c_n).$$

These two conditions are actually equivalent because of the following theorem.

Theorem 4. *Let K^n, C^n, p_K^n, p_C^n be defined as above. For any n, $K^nC^n = C^nK^n$ if and only if there exists a constant k such that $p_K^n = kp_C^n$.*

Proof: The sufficiency is obvious. We prove necessity by induction.

For $n = 1$, this is obviously true. Suppose $K^nC^n = C^nK^n$ implies there exists a constant k such that $p_K^n = kp_C^n$. We are going to show that $K^{n+1}C^{n+1} = C^{n+1}K^{n+1}$ implies $p_K^{n+1} = kp_C^{n+1}$.

Since

$$K^{n+1} = \begin{pmatrix} \tilde{K}^n_{11} & \tilde{K}_{12} \\ \tilde{K}^T_{12} & k_{n+1} \end{pmatrix},$$

where

$$\tilde{K}^n_{11} = K^n + \begin{pmatrix} 0 & 0 \\ 0 & k_{n+1} \end{pmatrix}, \quad \tilde{K}_{12} = \begin{pmatrix} 0 \\ -k_{n+1} \end{pmatrix},$$

where 0 represents various zero matrices or vectors with compatible dimensions. Similarly,

$$C^{n+1} = \begin{pmatrix} \tilde{C}^n_{11} & \tilde{C}_{12} \\ \tilde{C}^T_{12} & c_{n+1} \end{pmatrix},$$

where

$$\tilde{C}^n_{11} = C^n + \begin{pmatrix} 0 & 0 \\ 0 & c_{n+1} \end{pmatrix}, \quad \tilde{C}_{12} = \begin{pmatrix} 0 \\ -c_{n+1} \end{pmatrix}.$$

Therefore,

$$K^{n+1}C^{n+1} = \begin{pmatrix} \tilde{K}^n\tilde{C}^n + \begin{pmatrix} 0 & 0 \\ 0 & k_{n+1}c_{n+1} \end{pmatrix} & \tilde{K}^n\tilde{C}_{12} + \tilde{K}_{12}c_{n+1} \\ \tilde{K}^T_{12}\tilde{C}^n + \tilde{C}^T_{12}k_{n+1} & 2k_{n+1}c_{n+1} \end{pmatrix}$$

and

$$C^{n+1}K^{n+1} = \begin{pmatrix} \tilde{C}^n\tilde{K}^n + \begin{pmatrix} 0 & 0 \\ 0 & k_{n+1}c_{n+1} \end{pmatrix} & \tilde{C}^n\tilde{K}_{12} + \tilde{C}_{12}k_{n+1} \\ \tilde{C}^T_{12}\tilde{K}^n + \tilde{K}^T_{12}c_{n+1} & 2k_{n+1}c_{n+1} \end{pmatrix}.$$

Comparing elements of these two matrices, we have

$$\tilde{K}^n\tilde{C}_{12} = \tilde{C}^n\tilde{K}_{12},$$

from which we obtain

$$k_nc_{n+1} = c_nk_{n+1}.$$

Since it is assumed that $k_n = kc_n$, we see from the above equation

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

that $k_{n+1} = kc_{n+1}$. Therefore, $p_K^{n+1} = kp_C^{n+1}$. The proof is complete.

4 Concluding Remarks

In this paper, a general model has been derived for a combined elastic continuous structure and a finite number of discrete systems, with the objective to study conditions under which the combined system has normal modes. It is shown that unless a matching condition is satisfied, the system does not have normal modes even for Kelvin-Voigt damping. The matching condition is, however, very difficult to satisfy. Therefore, the combined systems rarely have normal modes even if the simple continuous system alone has normal modes. Note that this result discusses the nature of the zeros or eigenvectors of a transfer function of the system, not the poles. Hence the results apply to underdamped, overdamped, and critically damped modes (i.e., complex, real, distinct, or repeated poles).

We further note that our detailed analysis depends heavily on the form of damping which is assumed proportional (Kelvin-Voigt damping). However, our techniques readily generalize to imply similar conclusions about the lack of normal modes for general classes of proportionally damped systems, including those with Rayleigh damping.

In our treatment we have used the standard definition of "normal modes" that requires zero off-diagonal terms in modal approximations. Issues related to approximation of non-normal mode systems (i.e., those with nonzero off-diagonal elements) by normal mode (zero off-diagonal) approximations have been discussed, for example, in Bellos (1989), Bellos and Inman (1989), Huang and Ma (1994), and Ma and Caughey (1995).

Acknowledgment

This research was supported by the U.S. Air Force Office of Scientific Research in part (H.T.B.) under grant AFOSR F49620-95-1-0236 and in part (H.T.B, D.J.I.) under grant AFOSR F49620-93-0280. The second author would like to gratefully acknowledge Prof. K. Ito for numerous discussions on a number of topics while he was a visiting professor at North Carolina State University.

References

Banks, H. T., Gates, S. S., Rosen, I. G., and Wang, Y., 1988, "The identification of a distributed parameter model for a flexible structure," *SIAM J. Control and Optimization,* Vol. 26, pp. 743–762.

Banks, H. T., Smith, R. C., and Wang, Y., 1996, *Smart Material Structures: Modeling, Estimation and Control,* Masson/Wiley; Paris/Chichester.

Bellos, J., 1989, "Theoretical and Experimental Analysis of Non-Proportional Damping," Ph.D. dissertation, Department of Mechanical and Aerospace Engineering, State University of New York at Buffalo, Buffalo, New York.

Bellos, J., and Inman, D. J., 1989, "A survey on non proportional damping," *Shock and Vibration Digest,* Vol. 27, No. 10, pp. 7–12.

Bergman, L. A., and McFarland, D. M., 1988, "On the Vibration of a Point Supported Linear Distributed System," ASME *Journal of Vibration, Acoustics, Stress, and Reliability in Design,* Vol. 110, pp. 485–492.

Bergman, L. A., and McFarland, D. M., 1988, "Vibration of a class of complex discrete-distributed systems," *Proc. of the 29th AIAA Structures, Structural Dynamics and Materials Conference,* pp. 890–900.

Bergman, L. A., and Nicholson, J. W., 1985, "Forced Vibration of Damped Combined Linear System," ASME *Journal of Vibration, Acoustics, Stress, and Reliability in Design,* Vol. 107, pp. 275–281.

Bergman, L. A., and Nicholson, J. W., 1985, "On the free and forced torsional vibration of multi-disk shaft systems," *Proc. of the 26th AIAA Structures, Structural Dynamics and Materials Conference,* pp. 515–521.

Bishop, R. E. D., and Johnson, D. C., 1960, *The Mechanics of Vibration,* Cambridge University Press, Cambridge, UK.

Caughey, T. K., 1960, "Classical Normal Modes in Damped Linear Dynamic Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 27, pp. 269–271.

Caughey, T. K., and O'Kelley, M. E. J., 1965, "Classical Normal Modes in Damped Linear Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 32, pp. 583–588.

Chen, Y., 1963, "On beams with many elastic supports," *Journal of the Franklin Institute,* Vol. 276, No. 4, pp. 273–281.

Chen, Y., 1963, "On the Vibration of Beams or Rods Carrying a Concentrated Mass," ASME JOURNAL OF APPLIED MECHANICS, Vol. 30, pp. 310–311.

Davies, H. G., and Rogers, R. J., 1979, "The vibration of structures elastically restrained at discrete points," *Journal of Sound and Vibration,* Vol. 63, No. 3, pp. 437–447.

Dowell, E. H., 1979, "On Some General Properties of Combined Dynamical Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 46, pp. 206–209.

Ewins, D. J., 1984, *Modal Testing Theory and Practice,* Research Studies Press, Letchworth, UK.

Gupta, G. S., 1970, "Natural flexural waves and the normal modes of periodically supported beams and plates," *Journal of Sound and Vibration,* Vol. 13, No. 1, pp. 89–101.

Huang, J. H., and Ma, F., 1994, "On the Approximate Solution of Nonclassically Damped Linear Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 60, pp. 695–701.

Inman, D. J., 1989, *Vibration: With Control Measurement and Stability,* Prentice-Hall, Englewood Cliffs, NJ.

Inman, D. J., and Olsen, C. L., 1988, "Dynamics of Symmetrizable Nonconservative Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 55, pp. 206–212.

Jacquot, R. G., 1970, "Vibrations of elastic surface systems carrying dynamic elements," *Journal of the Acoustical Society of America,* Vol. 47, No. 5, Part 2, pp. 1354–1358.

Kasprzyk, S., 1984, "An analysis of vibrations of a discrete-continuous system of the type (∞, 2)," *Nonlinear Vibration Problems,* Vol. 22, pp. 199–214.

Lin, Y. K., 1962, "Free vibrations of a continuous beam on elastic supports," *Int. Journal of the Mechanical Sciences,* Vol. 4, pp. 400–423.

McFarland, D. M., 1990, "Analysis of Passive and Active Discrete-Distributed Linear Dynamical Systems Using Green's Function Methods," Ph.D. Dissertation, Department of Aeronautical and Astronautical Engineering, University of Illinois, Urbana, IL.

Ma, F., and Caughey, T. K., 1995, "Analysis of Linear Nonconservative Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 62, pp. 685–691.

Mead, D. J., 1971, "Vibration Response and Wave Propagation in Periodic Structures," ASME *Journal of Engineering for Industry,* Vol. 93, pp. 783–792.

Meirovitch, L., 1967, *Analytical Methods in Vibrations,* Macmillan, New York.

Miles, J. W., 1956, "Vibrations of beams on many supports," *ASCE Journal of the Engineering Mechanics Division,* pp. 1–9.

Nicholson, J. W., and Bergman, L. A., 1986, "Free Vibration of combined dynamical systems," *ASCE Journal of Engineering Mechanics,* Vol. 112, No. 1, pp. 1–13.

Nicholson, J. W., and Bergman, L. A., 1986, "Vibration of damped plate-oscillator systems," *ASCE Journal of Engineering Mechanics,* Vol. 112, No. 1, pp. 14–30.

Nicholson, J. W., and Bergman, L. A., 1985, "Vibration of thick plates carrying concentrated masses," *Journal of Sound and Vibration,* Vol. 103, No. 3, pp. 357–369.

Pan, H. H., 1965, "Transverse Vibration of an Euler Beam Carrying a System of Heavy Bodies," ASME JOURNAL OF APPLIED MECHANICS, Vol. 32, pp. 434–437.

Wloka, J., 1992, *Partial Differential Equations,* Cambridge University Press, Cambridge, UK.

APPENDIX

We indicate how to solve the eigenvalue problem of the operator A defined in Section 2. Equivalently, we want to solve for λ_n, $w_n(x)$ and z_n from system (10) which is recapped as follows for convenience:

$$EIw_n''''(x) + k_H(w_n(\alpha) - z_n)\delta(x - \alpha) = \rho\lambda_n w_n(x), \qquad 0 < x < l,$$

$$k_H(z_n - w_n(\alpha)) = m_H\lambda_n z_n$$

$$w_n(0) = w_n'(0) = w_n''(l) = w_n'''(l) = 0. \tag{23}$$

Since a delta function is contained in this equation, we may consider the problem on two separate regions $(0, \alpha^-]$ and $[\alpha^+, l)$. For $0 < x \leq \alpha^-$, system (23) reduces to

$$EIw_n''''(x) = \rho\lambda_n w_n(x), \quad 0 < x < \alpha^-$$

$$k_H(z_n - w_n(\alpha^-)) = m_H\lambda_n z_n \tag{24}$$

and for $\alpha^+ \leq x < l$, system (23) reduces to

$$EIw_n''''(x) = \rho\lambda_n w_n(x), \quad \alpha^+ < x < l$$

$$k_H(z_n - w_n(\alpha^+)) = m_H\lambda_n z_n. \tag{25}$$

At the point $x = \alpha$, the derivatives of $w_n(x)$ up to the second order must be continuous, so that

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$w_n(\alpha^-) = w_n(\alpha^+),$$

$$w_n'(\alpha^-) = w_n'(\alpha^+),$$

$$w_n''(\alpha^-) = w_n''(\alpha^+). \quad (26)$$

Integration of the first equation in system (23) yields

$$EIw_n'''(\alpha^+) - EIw_n'''(\alpha^-) = k_H(z_n - w_n(\alpha)), \quad (27)$$

which is the jump condition for the shear force. Substituting the second equation of system (23) to eliminate z_n in Eq. (27), we obtain

$$EIw_n'''(\alpha^+) - EIw_n'''(\alpha^-) = \frac{k_H m_H \lambda_n}{k_H - m_H \lambda_n} w_n(\alpha). \quad (28)$$

In addition to these conditions, $w_n(x)$ should of course satisfy the boundary conditions at both ends $x = 0$ and $x = l$, i.e.,

$$w_n(0) = w_n'(0) = w_n''(l) = w_n'''(l) = 0. \quad (29)$$

Let $\lambda_n = EI\beta_n^4$. Then the solution of Eq. (24) is given by

$$w_n(x) = C_1 e^{\beta_n x} + C_2 e^{-\beta_n x} + C_3 e^{j\beta_n x} + C_4 e^{-j\beta_n x}, \quad 0 < x < \alpha \quad (30)$$

and the solution of Eq. (25) is given by

$$w_n(x) = C_5 e^{\beta_n x} + C_6 e^{-\beta_n x} + C_7 e^{j\beta_n x} + C_8 e^{-j\beta_n x}, \quad \alpha < x < l, \quad (31)$$

where $C_i, i = 1, \ldots, 8$ are unknown constants to be determined. Since there are exactly eight boundary conditions as shown in Eqs. (26), (28), and (29), there exists a unique solution for C_i when λ_n satisfies some specific algebraic equation from which we can evaluate the eigenvalues. Once the C_i are determined, the eigenfunctions $w_n(x), 0 < x < l$ are readily obtained from Eqs. (30) and (31).

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

T. Wierzbicki

K. A. Trauth

Department of Ocean Engineering,
Massachusetts Institute of Technology,
Cambridge, MA 02139

A. G. Atkins

Department of Engineering,
University of Reading,
Whiteknights, Reading RG6 6AY,
England

On Diverging Concertina Tearing

A simplified theory of the process of concertina tearing of thin plates with two diverging cracks is developed. The cracks initiate from a local cut and propagate away from the initiation site as the loading punch is pushed in the in-plane direction. At the same time the plate strip between the two cracks folds back and forth in much the same way as the "concertina" collapse pattern of circular tubes, studied extensively in the literature. Closed-form solutions are derived for the punch force and the length of the folding wave. Postulating a differential equation for the direction of the crack growth, a path of the tearing fracture is determined. It is shown that the crack trajectories depend on the plate thickness and the size of the initiation cut but, perhaps surprisingly, are independent of the material properties of the plate. This means that the process of concertina tearing would be similar for metal, plastic, paper, leather, etc. A series of experiments on 0.4-mm-thick mild steel plate was performed showing a very good agreement with the theoretical predictions.

1 Introduction

Concertina tearing is a common failure mode of thin plates subjected to in-plane indentation loads. When a narrow punch advances in the plate, the material in front folds back and forth ahead of the tool with simultaneous tearing at the edges. If continued, the process leaves a torn-off and folded strip of the plate (Fig. 1), much resembling the "concertina" folding pattern of a circular tube (Alexander, 1960). Depending on the size of the plate and the existence of a precut or starter crack, the process of concertina tearing proceeds in one of two modes. If the damage zone is confined to a narrow strip of a plate between two clamped edges (or between heavy longitudinal stiffening in a structure) distanced b apart, the plate may progressively fold with two parallel tearing cracks and a characteristic wavelength $2H$. The force necessary to advance the indentation process fluctuates back and forth around a mean value. Each peak and valley corresponds to the formation of one distinct folding wave. A theory of the so-called bounded concertina tearing was developed by Wierzbicki (1995). When the plate width-to-thickness ratio b/t is in the range of 30–60, an extremely regular folding pattern is observed (Fig. 2). The cracks at a clamped boundary are advancing at approximately the same speed so that the damage process is both stable and symmetric.

A different failure pattern is developed when the tearing process is initiated from a suitable precut which is positioned away from clamped boundaries. Now, two diverging tears will emanate from the initiation point and propagate ahead of the advancing tool (Fig. 3).

The objective of the present paper is to develop a simple theory of diverging concertina tearing and determine trajectories of the diverging cracks. The solution is divided into two steps. First, a formation of a single buckle or fold is analyzed in order to find a relationship between the plate width and thickness and the length of the folding wave. Then, a criterion for the direction of crack growth is introduced and the local slope of the crack trajectory is calculated. Finally, the governing differential equation is solved using the initial condition. It is shown that the path of the advancing cracks depends on the plate thickness and the size of the initial precut and is independent on the mechanical properties of the plate. The solution is compared with small-scale tearing tests reported by Trauth (1994) showing an excellent agreement.

The above failure mode is frequently encountered in everyday life and is called a "bulletin board" tear pattern. When a pinned sheet of paper is suddenly pulled down, two diverging cracks may run upwards and the turn-off tongue is folded in an alternating way with increasing wavelength.

Besides its theoretical value, the present solution has various practical applications. For example, concertina tearing has been identified in small and intermediate scale grounding tests of ship bottom structures as well as in actual grounding accidents (Kitamura et al., 1978; Maxwell, 1993; Yahiaoui et al., 1994; Paik, 1994; Astrup, 1994; Kuroiwa et al., 1992).

2 Local Analysis

Formulation of a single lobe in the process of periodic folding of the plate is analyzed first. The plate thickness is denoted by t and the material is considered rigid-perfectly plastic, characterized by the flow stress σ_0 and the specific work of fracture R. The current distance between two advancing cracks is denoted by b (Fig. 1). For the purpose of local analysis, the distance b is assumed to be known and the cracks are considered parallel. The concept of diverging cracks is introduced in Section 9 where the "global" solution is derived.

The immediate objective is to find the length of the folding wave H and the average force F_m necessary to advance the deformation process as functions of t, b, σ_0, and R. The "local" problem has been described in great detail by Wierzbicki (1995). Here only the main points of that solution are presented necessary to follow the original and main part of the paper which deals with the determination of the trajectories of diverging tears. The solution is obtained by constructing a kinematically admissible displacement field with plastic deformations localized in the generalized hinge lines separating displacing and rotating plate elements. Such a field is shown in Fig. 4 in the underformed and deformed configuration. It was found that in order to achieve full geometrical compatibility, two active folds must interact at all times. Consequently, a full deformation cycle is composed of two phases. (A reader who wants to reconstruct the present solution is strongly encouraged to construct a paper model of two interacting folds, as described below.)

In the *first phase,* a "roof"-type mechanism is formed with two sets of diagonal hinge lines DGN and AFK) intersecting at an unknown distance ξ from the edge (Fig. 5). The angle of the out-of-plane rotation is denoted as θ. Forward movement of the loading edge of the plate in contact with the punch is

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Apr. 22, 1997; final revision, Jan. 19, 1998. Associate Technical Editor: R. C. Becker.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

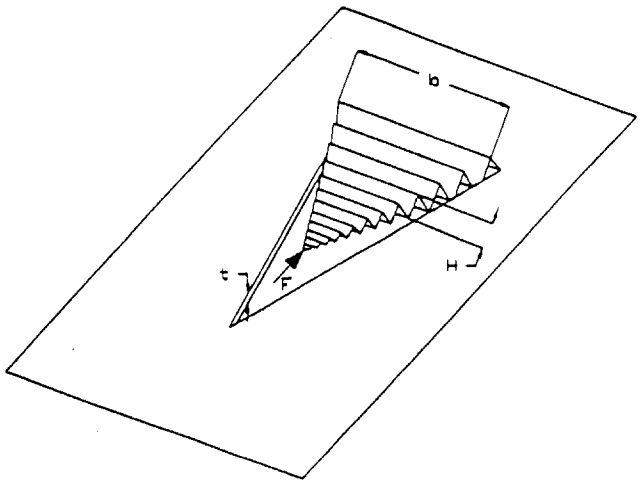


Fig. 1 A conceptual picture of diverging concertina tearing

denoted by u. Either θ or u can be considered as parameters of the process. As u or θ increases, there will be compressive deformation in the front trapezoidal elements. This is visualized in Figs. 4 and 5 as shaded areas. The amount of compression is proportional to the size of the overlapping zone Δ and can be related to the geometry of the one-degree-of-freedom roof mechanism and the process parameter θ by

$$\frac{\Delta}{H} = \frac{\xi}{H}\cos\theta\,\frac{\sqrt{\left(\frac{\xi}{H}\right)^2 - \sin^2\theta}}{\left(\frac{\xi}{H}\right)^2 + \cos^2\theta} \quad \cos\theta.$$

An analysis of the above function reveals that $\Delta(\theta)$ increases with (θ), reaches maximum at $\theta \approx 60$ deg, and decreases back to zero. This means that initial compressive deformations give way to tensile deformations. An approximate expression for the maximum overlap $\Delta_{\max}$ obtained from Eq. (1) is

$$\Delta_{\max} = \frac{H^2}{4\xi}.$$

In the *second phase* new hinge lines EF, GH are formed at the center of end triangular elements (KFA, NGD in Fig. 5). At the same time there is rebending at the diagonal hinge lines. The resulting displacement incompatibility between two adjacent elements, shown in Fig. 4, can be accommodated either by shear or tensile/compressive deformations at the back trapezoidal element. A consistent computational model with tension/compression deformation was developed earlier by one of the authors (Wierzbicki, 1995). Here it will be assumed that displacement compatibility is achieved through shear deformations (Section 6).

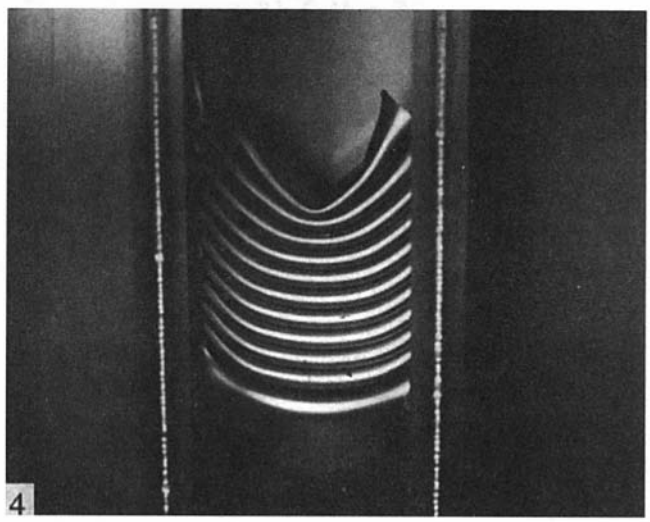


Fig. 2 A regular folding mode of a plate between two heavy longitudinal stiffners

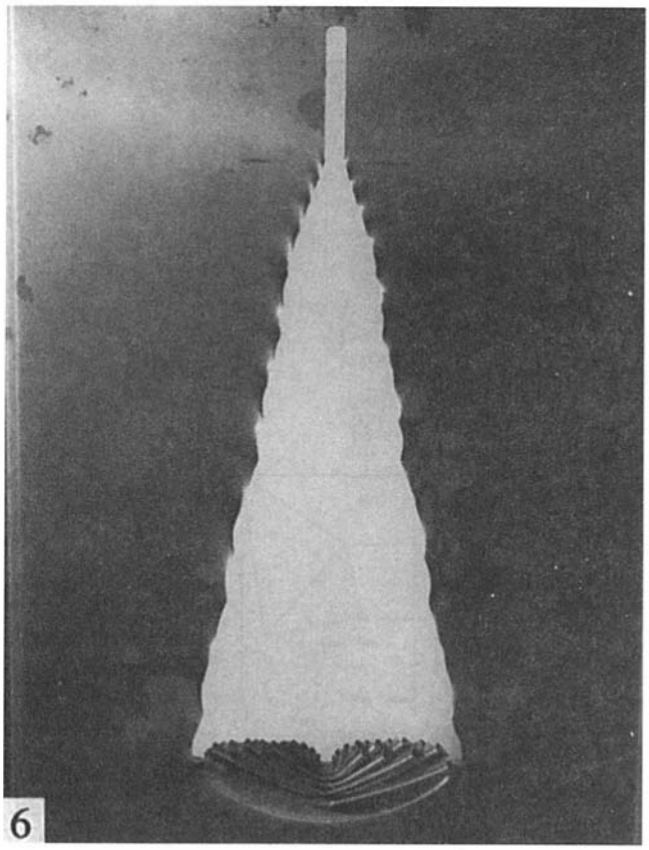


Fig. 3 A concertina tearing mode with two diverging cracks, originating from a local cut

3 Formulation of the Problem

Let us assume that the punch is pushed into the plate at a constant speed $\dot{u}$. In the previous section a kinematically admissible deformation field was constructed which involved two unknown parameters, H and ξ. At any instant the power input is assumed to be equal to the internal energy dissipation

$$F\dot{u} = \dot{E}_b + \dot{E}_m + \dot{E}_{sh} + \dot{E}_f \qquad (3)$$

where $\dot{E}_b$, $\dot{E}_m$, $\dot{E}_{sh}$, and $\dot{E}_f$ denote, respectively, rate of energy

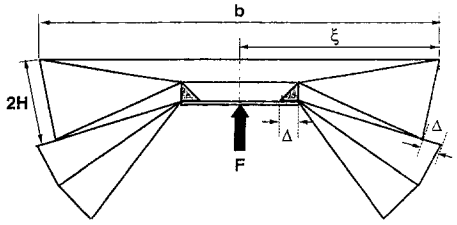


Fig. 4 Compatibility of two neighboring folding elements requires shear deformation to be developed along part of the boundary between these elements

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

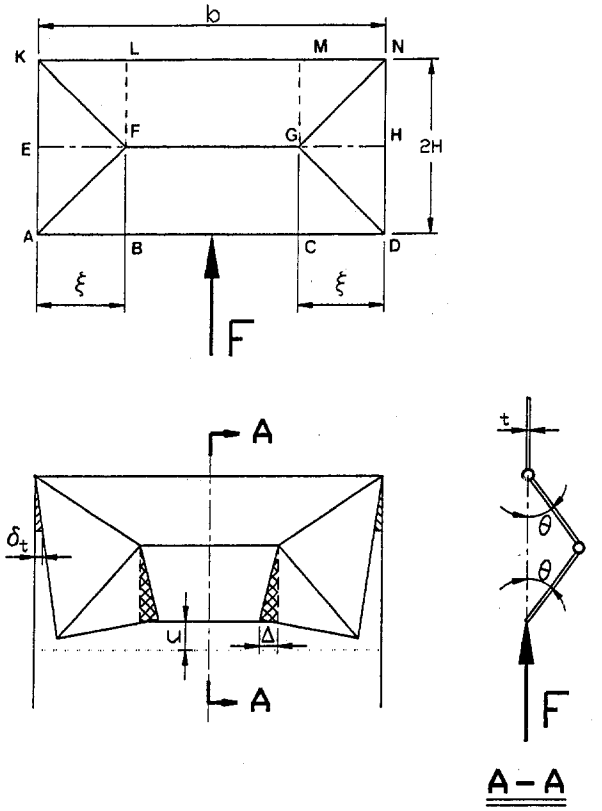


Fig. 5 Initial and current geometry of the deforming plate element in Phase I

dissipation in plate bending, plate membrane action, shear action, and plate fracture/tearing. For simplicity all terms in Eq. (3) are assumed to be uncoupled. Because the assumed kinematic model leads to one-dimensional strain fields in each deforming region, Eq. (3) can be integrated, without a loss in generality, over one deformation cycle ($0 \le u \le 2H$) to give

$$F_m 2H = E_b + E_m + E_{sh} + E_f \tag{4}$$

where E_b, E_m, E_{sh}, and E_f denote integrals of the respective rates. After integration the mean resisting force F_m becomes a function of two unknown parameters of the process, H and ξ. It is postulated that these parameters adjust so as to minimize the mean indentation force

$$\frac{\partial F_m}{\partial H} = \frac{\partial F_m}{\partial \xi} = 0. \tag{5}$$

The existence of the minimum principle in finite plasticity has been the subject of considerable debate in the literature (Calladine 1993, Abramowicz 1994). In most of the practical problems, good results are obtained when the postulate (5) were invoked. It is clear that the application of the minimum postulate would not lead to such accurate results if there were not some good physics behind it. The folding wavelength is determined early in the process. It is related to the buckling wavelength which magnitude depends in turn on the ratio of incremental bending and membrane energies. Any subsequent change in the wavelength would require activation of additional bending and membrane mechanisms with the associated expenditure of the energy increments. Therefore, from the energy point of view, it is beneficial for the structure to keep the wavelength constant during the entire folding process. Consequently, the wavelength, calculated by minimizing the average resisting force, comes up close to the wavelength obtained by optimizing the instantaneous response at the early (elastic-plastic) buckling state.

The analytical minimum of the function $F_m(H, \xi)$ does exist (Section 8) and Eq. (5) yields two nonlinear algebraic equations from which the solution can be found. The contribution of various resisting mechanisms to the right-hand side of Eq. (4) will now be calculated.

It should be pointed out that initially flat-plate elements fold into a complex pattern of double curvature shells. A question may be asked by a reader not familiar with the mechanism-type approach to the large deflection plastic problems whether the dissipated energy can be correctly calculated by assuming triangular and trapezoidal displacement fields.

It can be shown using a simple beam example that the bending energy calculated from the approximate triangular displacement field is exact while the membrane energy is underestimated by 22 percent. However, this error is later offset by considering a square yield condition with no interaction between the bending moments and membrane forces. So the actual error may in fact be very small.

4 Energy Dissipation due to Bending

The rate of bending energy is defined by

$$\dot{E}_b = \int_S M_{\alpha\beta}\dot{K}_{\alpha\beta} dS + \sum_{i=1}^{N} M_0 l_i \dot{\theta}_i \tag{6}$$

where $M_{\alpha\beta}$ are components of the bending moment tensor, and $\dot{K}_{\alpha\beta}$ the corresponding rate of curvature tensor. The first term in Eq. (6) represents the contribution of the continuous deformation field which extends over the deforming part of the body. The second term is the energy dissipation over the system of N stationary hinge lines. The length of the ith hinge line is denoted by l_i and the corresponding rate of rotation is $\dot{\theta}_i$. The plane-strain fully plastic bending moment

$$M_0 = \left(\frac{2}{\sqrt{3}}\sigma_0\right)\frac{t^2}{4}$$

is a function of the flow stress σ_0 and plate thickness t. According to the assumed kinematic model, the plate elements are rotating as rigid bodies and all plastic deformations are concentrated in a system of stationary hinge lines. Thus $\dot{K}_{\alpha\beta} = 0$ and the contribution of the continuous bending (first term in Eq. (6)) vanishes. Note that in the present one-degree-of-freedom model, hinge rotations are uniquely related to the process parameter. Therefore, Eq. (6) can be integrated in time to give

$$E_b = M_0 \sum_{i=1}^{N} l_i \theta_i \tag{7}$$

in which θ_i is the total relative rotation in the ith hinge line at the end of the full deformation cycle. Lengths and rotation angles in the first and second phases of the folding process are given in Table 1. Here, for simplicity, the length of a diagonal hinge has been approximated by its projection in the transverse direction.

Substituting the values from Table 1 into Eq. (8), the total bending energy becomes

Table 1 Contribution of bending deformation to the plate resisting force

Phase	Hinge	l_i	θ_i
Phase I	Edge	$2b$	π
	Central	b-2ξ	π
	Diagonal	4ξ	$\pi/2$
Phase II	Central	2ξ	π
	Diagonal	4ξ	$\pi/2$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\dot{E}_b = 2\pi M_0[b + 2\xi]. \tag{8}$$

5 Energy Dissipation due to Membrane Action

Using the Huber-Mises plane stress yield condition and the associated flow rule, the rate of membrane work (per unit volume) is expressed in terms of the components of the strain rate tensor by

$$\dot{E}_m = \frac{2\sigma_0 t}{\sqrt{3}} \sqrt{\dot{\epsilon}_{\zeta\zeta}^2 + \dot{\epsilon}_{\eta\eta}^2 + \dot{\epsilon}_{\zeta\zeta}\dot{\epsilon}_{\eta\eta} + \dot{\epsilon}_{\eta\eta}^2}. \tag{9}$$

According to the assumed kinematic model, the membrane of deformations are confined to the front and back trapezoidal elements (Fig. 4). The remaining trapezoidal and triangular elements rotate with respect to each other as rigid bodies. It is convenient to introduce a local coordinate system (ζ, η) which rotates with the deforming front trapezoid. The so-called velocity strain is defined by

$$\dot{\epsilon}_{\alpha\beta} = \tfrac{1}{2}(u_{\alpha,\beta} + u_{\beta,\alpha}), \quad \alpha, \beta = \zeta, \eta \tag{10}$$

where $\dot{u}_\alpha = (\dot{u}, \dot{v})$ are components of the velocity vector in the local coordinate system which rotate with rigid trapezoidal or triangular elements. Because the plate buckles and folds in front of the wedge, it is assumed that the trapezoidal element is inextensible in the η direction, i.e., $\dot{v} = 0$ and $\dot{\epsilon}_{\eta\eta} = 0$. The rate of shear work is neglected $\sigma_{\zeta\eta}\dot{\epsilon}_{\zeta\eta} = 0$ in the tension-dominant regions. The only nonvanishing component of the strain rate vector describes tension/compression in the transverse direction, $\dot{\epsilon}_{\zeta\zeta} = \partial\dot{u}_\zeta/\partial\zeta$. The present kinematic model suggests that plastic deformations are confined to a local triangular zone marked in Figs. 4 and 5 by a shaded area. The overlapping zones have been shaded only for illustrative purposes. In fact, plastic zones could extent at a finite distance on both sides of the generalized compression hinge. Under the above assumption, the expression for the rate of membrane energy per one side reduces to

$$\dot{E}_m = \frac{2}{\sqrt{3}}\sigma_0 t \int_0^H \int_{-\zeta(\eta)}^{+\zeta(\eta)} \frac{\partial\dot{u}}{\partial\zeta} d\zeta d\eta = \frac{2}{\sqrt{3}}\sigma_0 t \int_0^H \dot{u} d\eta \tag{11}$$

where $\dot{u}(\eta) = \dot{u}(+\zeta) - \dot{u}(-\zeta)$ is the relative transverse velocity between the left and right-hand sides of the trapezoidal element. In our model, this velocity is a linear function of the η-coordinate:

$$\dot{u}(\eta) = \dot{\Delta}\frac{\eta}{H} \tag{12}$$

where $\dot{\Delta}$ is the maximum (tip) velocity difference at the transverse edge of the plate. Introducing Eq. (12) into Eq. (11) and integrating the latter yields

$$\dot{E}_m = \frac{1}{\sqrt{3}}\sigma_0 H\dot{\Delta}t. \tag{13}$$

From Eq. (1) the tip velocity $\dot{\Delta}$ in a unique function of the process parameter θ. Therefore, Eq. (13) can be integrated in time

$$E_m = \frac{2H}{\sqrt{3}}\Delta_{\max}\sigma_0 t \tag{14}$$

where $\Delta_{\max}$ is given by Eq. (2) and a factor of two appears because of the strain reversal, explained in Section 2. Substituting Eq. (2) into Eq. (14), the final expression for the plastic energy dissipation per one side of the trapezoidal element is

$$E_m = \frac{2\sigma_0}{\sqrt{3}}\frac{H^3 t}{4\xi}. \tag{15}$$

This energy should be doubled for the entire front trapezoidal element,

$$E_m = M_0\frac{2H^3 t}{\xi}. \tag{16}$$

Note that the magnitude of the flow stress σ_0 may be different in the zones dominated by bending or membrane action. However, it was shown by Wierzbicki (1995) that average bending and membrane strains are approximately the same and equal to $\epsilon_{av} = 1.52(t/b)^{2/3}$. If the stress-strain diagram is described by the function $\sigma = f(\epsilon)$, the flow stress is defined $\sigma_0 = f(\epsilon_{av})$. It can be concluded that the magnitudes of flow stresses entering expressions for M_0 and N_0 are the same.

6 Energy Dissipation due to Shear

In the past, models of progressive folding of circular and square tubes were constructed in which one fold was formed at a time (Alexander, 1960; Wierzbicki and Huang, 1991). A distinctive feature of concertina tearing is that in order to achieve full geometrical compatibility, two active folds must interact at all times.

As a consequence, there is a relative shear displacement between the boundary of two subsequent folding elements (Fig. 4). This displacement is assumed to produce uniform shear strain $\epsilon_{\zeta\eta}$ in two *back* triangular elements of the area $S = (1/2)\xi H$, each. The normal strain rate $\dot{\epsilon}_{\zeta\zeta}$ is neglected in the shear dominated zone. Note that the *front* trapezoidal element is undergoing tension/compression. From Eq. (9) the rate of shear energy per one side is

$$\dot{E}_{sh} = \frac{\sigma_0 t}{\sqrt{3}}\int_S 2\dot{\epsilon}_{\zeta\eta} dS \tag{17}$$

where the shear component of the strain rate tensor is $\dot{\epsilon}_{\zeta\eta} = \partial\dot{u}/2\partial\eta$ and $dS = d\zeta d\eta$. (Because of inextensibility in the η-direction $\dot{v} \equiv 0$). After integration with respect to η Eq. (17) yields

$$\dot{E}_{sh} = \frac{\sigma_0 t}{\sqrt{3}}\int_0^\xi \dot{u} d\zeta. \tag{18}$$

It is assumed that the velocity varies linearly with the coordinates ξ along the crease of the back triangular element

$$\dot{u} = \dot{\Delta}\frac{\zeta}{\xi} \tag{19}$$

reaching maximum value $\dot{\Delta}$ at $\zeta = \xi$. Integrating with respect to ζ gives

$$\dot{E}_{sh} = \frac{\sigma_0 t}{2\sqrt{3}}\xi\dot{\Delta}. \tag{20}$$

Integrating again with respect to time and taking into account the strain reversal, the energy dissipation due to shear becomes

$$E_{sh} = \frac{\sigma_0 t}{\sqrt{3}}\xi\Delta_{\max} = \frac{\sigma_0 t H^2}{2\sqrt{3}}. \tag{21}$$

This energy should be doubled because there are two identical elements on both sides of the symmetry line.

Introducing the definition of the fully plastic bending moment, the shear contribution to the plastic energy dissipation takes a simple form

$$E_{sh} = 2M_0\frac{H^2}{t}. \tag{22}$$

7 Energy of Tearing

The concertina tearing problem involves combined plastic flow and fracture.

Following Akins (1988) and Stronge et al. (1983) (1984), the tearing strength of the plate can be characterized by a spe-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

cific work of fracture R at the crack tips. Like other mechanical properties, R may be rate, temperature, or environment-dependent, and R may change with crack growth. Note that the work of tearing, like the other terms in Eq. (3), is assumed to be uncoupled from the other energy-dissipating mechanisms occurring simultaneously. Methods of determining R experimentally under conditions of combined flow and fracture, similar to the present circumstances, are given by Cotterell and Mai (1982) whose tests are usefully gathered together in Atkins and Mai (1985 and 1988). The total work to tear off two plate segments along the length $2H$ each is

$$E_f = 4HtR. \tag{23}$$

Alternatively, the fracture energy can be calculated in terms of the crack-tip opening displacement (CTOD). Using $R \approx \sigma_0$ (CTOD), converts the CTOD relation to Eq. (23) (Wierzbicki and Thomas, 1993).

8 Folding Wavelength and Indentation Force

Substituting Eqs. (8), (15), and (17) into Eq. (4) the mean indentation force, normalized with respect to M_0, is

$$\frac{F_m}{M_0} = \pi \frac{b}{H} + 2\pi \frac{\xi}{H} + \frac{H^2}{\xi t} + \frac{H}{t} + 4\sqrt{3}\,\frac{R}{\sigma_0 t}, \tag{24}$$

where b, t, σ_0, and R are considered to be known. The unknown parameters of the deformation field H and ξ can be determined from the minimum postulate given by Eq. (5). Such postulate has been used successfully in the past in the crushing analysis of metal tubes (Alexander, 1960; Wierzbicki and Abramowicz, 1983) which is also a periodic process. Consideration of fracture does not change the minimum property because the last term in Eq. (18) depends on neither H nor ξ. Thus, Eq. (5) yields

$$\frac{1}{t} + \frac{2H}{\xi t} - \frac{2\pi\xi + \pi b}{H^2} = 0$$

$$\frac{2\pi}{H} - \frac{H^2}{\xi^2 t} = 0. \tag{25}$$

Eliminating the parameter ξ between the above equations, the width of the fold and the mean crushing force can be expressed in terms of the folding wavelength by

$$b = \frac{H^2}{\pi t} + \sqrt{\frac{2H^3}{\pi t}} \tag{26}$$

$$\frac{F_m}{M_0} = \pi \frac{b}{H} + \sqrt{\frac{8H\pi}{t}} + \frac{H}{t} + 4\sqrt{3}\,\frac{R}{\sigma_0 t}. \tag{27}$$

The above system of equations furnishes an exact parametric solution to the local problem with H as a parameter.

9 Trajectory of Diverging Cracks

At first glance, the final deformation pattern of the plate suggests that the tearing process proceeds in the out-of-plane Mode II fracture. However, a close examination of the sequential failure process in many tested pieces has revealed that a true tearing failure is a combination of all three fracture modes, with the in-plane Mode I fracture being dominant.

In order to predict the path of the diverging tears which propagate away from the initiation site, a criterion for crack trajectory must be introduced. It is postulated in this paper that the material tears in the direction normal to the direction of the principal tensile stress. Furthermore, it is assumed that the principal stress develops as a result of the diagonal tension early in the folding process when the new fold is almost flat. The diagonal tension acts between the symmetry point O at the back of the current position of the crack tip. As the fracture process

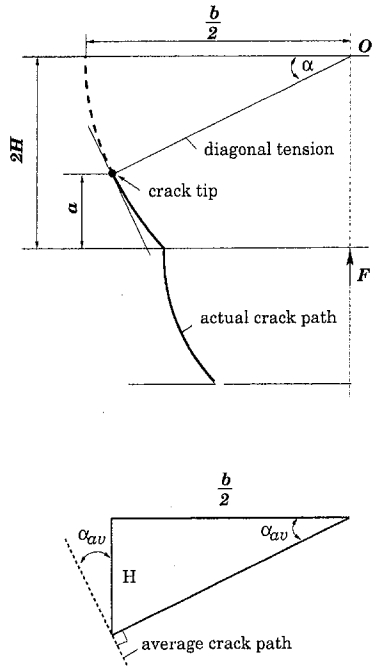


Fig. 6 Tearing procedes along direction of diagonal tension. An actual crack trajectory and an average slope.

proceeds, the instantaneous angle α of the diagonal tension is defined by

$$tg\,\alpha = \frac{2H - a}{b/2} \tag{28}$$

where a is the length of the crack, shown in Fig. 6. It is seen that the angle α is maximum at $a = 0$ and diminishes to zero for $a = 2H$. Thus, the fracture process leaves a tear in the shape of a circular arc, Fig. 6. This sequence of events is repeated in the next fold etc. The above failure pattern is fully confirmed by experiments which clearly show small waves in a form of circular arches superimposed on the diverging trajectory (Fig. 3). The average local slope of the crack path corresponds to that obtaining when the crack tip has moved a half cycle, i.e., at H. This gives a differential equation for the crack trajectory

$$\frac{d\left(\frac{b}{2}\right)}{dx} = \frac{H}{b/2} \tag{29}$$

where x is the coordinate along the symmetry line originating at the tear initiation point. By the chain rule of differentiation, Eq. (28) can be transformed to

$$\frac{db}{dx} = \frac{db}{dH}\frac{dH}{dx} = \frac{4H}{b}. \tag{30}$$

The tear width $b(H)$ and its gradient db/dH are defined by Eq. (26). Therefore, Eq. (30) furnishes a first-order nonlinear differential equation for $H(x)$

$$\frac{d\bar{H}}{d\bar{x}} = \frac{1}{0.0512\bar{H}^2 + 0.224\bar{H}^{1.5} + 0.24\bar{H}} \tag{31}$$

where all geometrical variables are normalized by the plate thickness

$$\bar{H} = \frac{H}{t}, \ \bar{b} = \frac{b}{t}, \ \bar{x} = \frac{x}{t}, \ \bar{F}_m = \frac{F_m}{M_0}. \tag{32}$$

Separating the variables, the solution, satisfying the initial condition $H = H_0$ at $x = 0$, is

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 2 Summary of test configuration

Test number	Wedge shape	b_o, Width of initial cut [mm]	Initiation site	Crosshead velocity	Observed failure mode
1	Blunt	4	Notched Free Edge	1 in./min	Diverging Concertina
2	Blunt	4	Notched Free Edge	1 in./min	Diverging Concertina
3	Blunt	4	Precut Hole	1 in./min	Diverging Concertina
4	Blunt	4	Precut Hole	1 in./min	Diverging Concertina
5	Rounded	4	Precut Hole	1 in./min	Diverging Concertina
6	Rounded	4	Precut Hole	1 in./min	Double Mode
7	Rounded	4	Precut Hole	1 in./min	Double Mode
8	45 deg Rounded	20	Precut Hole	2 in./min	Diverging and Concertina
9	45 deg Rounded	20	Precut Hole	1 in./min	Diverging Concertina
10	45 deg Rounded	20	Notched Free Edge	1 in./min	Bounded Concertina

$$\bar{x} = 0.017(\bar{H}^3 - \bar{H}_0^3) + 0.09(\bar{H}^{2.5} - \bar{H}_0^{2.5}) + 0.12(\bar{H}^2 - \bar{H}_0^2). \quad (33)$$

Together with a dimensionless version of Eq. (26)

$$\bar{b} = 0.32\bar{H}^2 + 0.8\bar{H}^{1.5} \quad (34)$$

the above system provides an exact solution of the global tearing problem in a parametric form. The length of the first fold can be calculated from a given initial width of the tear b_0 using Eq. (34). For example $\bar{b}_0 = 50$ corresponds to $\bar{H}_0 = 93$.

Both $b(x)$ and $H(x)$ are increasing functions of the coordinates x. Consequently, from Eq. (27), the mean indentation force increases as the process of concertina tearing progresses.

10 Approximate Solution

In order to relate explicitly the tear width b to the in-plane punch indentation travel x, Eq. (33) must be inverted. This can be done only in an approximate way. It was found that the following power fit gives an adequate accuracy.

$$\bar{b} = 2.38\bar{H}^{3/2} \quad \text{or} \quad \bar{H} = 0.56\bar{b}^{2/3} \quad (35)$$

Substituting the above expression into Eq. (29) and integrating under the same initial condition yields

$$\bar{b} = (3\bar{x} + \bar{b}_0^{4/3})^{3/4}. \quad (36)$$

The approximate solution slightly underpredicts the tear width as compared with the exact solution but the difference is small. For example, taking maximum values used in the experiments, $\bar{b}_0 = 50$ and $\bar{x} = 500$, the error is about four percent. Therefore, all subsequent plots will be based on the approximate solution.

Substituting Eq. (36) into Eq. (27), the mean crushing force becomes only a function of $\bar{b}$,

$$\frac{F_m}{M_0} = 9.35\bar{b}^{1/3} + 0.56\bar{b}^{2/3} + 4\sqrt{3}\,\frac{R}{\sigma_0 t}. \quad (37)$$

From Eqs. (37) and (36) the dimensionless mean indentation force can be expressed in terms of the punch travel $\bar{x}$. In the next section predictions of Eqs. (35) and (36) will be compared with test results.

11 Experimental Verification and Discussion

Ten tearing tests were conducted in a screw-driven universal testing machine with a long stroke. In all cases thin metal plates of thickness $t = 0.4$ mm were mounted in a heavy U-shaped frame. The plates were secured to the frame by a row of bolts and specially designed drawbeads to prevent in-plate slippage. In three of the plates a slit of width b_0 was cut perpendicular to the direction of travel. The remaining specimens had a precut hole of a rectangular shape of width b_0. In all cases, b_0 was made slightly greater than the width of the punch. The slit or hole began 20 mm from the top of the plate and was approximately 20 mm long. Two types of punch geometries were used. One was a square (blunt) and the other one was semicylindrical with a radius equal to half of the width of the punch.

It was found through extensive experimentation that the actual shape and size of the punch has little effect on the process of concertina tearing. This is explained by the fact that while material is piling up in front of the punch, the "active" plastic flow and fracture processes which determine the development of the concertina tear, occur at remote locations with respect to

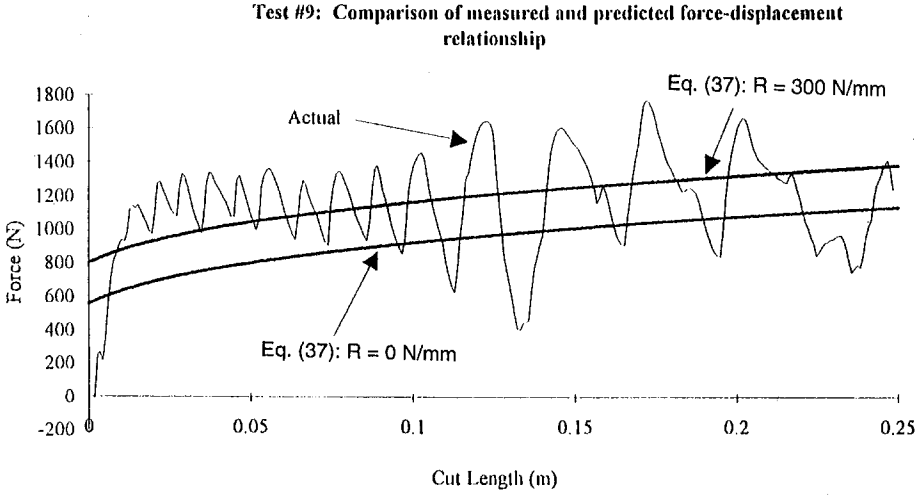


Fig. 7 Comparison of measured and predicted force-displacement relationship in Test No 8

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

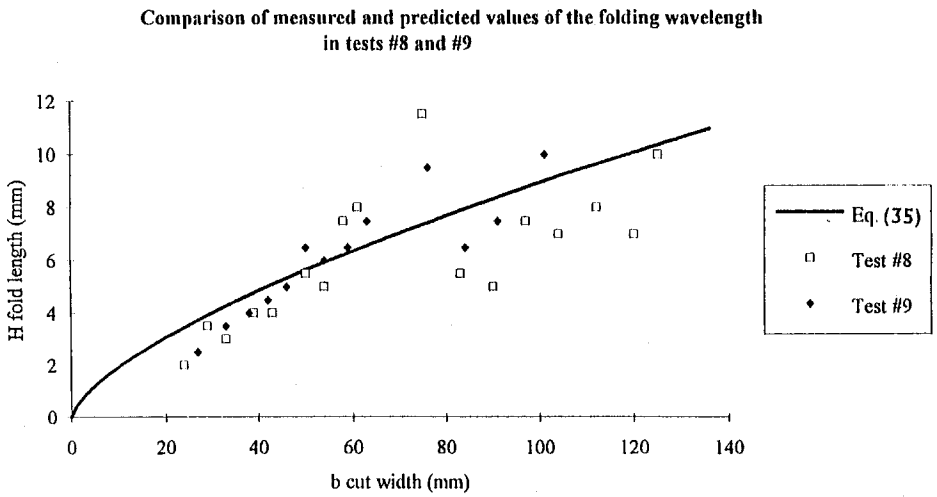


Fig. 8 Comparison of measured and predicted values of the folding wavelength in Tests No 8 and 9

the punch. For comparison, one test (Test No. 10) was performed with two longitudinal stiffeners welded to the plate at a distance $b = 53$ mm. This resulted in the bounded concertina tearing with two parallel tears, Fig. 2. In all remaining cases a nice and highly repeatable pattern of diverging tears, shown in Fig. 3, was produced. A summary of experimental configuration is given in Table 2.

Measured in all tests were the force-displacement diagram, the trajectory of the diverging cracks, and the length of the folding wave. Stress-strain diagrams were determined using ASTM tensile specimens cut at orientations of 0 deg and 90 deg to the sheet rolling direction. The average strain in the present concertina tearing process varied in the range $0.05 < \epsilon_{av} < 0.12$, depending on the (b/t) ratio. In the calculation, the flow stress corresponding to ten percent strain was used, namely $\sigma_0 = 290$ MPa. The specific work of fracture was not measured in the present series of tests. Therefore all results were plotted for a range of values bounded by $R = 0$ and $R = 300$ kJ/m^2. Owing to the high reproducibility of the experimental results, comparisons are made in this paper with one representative test run only. A detailed description of all test results can be found in the reports by Yahiaoui et al. (1994) and Trauth (1994).

In Fig. 7 is plotted the indentation force versus the punch travel. A thin jagged solid line represents recorded data from Test No. 9. The peaks and valleys correspond to the formation of subsequent folds. As the wavelength increases with x, the horizontal distance between two consecutive peaks is larger. In addition, there is a slow upward shift of the mean indentation force as the cut length increases. The predictions of the present solution, Eqs. (36) and (37) are shown by a smooth solid line. It is seen that the curve corresponding to $R = 300$ kJ/m^2 follows rather closely an average force from the test.

The width and wavelength of all folds were measured in each test and represented by points in Fig. 8. The solid line is the theoretical curve based on Eq. (35). Comparison of experimental and theoretical trajectories of diverging tears is shown in Fig. 9. A short length variation in the shape of the torn edge (circular arcs) has been filtered out from the experimental points. The predicted path of the cracks, given by Eq. (36), closely follows the experimental points. The average slope of the diverging crack calculated over the internal ($\bar{x} = 0$, $\bar{x} = 500$) with $\bar{b}_0 = 50$ equals $\alpha = 12.2$ deg. This compares rather favorably with the measured average slope which varied in the range $\alpha_{exp} = 13{:}5$ deg $-$ 14.5 deg in all nine experiments. Finally, it must be mentioned that the crack trajectory does not depend on either yield stress or specific work of fracture. A justification of the above seemingly unexpected result comes from the consideration of the fracture criterion. The crack path depends on the direction of principal stresses. These, in turn, are governed by the geometry of a single fold (width b and height H). In the present solution the wavelength depends on the ratio of bending to membrane energy which in the first

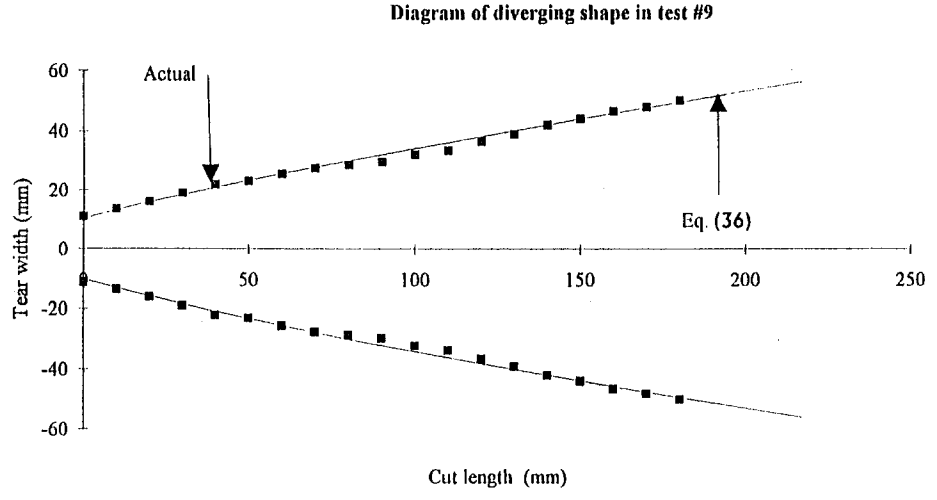


Fig. 9 Comparison of theoretical and experimental trajectories of diverging cracks in Test No 9

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

approximation is independent of the yield stress. Therefore fracture follows the process of plastic folding. This is typical of many similar problems with prescribed displacement boundary conditions.

12 Conclusions

In this study a closed-form solution was derived for the process of diverging concertina tearing initiating from a local cut. The most significant result of the present analysis is prediction of the path of two diverging cracks. Also found were the magnitude of the indentation force and the length of the folding wave. All the above quantities assumed the smallest values at the initiation site and increased with a travel of the advancing punch. A series of tests was conducted to validate the theoretical solution showing a very good reproducibility of the deformation pattern and associated forces. The theory was shown to correlate well with test results giving errors less than $5 \sim 10$ percent. The present computational model captures all of the main features of the deformation process such as in-plane tearing, out-of-plane folding, interaction between two subsequent folding waves, and strain reversal. A high degree of correlation between the theory and experiments proved validity of the underlying modeling assumptions.

An interesting feature of the present analysis is that the plate resistance is independent in the punch geometry and the friction coefficient. This is in contrast to wedge cutting of a sheet plate (so-called plate central separation mode) where all of the above parameters influence the plate cutting force. The difference is that in the process of concertina tearing the material piles up in front of the punch rather than curls and presses back on the wedge in the wake of the crack tip.

A theoretical analysis of the plate cutting problem by the wedge in the initiation and steady-state mode was developed, respectively, by Wierzbicki and Thomas (1993) and Zheng and Wierzbicki (1996).

Acknowledgment

The work presented herein was supported partially by the Joint MIT/Industry Consortium in Tanker Safety and by the Department of Ocean Engineering of the Massachusetts Institute of Technology. Both sources are gratefully acknowledged.

References

Abramowicz, W., 1983, "The effective crushing distance in axially compressed thin-walled metal columns," *Int. J. Impact Engng.*, Vol. 1, pp. 309–317.

Abramowicz, W., 1994, private communication.

Alexander, J. M., 1960, "An approximate analysis of the collapse of thin cylindrical shell under axial loading," *Q. J. Mech. Appl. Math.*, Vol. 13, No. 1, pp. 10–15.

Astrup, O., 1994, "Cutting of thick plates by a wedge—An experiment study," Report No. 27, Joint MIT/Industry Program on Tanker Safety.

Atkins, A. G., 1988, "Scaling in combined plastic flow and fracture," *Int. J. Mech. Sci.*, Vol. 30, p. 173.

Atkins, A. G., and Mai, Y.-W., 1985 and 1988, *Elastic and Plastic Fracture*, Ellis Horwood, Chichester, UK.

Calladine, C. R., 1993, "Some problems in propagating plasticity," *Proc. IMPLAST '93 Symposium on Plasticity and Impact Mechanics*, I.T.T. Delhi, Dec. 11–14.

Cotterell, B., and Mai, Y.–W., 1982, *Advances in Fracture Research*, (ICF-5), D. Francois ed., Vol. 4, Pergamon, London, UK, pp. 1683–1695.

Jones, N., and Jouri, W. S., 1987, "A study of plate tearing for ship collision and ground damage," *J. Ship Res.*, Vol. 31, p. 253.

Kitamura, K., Okumoto, Y., and Shibue, T., 1978, "On the model tests of double bottom strength for stranding (in Japanese)," *J. Soc. Naval Architects Japan*, Vol. 143, pp. 346–356.

Kuroiwa, T., Kawamoto, Y., and Yuhara, T., 1992, "Study on damage of ship bottom structures due to grounding," *Proc. Conf. on Prediction Methodology of Tanker Structural Failures*, Tokyo, Japan.

Lu, G., and Calladine, C. R., 1960, "On the cutting of a plate by a wedge," *Int. J. Mech. Sci.*, Vol. 32, pp. 295–313.

Maxwell, L. M., 1993, "Effect of rock geometry in the failure mode of plates and the forces in grounding experiments," M.S. thesis, Department of Ocean Engineering, MIT; also Report No. 15 Joint MIT/Industry Program on Tanker Safety.

Paik, J. K., 1994, "Cutting of longitudinally stiffened steel plate by a rigid wedge," *J. Ship Res.*, submitted for publication.

Stronge, W. J., Yu, T. X., and Johnson, W., 1983, "Long stroke energy dissipation in splitting tubes," *Int. J. Mech. Sci.*, Vol. 25, pp. 637–647.

Strong, W. J., Yu, T. X., and Johnson, W., 1984, "Energy dissipation by splitting and curling tube," *Structural Impact and Crashworthiness*, J. Morton, ed., Vol. 2, Applied Science, UK, pp. 576–587.

Trauth, K., 1994, "Validation of concertina tearing solution," Report No. 34, Joint MIT/Industry Program on Tanker Safety.

Wierzbicki, T., and Abramowicz, W., 1983, "On the Crushing Mechanics of Thin-Walled Structures," ASME JOURNAL OF APPLIED MECHANICS, Vol. 5, pp. 727–734.

Wierzbicki, T., and Thomas, P., 1993, "Closed-form solution for wedge cutting force through thin metal sheets," *Int. J. Mech. Sc.*, Vol. 35, No. 4, pp. 209–229.

Wierzbicki, T., 1995, "Concertina tearing of metal plates," *Int. J. Solids Structures*, Vol. 32, pp. 2923–2943.

Yahiaoui, M., Bracco, M., Little, P., and Trauth, K., 1994, "Experimental studies on scale models for grounding," Report No. 19, Joint MIT/Industry Program on Tanker Safety.

Zheng, Z.-M., and Wierzbicki, T., 1996, "A theoretical study on the steady-state wedge cutting through thin metal plates," *Int. J. Fracture*, Vol. 78, No. 1.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Tangential Loading of General Three-Dimensional Contacts

M. Ciavarella[1]
Department of Engineering Science,
Oxford University,
Parks Road,
Oxford OX1 3PJ, England

A general three-dimensional contact, between elastically similar half-spaces, is considered. With a fixed normal load, we consider a pure relative tangential translation between the two bodies. We show that, for the case of negligible Poisson's ratio, an exact solution is given by a single component of shearing traction, in the direction of loading. It is well known that, for full sliding conditions, the tangential force must be applied through the center of the pressure distribution. Instead, for a full stick case the tangential force must be applied through the center of the pressure distribution under a rigid flat indenter whose planform is the contact area of the problem under consideration. Finally, for finite friction a partial slip regime has to be introduced. It is shown that this problem corresponds to a difference between the actual normal contact problem, and a corrective problem corresponding to a lower load, but with same rotation of the actual normal indentation. Therefore for a pure translation to occur in the partial slip regime, the point of application of the tangential load must follow the center of the "difference" pressure. The latter also provides a complete solution of the partial slip problem. In particular, the general solution in quadrature is given for the axisymmetric case, where it is also possible to take into account of the effect of Poisson's ratio, as shown in the Appendix.

1 Introduction

Transmission of loads and guiding of components is very often accomplished by mechanical contact, in a variety of possible geometries. It becomes important, then, to quantify traction distributions to improve our understanding of the "strength" of the contact as defined usually by the maximum value of the von Mises parameter. Alternatively, depending on the application, the quantification of fretting damage and surface frictional energy dissipation, as well as crack initiation and crack propagation. Also, in many applications of robotics and, in general, in production and manufacturing, such as positioning of objects and workpieces, it is of paramount importance to precisely quantify the elastic deformations, and the relative displacements induced into bodies. Indeed, it may be desired to produce *pure relative translation* of the two bodies. It is known that, in a general contact in *gross or full sliding,* the condition to obtain pure relative translation is the application of a tangential force through the *centroid* of the pressure, i.e., the point of application of the normal force[2] (MacMillan, 1936). However, the condition for full sliding cannot be reached, other than through an increase of the tangential force from zero. Now, when a tangential load is applied to a general contact between elastic bodies, a complex frictional shearing traction distribution arises to equilibrate the applied load. The contact area is divided into regions of micro-slip and regions of adhesion, according to Coulomb's law. Therefore it is possible that, on applying the tangential force through the centroid of pressure, there is an undesired relative rotation until the full sliding conditions are reached. To investigate this possibility, we need to consider the contact problem in the partial slip regime, evolving from the fully adhesive condition in the presence of normal load only, towards the full sliding conditions, when the tangential load is fully applied.

There is a number of results available in the literature for (frictionless) normal contact problems, at least in the case where the contacting bodies can be approximated as half-spaces. For a circular contact area, that practically requires axisymmetry of the geometry and boundary conditions, a general solution in quadrature is known (see, for example, Scthaerman, 1949). In the case of general shape of the (single) contact area, approximate solutions are known (Fabrikant, 1986), and the normal compliance and relation load-contact area can be also predicted accurately by ad hoc methods (Barber and Billings, 1990). There remains the far more complicated case of multiple contact area, which is of particular importance when rough surfaces are in contact. In this case, recourse to numerical methods to solve the governing equations is in practice not avoidable. Indeed, several solution techniques have been developed (see, for example, Kalker, 1990), although it is not always easy to achieve the desired accuracy, especially regarding localized regions, for example when singularities are expected at the contact area boundary, and in that case they may also change intensity with position (Fabrikant et al., 1988), in the case of sharp corners.

Moving to the case of interest for the present investigation, where a tangential action is applied with a finite friction coefficient, the number of known solutions is largely inferior, even in the simplest case of elastically similar materials. The problem corresponds to the finding of a shearing traction distribution such that there is rigid-body displacement in the entire stick zone, fulfilling Coulomb's law in both stick and slip areas. Cattaneo (1938) solved the three-dimensional contact problem for second-order surfaces (Hertzian contact). The solution is obtained by making an "educated" guess of a distribution of tractions that fulfills the requirements for displacement fields. It is little known that Cattaneo also solved the case of an axisymmetric contact with distance function of fourth degree, in 1947 (Cattaneo, 1947a, 1947b). Later on, Mindlin (1949) independently reobtained the solution of the Hertzian contact, and in Mindlin et al. (1952) and Mindlin and Deresiewicz (1953) also gave solutions to the unloading, cyclic, and oblique loading problems. Deresiewicz (1957) gave the treatment for the oscil-

[1] Presently at the Mechanical Engineering Department, University of Southampton, Highfield SO17 1BJ, UK.

[2] This is actually true, a rigori, only in the absence of Poisson's effect, as we will see in the discussion for the general case $\gamma \neq 0$ below.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, June 23, 1997; final revision, June 18, 1998. Associate Technical Editor: J. T. Jenkins.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

lating tangential forces in elliptical contact area case, while many experimental results by Mindlin himself (Mindlin et al., 1952), and more extensively, by Johnson (1955), were in very good agreement with the theory in two respects: the tangential compliance, and the appearance of annular regions of slip, where wear resulted. Other results are known for twisting problems (Lubkin, 1951; Cattaneo, 1955; Pacelli, 1956; Hetenyi et al., 1958; Hills et al., 1986) in Hertzian contacts; it is well known that, even in the case of pure twist of Hertzian contacts, the equations become quite cumbersome, in particular in the case of elliptical contact (Pacelli, 1956), and therefore we do not address this configuration. In the case of tangential loading with complex stick-slip patterns, however, the computational effort of a fully numerical solution is, even for ad hoc procedure (Kalker, 1990), very intensive, and it is still difficult to go under certain levels of numerical error.

For the shifting partial slip problem, apart from the geometry already solved by Cattaneo, to the best of the author's knowledge, only the case of a cone indenter has been solved to date (Truman et al., 1995). This is probably because the Cattaneo-Mindlin procedure involves the explicit calculation of the displacements, whereas a general method would ideally avoid completely this step, formulating the problem directly in terms of integral equations. In a recent paper (Ciavarella, 1998a, b), the author has treated the case of partial slip in the general *plane* contact between elastic isotropic bodies, showing in particular a connection in any plane partial slip contact problem with two corresponding frictionless plane contact problems, for which solutions in quadrature are known for single, multiple, and periodical contact; this result has permitted new exact solutions to many plane partial slip problems, for which the Cattaneo-Mindlin procedure, involving the explicit calculation of displacements, would be prohibitively complicated.

In this paper it is shown that in the three-dimensional case, for negligible elastically dissimilarity and combined Poisson's ratio, similar results apply, as a single distribution of shearing tractions (i.e., shear only in the direction of the applied force) solves the problem. This permits one to obtain several properties of the tangential loading regime, and provides a means to do actual calculations either analytically or numerically in a much simpler way. It is known that, in the presence of Poisson's effect, even the Cattaneo-Mindlin solution for Hertzian contacts is *approximate,* in that it neglects the presence of a small shearing distribution in the transverse direction (Munisamy et al., 1992). In view of this, attention will be concentrated on contact problems for the case of no Poisson's effect ($\gamma = 0$ as below).

2 Formulation

Let us start from the governing equations for the case of normal indentation. In general, there is coupling between tangential and normal displacements, and therefore between pressure and shearing tractions. In that case, exact analytical solutions are not known even in the simplest geometrical and loading configurations. However, the effect of coupling is generally small (Goodman, 1962), and indeed there is a vast category of situations where the contacting bodies material are of same material, or the materials are elastically similar: specifically, the case where Dundurs' constant, β, is zero

$$\beta = \frac{\mu_2(\kappa_1 - 1) - \mu_1(\kappa_2 - 1)}{\mu_2(\kappa_1 + 1) + \mu_1(\kappa_2 + 1)} = 0, \tag{1}$$

where κ is the Kolosov's constant, given by $\kappa = (3 - 4\nu)$ as in the case of plane-strain conditions, (ν_i, μ_i are the Poisson's ratio and shear modulus of the material of body i). In this case, the equations relating normal and tangential traction become uncoupled, and the contact pressure can be calculated a priori. Let us write relative surface displacements as $\mathbf{u} = \mathbf{u}_1 - \mathbf{u}_2$. Then, the integral equation relating normal displacements to the pressure is (Johnson, 1985)

$$\frac{2\pi}{A} u_z(x, y) = \iint_S \frac{p(\xi, \eta)}{\rho} d\xi d\eta, \quad (x, y) \in S \tag{2}$$

where S is the contact area, as yet undetermined, ρ is the distance between field and integration point

$$\rho = \sqrt{(\xi - x)^2 + (\eta - y)^2}, \tag{3}$$

and A is the "composite compliance" of the bodies[3]

$$A = \left(\frac{1 - \nu_1}{\mu_1} + \frac{1 - \nu_2}{\mu_2} \right). \tag{4}$$

This integral equation is sufficient to solve the normal contact problem, imposing the appropriate boundary conditions. In particular, the equality (inequality) relating to contact over the area S (noninterpenetration condition exterior to S) is

$$u_z(x, y) = \delta_z - [\theta_x y + \theta_y x] - [f_1(x, y) - f_2(x, y)], \quad (x, y) \in S \tag{5}$$

$$u_z(x, y) > \delta_z - [\theta_x y + \theta_y x] - [f_1(x, y) - f_2(x, y)], \quad (x, y) \notin S \tag{6}$$

where $f_1(x, y), f_2(x, y)$ are the functions describing the profiles of the contacting bodies in the undeformed configuration, and $(\delta_z, \theta_x, \theta_y)$ is the given rigid-body motion necessary to bring the two body into contact. Given $(\delta_z, \theta_x, \theta_y)$, the distribution of pressure $p(x, y)$ is determined, from which a resultant load, and moment $\mathbf{M}$ (of components M_x and M_y) with respect to the coordinate system (ξ, η) can be calculated as

$$P = \iint_S p(\xi, \eta) d\xi d\eta, \quad M_x = \iint_S p(\xi, \eta) \xi d\xi d\eta,$$

$$M_y = \iint_S p(\xi, \eta) \eta d\xi d\eta. \tag{7}$$

Of course, there is a point where the resultant P can be applied without any moment, and it is called the *center of gravity* of the normal pressure distribution (simply, the *centroid* of the pressure), and indicated by N; any coordinate system (ξ_0, η_0) through such a point N will satisfy

$$\iint_S p(\xi_0, \eta_0) \xi_0 d\xi_0 d\eta_0 = 0,$$

$$\iint_S p(\xi_0, \eta_0) \eta_0 d\xi_0 d\eta_0 = 0. \tag{8}$$

Although there is no exact method to solve the contact problem in general, and in particular for other than circular or elliptical contact areas S[4], remarkable is Barber's method (see Barber et al., 1990) which is based on a variational formulation, using the theorem that the exact shape of the contact area is the one that maximizes the load, for a given indentation; a numerical method is then needed anyway, to find this maximum.

On applying a monotonically increasing tangential load in direction x, there will in general be stick and slip zones. The shearing tractions are strictly related to the limiting value of Coulomb friction (see below). The other two integral equations defining the problem both relate to displacement of particles parallel with the surface, one in the direction of the tangential force, and the other perpendicular to it. They are

[3] Notice that A corresponds to $2/E^*$ in Johnson's notation (Johnson, 1985).

[4] In fact, for circular area, the problem being axisymmetric, a solution in quadrature is possible (Schtaerman, 1949). For elliptical contact area, it is difficult to think of any non-Hertzian problem that produces exactly this shape.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\frac{2\pi}{A}u_x(x,y)=\iint_S\frac{q_x(\xi,\eta)}{\rho}d\xi d\eta+\gamma\iint_S\left[\frac{(\xi-x)^2}{\rho^3}q_x(\xi,\eta)-\frac{(\xi-x)(\eta-y)}{\rho^3}q_y(\xi,\eta)\right]d\xi d\eta \quad (9)$$

$$\frac{2\pi}{A}u_y(x,y)=\iint_S\frac{q_y(\xi,\eta)}{\rho}d\xi d\eta+\gamma\iint_S\left[\frac{(\eta-y)^2}{\rho^3}q_y(\xi,\eta)+\frac{(\xi-x)(\eta-y)}{\rho^3}q_x(\xi,\eta)\right]d\xi d\eta \quad (10)$$

for $(x, y) \in S$, where

$$\gamma=\left(\frac{\nu_1}{\mu_1}+\frac{\nu_2}{\mu_2}\right)\Big/A. \quad (11)$$

It is evident that, even though we are considering the $\beta = 0$ case, there is a second coupling, due to γ, and it is between the distributions of shearing tractions themselves. Indeed, this system of coupled equations gives a solution $[q_x(x, y), q_y(x, y)]$ for any couple of assigned functions $[u_x(x, y), u_y(x, y)]$ in the domain S. Therefore, in the general case, even if we impose, say, $u_y(x, y) = 0$, due to the coupling, it is not necessary that $q_y(x, y) = 0$. Also, again due to the coupling, there is no analogy between the integral equation for the normal displacements (2) and the tangential correspondents (9), (10). Therefore, in order to avoid coupling, and obtain an equation similar to (2), let us consider the case that the composite Poisson's ratio of the bodies is zero[5] $\gamma = 0$: equations in tangential direction simplify to

$$\frac{2\pi}{A}u_x(x,y)=\iint_S\frac{q_x(\xi,\eta)}{\rho}d\xi d\eta \quad (12)$$

$$\frac{2\pi}{A}u_y(x,y)=\iint_S\frac{q_y(\xi,\eta)}{\rho}d\xi d\eta \quad (13)$$

which are two independent problems, each one analogous to a normal contact problem in (2). The total shearing forces are

$$Q_x=\iint_S q_x(\xi,\eta)d\xi d\eta,\quad Q_y=\iint_S q_y(\xi,\eta)d\xi d\eta. \quad (14)$$

The condition we look for is the *pure relative translation.*[6] Therefore, we seek here for conditions under which a simple solution (i.e., with only one component of shearing traction distribution) for the pure relative translation problem is possible. Let us denote with x the direction where the resultant force is applied, i.e., $Q_y = 0$: it is clear that a priori we do not know whether this is also the direction of the relative translation. However, as the shearing tractions in the two bodies are equal and opposite for Newton's third law, the relative tangential displacement of surface particles must be constant within the stick zone, and equal to the imposed rigid translation

$$u_x(x,y)=\delta_x;\quad u_y(x,y)=\delta_y,\quad (x,y)\in S_{\text{stick}} \quad (15)$$

where δ_y is expected to be zero under conditions of symmetry. Moreover, Coulomb's law requires that the shearing traction must be less than the limiting value in the stick zone, i.e.,

$$|\mathbf{q}(x,y)|<fp(x,y),\quad (x,y)\in S_{\text{stick}}. \quad (16)$$

Within the slip zones the shearing traction is limited by friction again, i.e.,

$$|\mathbf{q}(x,y)|=fp(x,y),\quad (x,y)\in S_{\text{slip}} \quad (17)$$

where the vectorial notation is necessary because, in general, $\mathbf{q}(x, y)$ has two components. The shear traction must always oppose the direction of relative change in the direction of slip, i.e. under monotonic tangential loading,

$$\frac{\mathbf{q}(x,y)}{|\mathbf{q}(x,y)|}+\frac{\partial}{\partial t}\frac{\mathbf{u}_t(x,y)}{|\mathbf{u}_t(x,y)|}=0,\quad (x,y)\in S_{\text{slip}}, \quad (18)$$

where $\mathbf{u}_t = \mathbf{u}_x + \mathbf{u}_y$ is the relative tangential displacement, and the variable time t is introduced, but it is clear that we are considering a quasi-static formulation. The above equations provide the framework for solving the problem.

3 Solution

As in the normal loading phase there was no tendency for surface particles to slip, the initial stick zone envelopes the entire contact. A monotonically increasing tangential remote relative displacement will therefore give rise to a *receding* problem, according to Dundurs' classification (Dundurs, 1975), and therefore there is no need to follow the entire loading path, and we can solve directly for any particular value of δ_x. We only need to distinguish between the limit full stick conditions, and finite friction conditions.

3.1 Full Stick Conditions. We shall assume that the shearing distribution will have act only in the direction of the tangential displacement. Writing the boundary condition (15) in the integral equation (12), we have

$$\delta_x=\frac{A}{2\pi}\iint_S\frac{q_x(\xi,\eta)}{\rho}d\xi d\eta\quad (x,y)\in S. \quad (19)$$

Equation (13) is identically satisfied, as $u_y(x, y)$ depends only on $q_y(x, y)$, which is assumed to be zero. Integral Eq. (19) then, is formally equivalent to the problem for a *rigid flat* punch of planform S, indenting the half-space with a rigid vertical displacement. Therefore, q_x is proportional to the distribution of pressure, which is singular along the boundary of S, of the cited normal indentation problem; this implies also that a pure translation solution is possible only if the line of application of the resultant Q_x is through the center of the *rigid flat* punch of planform S pressure distribution, that we call the tangential center T (in the full stick conditions). Moreover, this solution will satisfy the problem *exactly.* We can therefore invoke *uniqueness* of the elastic solution to prove that there will be no other solution of the tangential shift problem, perhaps with a self-equilibrated distribution q_y of shearing tractions. It is interesting to study in more detail the tangential center T; we have proved that, if the direction of the tangential force Q_x does not pass through this point, there will be relative twist of the contacting bodies, because, for uniqueness, the conditions $u_x(x, y) = \delta_x$; $u_y(x, y) = 0$ cannot be satisfied in the entire contact area by any other distribution of shearing tractions than the given q_x distribution. It is clear that for a fairly general profile, for which the shape of the contact area S changes continuously increasing the load, and consequently so does the tangential center T, for each value of the normal load the set of possible direction of tangential load for which there is pure translation is the set of lines intersecting in such point T. Only on condition of symme-

[5] For materials having positive Poisson's ratio, this is only possible if both $\nu_1 = \nu_2 = 0$.

[6] Things become much more complicated when a relative tangential rotation is present. In the simplest case of "pure spin," only for the Hertzian circular and elliptical contacts a solution is known, although the results are not in closed form (Lubkin, 1951; Cattaneo, 1955; Pacelli, 1956; Hetenyi et al., 1958; Hills et al., 1986).

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

try the direction of application of Q_x for pure translation will not change. To note that the result applies equally to connected, as well as multiply connected areas S.

Again, no method exists to provide a solution in the general case of S. This aside, the analogy is sufficient to validate the assumption that full stick implies a singular distribution of shearing traction, that cannot be sustained by a *finite* friction coefficient. Therefore, the mathematical problem of full stick is practically non very meaningful in this context, unless we consider the case where the bodies are welded together, i.e., we have an external crack problem with the singularity giving a stress intensity factor.

3.2 Partial Slip Conditions. Let us then move to a partial slip regime, that arises if friction is finite. We assume a shearing traction distribution given by the sum of two components, a full sliding term, and a correction, as

$$q_x(x, y) = fp(x, y) - q_x^*(x, y), \quad (x, y) \in S \tag{20}$$

where $q_x^*(x, y) = 0$ in the region of slip, i.e., for $(x, y) \in S_{\text{slip}}$, whereas it is at yet undetermined in the stick region. Writing again the integral equation for relative displacement in the tangential direction (12),

$$\delta_x = \frac{A}{2\pi}\iint_S \frac{fp(\xi, \eta)}{\rho}\, d\xi d\eta - \frac{A}{2\pi}\iint_{S_{\text{stick}}} \frac{q_x^*(x, y)}{\rho}\, d\xi d\eta, \quad (x,y) \in S \tag{21}$$

whereas Eq. (13) is, again, identically satisfied. Now, as the normal loading equation (2) continues to hold, it follows that $q_x^*(x, y)$ is the solution of the integral equation

$$\frac{A}{2\pi}\iint_{S_{\text{stick}}} \frac{q_x^*(x, y)/f}{\rho}\, d\xi d\eta = u_z(x, y) - \delta_x/f, \quad (x, y) \in S_{\text{stick}} \tag{22}$$

which can be recognized as being of the same kind as the original equation involving the contact pressure (2), for a lower value of the vertical rigid-body displacement, giving the resulting "contact" area (i.e., the stick area) to be a subregion of the actual contact area. In particular, it is important to note that the rotation is the one fixed by the actual normal contact problem, i.e., by Eq. (5).

The analogy is so far only partly proved, in that we need to check that the inequalities are automatically satisfied by this analogy. Inequality (6) corresponds to the condition that the relative tangential displacement, for the chosen sign of the tangential load, has to be positive: this results from a theorem in normal contact which states that there is no point at which an increase of normal force causes a decrease in the local contact pressure (Barber, 1992, Section 25.3.2). Equation (13) proves that the direction of slip in the slip zones is satisfied, as $u_y = 0$, whilst the definition of $q_x^* = 0$ satisfies implicitly the limiting conditions of Coulomb's law in the slip zones. As the solution found is the exact solution to the problem, and the solution is unique, there cannot be a solution with $q_y(x, y) \neq 0$ in any point of the contact area. Thus, any partial slip solution for a contact problem, for elastically similar materials, and of any shape (simply or multi-connected) of contact area, can be completed immediately, providing the corresponding normal load is itself known.

Let us now consider the implications on the direction of the force Q_x. First, consider that the full sliding component, with tangential force fP, clearly passes through the center of the normal pressure distribution N, in the direction x. Then, we define a center of corrective pressures by T^*. The conditions for determining the centre of the shearing tractions in the actual partial slip tangential problem are given by elementary vector analysis. In particular, if the distance in the y-direction between N and T^* is d^*, then it is immediately possible to find the distance d in the y-direction between N and T, which is

$$d = \left(1 - \left|\frac{Q_x}{fP}\right|\right) d^* \leq d^* \tag{23}$$

and in particular at the first stages of application of Q_x, i.e., when $Q_x/fP \simeq 0$, it is $d \simeq d^*$ but d^* is expected to be small, as the corrective problem is only slightly different from the actual normal contact. At the other limit, when $Q_x/fP \simeq 1$, close to the full sliding, even if d^* is large, $d \simeq 0$, as expected because the corrective problem becomes of vanishing small load. Therefore, if we look for pure translation *without* moving the direction of application of the tangential load, we can distinguish between three situations, depending on what happens to the contact on unloading the normal load with the fixed rotation of Eqs. (5), (6):

1 T^* does not move—i.e., it rests coincident with N. In this case, all directions passing through this point give possible x-direction of application of tangential load that give pure translation solutions.

2 T^* moves along a line. In this case, only this line gives a possible direction of tangential loading that gives a pure translation.

3 T^* moves along a curve. There is no translation solution, for any direction of application of Q_x.

It is clear that, in general, it is possible to obtain pure translation in any direction, just following the path of the point T, i.e., applying the tangential load at distance d varying according to Eq. (22).

Moreover, several elementary properties can be deduced:

- Flat regions are either entirely in full stick or are in full slip conditions, as in unloading they either stay in contact, or loose contact simultaneously. In the limiting case of an entirely flat contact, the behavior is full stick/full slip.
- The points that loose contact first in the normal unloading, with the fixed rotation of the actual normal contact problem, are the first to slip when the tangential force is imposed.
- The tangential approach can be calculated from the superposition of the two normal approaches of the normal contact problems, as

$$\frac{\alpha_t}{f} = \alpha_n - \alpha_n^* \tag{24}$$

- The shape of the stick zone is the one that maximizes the corrective load for a given corrective normal approach, therefore the shape of the slip area is the one that minimize the full sliding component.
- If the indenter profile is symmetrical and self-similar, the stick zone is similar to the initial shape, and the corrective solution is always of the same functional form of the normal pressure in the contact area.

4 Examples

This simple result permits new solutions to be found immediately from know solutions of normal contact problems. Also, numerical methods for simple frictionless normal contact problems, permit the solution of partial slip problems, with evident advantages in terms of computational time. Some solutions are worked out in the following paragraphs.

4.1 Axisymmetric Contact. In this case the solution for the normal contact problem of any continuous gap function $z(r) = f_1(x, y) - f_2(x, y)$ is known in quadrature (Scthaerman, 1949). The simplest cases have been already worked out, namely second and fourth-order contacts (Cattaneo, 1938,

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

1947a, 1947b), and the cone indenter (Truman, 1995). In the general case, supposing $p(a)$ is bounded (the so-called *incomplete contact conditions*),

$$p(r) = -\frac{2}{\pi A}\int_r^a \frac{F'(s)ds}{\sqrt{s^2 - r^2}}, \quad 0 \le r \le a, \tag{25}$$

where

$$F(r) = \alpha_n - r\int_0^r \frac{z'(t)dt}{\sqrt{r^2 - t^2}}, \quad 0 \le r \le a \tag{26}$$

and the approach of remote points in normal direction α_n is given by

$$\alpha_n = a\int_0^a \frac{z'(t)dt}{\sqrt{a^2 - t^2}}. \tag{27}$$

The equilibrium condition between the applied load and the pressure distribution can be written as

$$P = \int_0^a 2\pi r p(r)dr = \frac{4}{A}\int_0^a \frac{z'(t)t^2dt}{\sqrt{a^2 - t^2}}. \tag{28}$$

These equations translate immediately into partial slip solution (in the Cattaneo-Mindlin approximate sense), and so

- corrective shear

$$q^*(r)/f = -\frac{2}{\pi A}\int_r^c \frac{F'(s)ds}{\sqrt{s^2 - r^2}}, \quad 0 \le r \le c, \tag{29}$$

- corrective approach of remote points

$$\alpha_n^* = c\int_0^c \frac{z'(t)dt}{\sqrt{c^2 - t^2}}, \tag{30}$$

- relation dimension of stick area and corrective load

$$Q^*/f = \frac{4}{A}\int_0^c \frac{z'(t)t^2dt}{\sqrt{c^2 - t^2}}. \tag{31}$$

In the Appendix this solution is also extended to the case with Poisson's effect, and in the paper (Ciavarella, 1998c), the case of rounded punches like a flat with rounded corners or a cone with rounded tip are discussed in detail.

4.2 General Simply Connected Area. There is no *exact* analytical solution for any contact of general simply connected area S, i.e., different from circular or elliptical shape. Therefore, we can only make use of numerical solutions. In particular, for a full stick problem, we need a solution for the complete contact for a rigid flat indenter of planform S, whereas for the partial slip conditions we need the solution of normal contact for indenters of smooth profile, and for two values of the normal load. For the former, Fabrikant provides an approximate analytical solution (Fabrikant, 1986), whereas for the latter, Barber and Billings (1990) give interesting methods to compute the relation load-penetration, the latter being an integral property. These methods can be used immediately to compute relation load-displacement in the tangential direction in the partial slip problem. Also, Conway et al. (1968) provide a series of numerical results for the relationship between load and penetration for flat punches of arbitrary cross section which are of immediate interest now in terms of the tangential contact problems.

Although having made recourse to a numerical approach, it is still advantageous when solving problems with $\gamma \neq 0$, the use of the superposition of normal contact problems, although this will give only a rough approximate solution (the approximation involved is not only the Cattaneo-Mindlin one, the problem being nonaxisymmetric), as the complete investigation requires the solution of a coupled system of two integral equations.

5 Discussion

We know that if $\gamma \neq 0$, in the case of Hertzian contacts, a solution of the kind just developed, i.e., with the shearing distribution acting in the direction of the tangential force *only*, was given by Cattaneo (1938) and Mindlin (1949) for the Hertzian three-dimensional case, although they did not notice that the solution is approximate, in the sense that there is a small *disalignment* between relative tangential displacements and shearing tractions.[7] The question, therefore, is whether there are more general cases where this solution can be developed under this Cattaneo-Mindlin approximation.

The most important cases are the plane contact, solved in Ciavarella (1998a, b), and the general axisymmetric single contact case, which is shown in Ciavarella (1998c), and the main results are briefly reported in the Appendix. This result is used in Ciavarella (1998c) for the solution of the flat and conical rounded indenters in great detail. More general cases are not only unlikely to be solved, but are also of limited interest as the surfaces required to produce such general smooth contact will be of rather special form, and the shape is also likely to change with the load.

6 Conclusions

Partial slip three-dimensional contact has been considered with new general results, in the framework of contact of bodies that can be elastically approximated as half-spaces. For the solution with a single component of shearing traction, to satisfy the problem exactly, Dundurs' constant β and combined Poisson's ratio γ have to be zero. In this case, simple properties and results can be obtained. In particular, the general solution for axisymmetric contacts has been given. The more general case has been discussed, and the effect of Poisson's ratio is also included rigorously in the general axisymmetric case.

Acknowledgments

The author wishes to thank Prof. J. R. Barber, of University of Michigan, for *invaluable* opinions on the work. Also, Dr. V. I. Fabrikant, for discussions on the general shape of the contact area. Also, Dr. D. A. Hills, of Oxford University and Lincoln College, Oxford, who made my visit there possible, and for other numerous suggestions on the work. Finally, to a referee for directing me to the paper by Conway et al. (1968).

References

Barber, J. R., 1992, *Elasticity,* Kluwer, Dordrecht, The Netherlands.

Barber, J. R., and Billings, D. A., 1990, "*An approximate solution for the contact area and elastic compliance of a smooth punch of arbitrary shape,*" *Int. J. Mech. Sci.,* Vol. 32, No. 12, pp. 991–997.

Cattaneo, C., 1938, "Sul contatto di due corpi elastici: distribuzione locale degli sforzi," *Rendiconti dell'Accademia Nazionale dei Lincei,* Vol. 27, pp. 342–348, 434–436, 474–478.

Cattaneo, C., 1947a, "Teoria del contatto elastico in seconda approssimazione," *Rendiconti di Matematica e delle sue applicazioni, Serie V,* Vol. 6, pp. 504–512.

Cattaneo, C., 1947b, "Teoria del contatto elastico in seconda approssimazione: Compressione obliqua," *Rendiconti Seminario Facolta' Scienze Università di Cagliari,* Vol. 17, pp. 13–28.

Cattaneo, C., 1955, "Compressione e torsione nel contatto tra corpi elastici di forma qualunque," *Ann. Scuola Norm. Sup. Pisa, Serie III,* Vol. 9, pp. 23–42.

Ciavarella, M., 1998a, "The Generalized Cattaneo Partial Slip Plane Contact Problem. I—Theory," *Int. J. Solids and Struct,* Vol. 35, No. 18, pp. 2349–2362.

Ciavarella, M., 1998b, "The Generalized Cattaneo Partial Slip Plane Contact Problem. II—Examples," *Int. J. Solids and Struct,* Vol. 35, No. 18, pp. 2363–2378.

[7] The exact solution implies a slightly noncircular stick zone, and requires a numerical nonaxisymmetrical solution, but the effort to do so is not justified by the small correction achieved (Munisamy et al., 1994). In the more general case of dissimilar elastic materials ($\beta \neq 0$), pressure distribution and shearing tractions in normal and a fortiori tangential problem will present a complicated pattern, and the stick-slip boundary a complicated shape, very much dependent on the actual loading path. This clearly requires an intensive numerical work.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Ciavarella, M., 1998c, "Indentation by nominally Flat or Conical Indenters with rounded Corners," *Int. J. Solids Struct.*, submitted for publication.

Conway, H. D., and Farnham, K. A., 1968, "The relationship between load and penetration for a rigid, flat-ended punch of arbitrary cross section," *Int. J. Engg. Sci.*, Vol. 6, pp. 489–496.

Deresiewicz, H., 1957, "Oblique Contact of Nonspherical Bodies," ASME JOURNAL OF APPLIED MECHANICS, Vol. 24, pp. 623–624.

Dundurs, J., 1975, "Properties of elastic bodies in contact," *Mechanics of Contact Between Deformable Bodies*, A. D. de Pater and J. J. Kalker eds., Delft University Press, Delft, The Netherlands.

Fabrikant, V. I., 1986, "Flat punch of arbitrary shape on an elastic half-shape," *Int. J. Engng. Sci.*, Vol. 24, pp. 1731–1738.

Fabrikant, V. I., and Sankar, T. S., 1988, "Singularities at angular points in elastic contact problems," *Comm. in Appl. Num. Meth.* Vol. 4, No. 2, pp. 173–178.

Goodman, L. E., 1962, "Contact Stress Analysis of Normally Loaded Rough Spheres," ASME JOURNAL OF APPLIED MECHANICS, Vol. 29, pp. 515–522.

Hetenyi, M., and McDonald, P. H., 1958, "Contact Stresses Under Combined Pressure and Twist," ASME JOURNAL OF APPLIED MECHANICS, Vol. 25, pp. 396–401.

Hills, D. A., Nowell, D., and Sackfield, A., 1993, *Mechanics of Elastic Contacts*, Butterworth-Heinemann, Oxford.

Hills, D. A., and Sackfield, A., 1986, "The Stress Field Induced by a Twisting Sphere," ASME JOURNAL OF APPLIED MECHANICS, Vol. 54, pp. 8–14.

Johnson, K. L., 1955, "Surface interaction between elastically loaded bodies under tangential forces," *Proc. R. Soc.*, Lond., Vol. A230, pp. 531–548.

Johnson, K. L., 1985, *Contact Mechanics*, Cambridge University Press, Cambridge, UK.

Kalker, J. J., 1990, *Three-dimensional elastic bodies in rolling contact*, Kluwer, Dordrecht, The Netherlands.

Lubkin, J. L., 1951, "The Torsion of Elastic Spheres in Contact," ASME JOURNAL OF APPLIED MECHANICS, Vol. 73, pp. 183–187.

MacMillan, W. D., 1936, *Dynamics of rigid bodies*, McGraw-Hill, New York.

Mindlin, R. D., 1949, "Compliance of Elastic Bodies in Contact," ASME JOURNAL OF APPLIED MECHANICS, Vol. 71, pp. 259–268.

Mindlin, R. D., Mason, W. P., Osmer, T. F., and Dereciewicz, H., 1952, "Effects of an oscillatory tangential force on the contact surfaces of elastic spheres," *Proc. 1st Nat. Congress of Applied Mechanics*, Vol. 203–208.

Mindlin, R. D., and Dereciewicz, H., 1953, "Elastic Spheres in Contact Under Varying Oblique Forces," ASME JOURNAL OF APPLIED MECHANICS, Vol. 75, pp. 327–344.

Munisamy, R. L., Hills, D. A., and Nowell, D., 1994, "Static Axisymmetrical Herzian Contacts Subject to Shearing Forces," ASME JOURNAL OF APPLIED MECHANICS, Vol. 61, pp. 278–283.

Pacelli, M., 1956, "Contatto con attrito tra due corpi elastici di forma qualunque: compressione e torsione," *Ann. Scuola Norm. Sup. Pisa, Serie III*, Vol. 10, pp. 155–184.

Scthaerman, I. Ya., 1949, *Contact Problem of the Theory of Elasticity*, Moscow, Leningrad: Gostekhteoretizdat (available from the British Library in a English translation by Foreign Technology Div., FTD-MT-24-61-70, 1970).

Truman, C. E., Sackfield, A., and Hills, D. A., 1995, "Contact mechanics of wedge and cone indenters," *Int. J. Mech. Sci.*, Vol. 37, No. 3, pp. 261–275.

APPENDIX

The Axisymmetric Case With $\gamma \neq 0$

In (Ciavarella, 1998c) it is proved that in the general case of an axisymmetrical problem, a solution of the kind just developed can be obtained. Here the proof is not repeated, but the main results are given, extending therefore what was obtained in the main part of the paper. It is true that the case $\gamma = 0$ is powerfully proved in general terms, and therefore a numerical solution of any partial slip problem can now be developed easily as the superposition of far simpler frictionless contact, but regarding strictly analytical results, only axisymmetrical problems may be of practical interest. The solution of the problem for $\gamma \neq 0$, following what obtained in Ciavarella (1998c), is

$$\frac{2\pi}{A}\iint_{S_{\text{stick}}} \frac{q_x^*(r')/f}{\rho}\, dS = u_z(r) - \frac{\delta_x}{f\left(1+\frac{\gamma}{2}\right)}, \qquad 0 \le r \le c \quad (32)$$

where c is the size of the circular stick area Sstick, r' is the radial coordinate of the integration point; the field point has radial coordinate r; and the distance between them is ρ. The latter integral equation means that with the corrective part, $q_x^*(r')/f$ can be found from a normal contact problem of the kind (2) for a *reduced* indentation. Also inequalities in the slip area translate to the no-interpenetration condition of the corrective contact problem.

All the relevant quantities can be obtained from the superposition of the normal load solutions, using the corrective load defined by $Q_x^* = fP - Q_x$. The approach in the tangential direction can be related to the superposition of approaches in normal directions through the factor $(1 + (\gamma/2))$, i.e.,

$$\frac{\alpha_t}{f} = \left(1+\frac{\gamma}{2}\right)(\alpha_n - \alpha_n^*). \quad (33)$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Hybrid Stochastic Finite Elements and Generalized Monte Carlo Simulation

R. Ghanem
Civil Engineering Department,
The Johns Hopkins University,
Baltimore, MD 21218

A procedure is developed to integrate analytical solutions to problems featuring random media with Monte Carlo simulations in order to improve the efficiency of the simulations. This is achieved by developing a common theoretical framework that encompasses Monte Carlo procedures as well as various expansion solution techniques. This framework can be perceived as a natural extension of hybrid deterministic finite element procedures whereby refinement is achieved by simultaneously increasing the number of elements as well as the degree of interpolation within each element.

1 Introduction

The mechanics of heterogeneous media presents a number of issues that, although long recognized, remain unresolved. The particular form of these issues is intimately related to the mathematical framework used to describe the heterogeneity of the medium. The treatment presented in this paper will refer to a probabilistic framework. Within this class of models, one of the most important, though elusive, issues seems to be the meaning to associate to the solution to a problem involving random media. This issue has immediate bearing on the identification of quantities to be solved for in the course of such an analysis. A large number of research in this field, for example, has aimed at evaluating approximations to the statistical moments of certain solution variables. These moments have usually been restricted to second-order moments, while higher order moments have only occasionally been evaluated. Recent research developments in stochastic finite elements have also headed in a similar direction (Deodatis 1989, 1991; Shinozuka, 1987). Even when a Monte Carlo simulation is relied upon for the analysis (Shinozuka, 1976), the results from the simulations are still used to synthesize these statistical moments. The information contained in these moments is valuable and, if evaluated accurately, provides useful, though limited, data for decision making. Mathematically, statistical moments can be viewed as yielding the various norms of the solution variables, and not the solution itself.

In a series of recent articles (Ghanem and Spanos 1990, 1991a, b, 1993) a new technique has been presented that addresses the important issue of endowing an explicit meaning to the solution of problems involving stochastic media. Specifically, the solution is identified with its coordinates with respect to a basis in an appropriate Hilbert space. A procedure has been developed for evaluating these coordinates. The random aspect of the problem is viewed as representing a new dimension. Unlike the other dimensions in the problem, where Lebesgue measure is the naturally induced measure, a probability measure is naturally induced along this new dimension. Moreover, whereas Euclidian discretization and an associated grid (finite elements or finite difference) can be readily associated with any variable defined on the Euclidian dimensions, no such expeditious geometrical interpretation can be associated with discretizations along the random dimension. Indeed, a global (i.e., of nonbounded support) spectral discretization is used along this dimension. This discretization is associated with basis functions having global support over the random dimension. A procedure has been developed for evaluating the projection of the solution process on a suitable basis in an appropriate Hilbert space. This provides a convergent expansion of the solution process in terms of a basis of random variables. This expansion can then be utilized to evaluate norms of the solution, resulting in statistical moments, or to evaluate integrals over domains defined by the solution, resulting in probabilities of various events. In this paper, the spectral formulation just described is generalized to include Monte Carlo simulation as a special case, thus allowing for it to be combined with a Monte Carlo simulation solution of the problem.

2 Representation of the Randomness in the System

Consider the scalar random process $h(\mathbf{x}; \theta)$ with mean

$$\bar{h}(\mathbf{x}) = E\{h(\mathbf{x}; \theta)\} \tag{1}$$

and covariance

$$C^h(\mathbf{x}, \mathbf{y}) = E\{(h(\mathbf{x}; \theta) - \bar{h}(\mathbf{x}))(h(\mathbf{y}; \theta) - \bar{h}(\mathbf{y}))\}. \tag{2}$$

Here θ is used to reference random quantities and can therefore be considered to index a new dimension, referred to henceforth as the random dimension. The covariance function is a symmetrical and positive definite function. Therefore it can be represented using the following expansion:

$$C^h(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{\infty} \lambda_n^h \phi_n^h(\mathbf{x}) \phi_n^h(\mathbf{y}), \tag{3}$$

where $\phi_n^h(\mathbf{x})$ and λ_n^h denote, respectively, the eigenfunctions and eigenvalues obtained as the solution to the following integral equation:

$$\int_{\mathcal{D}} C^h(\mathbf{x}, \mathbf{y}) \phi_n^h(\mathbf{x}) d\mathbf{x} = \lambda_n^h \phi_n^h(\mathbf{y}), \tag{4}$$

and $\mathcal{D}$ denotes the spatial domain of definition of the stochastic process $h(\mathbf{x}, \theta)$. Due to the symmetry of the covariance function, its eigenfunctions are orthogonal. Furthermore, they are complete in the space L_2 of square-integrable functions, and can therefore be used to expand the process $h(\mathbf{x}; \theta)$. The resulting expansion is known as the Karhunen-Loeve expansion (Loeve, 1948; Spanos and Ghanem, 1989) and is given by

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, July 2, 1997; final revision, Jan. 10, 1998. Associate Technical Editor: M. Shinozuka.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$h(\mathbf{x};\theta) = \bar{h}(\mathbf{x}) + \sum_{n=1}^{\infty} \sqrt{\lambda_n^h}\,\xi_n^h(\theta)\phi_n^h(\mathbf{x}). \quad (5)$$

It can be shown that the above expansion is optimal in the mean-square sense. In the case where the process is Gaussian, the random variables $\xi_n^h(\theta)$ form an orthogonal Gaussian vector of mean zero and covariance equal to the identity matrix. Expansions similar to the Karhunen-Loeve expansion have been used in the literature to represent stationary process extending over infinite domains (Shinozuka and Deodatis, 1991). Furthermore, expansions that rely on orthogonal functions have also been used. These functions have been associated with self-adjoint operators (Elishakoff, 1979), orthogonal polynomials (Zhang and Ellingwood, 1994), or trigonometric polynomials (Grigoriu, 1993). In these cases, however, the orthogonality of the random variables used in the expansion is lost but can be recovered by solving an associated matrix eigenvalue problem, effectively transforming the problem back into a vector Karhunen-Loeve expansion.

3 Representation of the Response Process

Since the covariance of the response process is not known a priori, the Karhunen-Loeve expansion cannot be used for its representation. Instead, the polynomial chaos expansion is used to represent the response process (Ghanem and Spanos, 1991a). This expansion is based on a theorem by Cameron and Martin (1947), and provides a means for expanding second-order random variables in terms of orthogonal polynomials in uncorrelated standard Gaussian variables. Second-order random variables refer to random variables with finite variance. Thus, a general second-order random variable $X(\theta)$, viewed as a function of θ, can be represented in the form,

$$X(\theta) = a_0\Gamma_0 + \sum_{i_1=1}^{\infty} a_{i_1}\Gamma_1(\xi_{i_1}(\theta)) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} a_{i_1 i_2}\Gamma_2(\xi_{i_1}(\theta), \xi_{i_2}(\theta)) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}\Gamma_3(\xi_{i_1}(\theta), \xi_{i_2}(\theta), \xi_{i_3}(\theta)) + \ldots. \quad (6)$$

In this equation, the symbol $\Gamma_n(\xi_{i_1}, \ldots, \xi_{i_n})$ denotes the polynomial chaos of order n in the variables $(\xi_{i_1}, \ldots, \xi_{i_n})$. The general expression for generating this polynomial is given by

$$\Gamma_n(\xi_{i_1}, \ldots, \xi_{i_n}) = e^{(1/2)\boldsymbol{\xi}^T\boldsymbol{\xi}}(-1)^n \frac{\partial^n}{\partial\xi_{i_1}\ldots\xi_{i_n}} e^{-(1/2)\boldsymbol{\xi}^T\boldsymbol{\xi}}, \quad (7)$$

where $\boldsymbol{\xi}$ denotes the vector consisting of the n random variables $(\xi_{i_1}, \ldots, \xi_{i_n})$. Equation (6) can be rewritten, for notational convenience, in the following form:

$$X(\theta) = \sum_{j=0}^{\infty} \hat{a}_j\Psi_j(\boldsymbol{\xi}), \quad (8)$$

where there is a one-to-one correspondence between the functions $\Gamma_n(\xi_{i_1}, \ldots, \xi_{i_n})$ and $\Psi_j(\boldsymbol{\xi})$, as well as between the coefficients in both expansions. Noting that the polynomial chaoses form an orthogonal basis in the L_2 space of random variables, the coefficients $\hat{a}_j$ in this last expansion can be explicitly obtained in the form

$$\hat{a}_j = \frac{\langle X(\theta), \Psi_j\rangle}{\langle\Psi_j, \Psi_j\rangle}, \quad j = 0, 1, \ldots, \quad (9)$$

where $\langle ., .\rangle$ denotes the inner product in the Hilbert space of random variables. This inner product is defined as the expected value of the product of its two arguments. Specifically, $\langle X, Y\rangle = E[XY]$. The notation $\langle XY\rangle$ will also be used to denote this same quantity.

4 Spectral Stochastic Finite Element Method

A standard form resulting from the spatial discretization of partial differential equations is given by the following equation:

$$\mathbf{K}(\mathbf{u}, h(\theta))\mathbf{u}(\theta) = \mathbf{f}(\theta), \quad (10)$$

where the nonlinear dependence of matrix $\mathbf{K}$ on the solution vector $\mathbf{u}$, and the random nature of the various quantities involved are explicitly shown. Moreover, the coefficient matrix $\mathbf{K}$ is random through its dependence on a vector stochastic process, $\mathbf{h}$. Equation (10) could be associated with either a load-step or a time-step increment. In the latter case, the right-hand side would involve quantities evaluated at the previous time step. Representing the process h in terms of its Karhunen-Loeve expansion, and using the polynomial chaos expansion (Ghanem and Spanos, 1991) for the solution process, results in the following equation:

$$\mathbf{u} = \sum_{j=0}^{N} \mathbf{u}_j\Psi_j \quad (11)$$

$$\sum_{j=0}^{N} \Psi_j\mathbf{K}(\mathbf{u}, \boldsymbol{\xi}(\theta))\mathbf{u}_j = \mathbf{f}(\theta), \quad (12)$$

where the vector of uncorrelated random variables, $\boldsymbol{\xi}$, now replaces the stochastic process h. The error in the above equation is constrained to be orthogonal to the approximating space. This requirement results in the following set of equations

$$\sum_{j=0}^{N} \langle\Psi_k\Psi_j\mathbf{K}(\mathbf{u}, \boldsymbol{\xi}(\theta))\rangle\mathbf{u}_j = \langle\Psi_k\mathbf{f}(\theta)\rangle \quad k = 0, 1, \ldots. \quad (13)$$

Clearly, since $\mathbf{K}$ depends on the solution $\mathbf{u}$, it is not a linear functional of the basic random variables $\{\xi(\theta)\}$, and it must therefore be expanded with respect to the polynomial chaos basis. Representing $\mathbf{K}$ in its polynomial chaos expansion,

$$\mathbf{K} = \sum_{i=0}^{L} \Psi_i\mathbf{K}_i, \quad \mathbf{K}_i = \frac{\langle\Psi_i\mathbf{K}\rangle}{\langle\Psi_i^2\rangle}, \quad (14)$$

Eq. (13) becomes

$$\sum_{j=0}^{N}\sum_{i=0}^{L} \langle\Psi_k\Psi_j\Psi_i\rangle\mathbf{K}_i\mathbf{u}_j = \langle\Psi_k\mathbf{f}(\theta)\rangle \quad k = 0, 1, \ldots. \quad (15)$$

5 Generalized Monte Carlo Simulation

The formalism developed in the previous section can be seen to encompass the Monte Carlo simulation technique as a special case. In particular, if the functions $\Psi_i(\theta)$, used in Eq. (8), instead of referencing the polynomial chaos basis, referred to delta functions $\delta(\theta - \theta_i)$, a collocation-like procedure is obtained along the random dimension θ. The coefficients of the delta functions in this expansion correspond to individual realizations of the process being represented; these realizations being sampled at the coordinates $\{\theta_i\}$. Implementing this expansion into the formulation presented in this paper results in a set of algebraic equations to be solved, at every iteration of every time-step, of the following general form:

$$\begin{bmatrix} \mathbf{K}^{(1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}^{(2)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}^{(\text{nsimul})} \end{bmatrix} \begin{Bmatrix} \mathbf{u} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_{\text{nsimul}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f}_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{f}_{\text{nsimul}} \end{Bmatrix}. \quad (16)$$

The off-diagonal submatrices, in these equations, are all zero,

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

due to the fact that the coefficients $d_{ijk} = \langle \Psi_i \Psi_j \Psi_k \rangle$, in this case, are all zero unless $i = j = k$. It is clear, then, that in this case, the number of unknown subvectors, nsimul, is equal to the number of independent simulations to be performed in the context of a Monte Carlo simulation procedure. The corresponding equation associated with the polynomial chaos basis has the general form

$$\begin{bmatrix} \mathbf{K}^{(11)} & \mathbf{K}^{(12)} & \cdots & \mathbf{K}^{(1N)} \\ \mathbf{K}^{(21)} & \mathbf{K}^{(22)} & \cdots & \mathbf{K}^{(22)} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{K}^{(N1)} & \mathbf{K}^{(N2)} & \cdots & \mathbf{K}^{(NN)} \end{bmatrix} \begin{Bmatrix} \mathbf{u}_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_N \end{Bmatrix} = \begin{Bmatrix} \mathbf{f}_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{f}_N \end{Bmatrix}. \quad (17)$$

In this equation, $\mathbf{K}^{(jk)} = \sum_{i=1}^{L} \langle \Psi_i \Psi_j \Psi_k \rangle \mathbf{K}_i$, and coupling occurs between the various submatrices, and the index N refers to the number of terms retained in the polynomial chaos expansion. The above discussion can be used as a vehicle to assess the relative efficiency of the procedure proposed in this paper in comparison to a traditional Monte Carlo simulation. Noting that, in general, *nsimul* is much larger that N, suggests that depending on the solution procedure used (matrix storage algorithm, iterative solution procedure, etc. . . .), there may be a threshold level for which one procedure may be more efficient than the other. It is noted in this context that the number of iterations required in any iterative scheme is dependent on the condition number of the algebraic system being solved, which itself depends on the level of variability in the parameters of the system, being worse with increasing fluctuations in the properties of the system. Given this complex dependence of the computational efficiency on the level of uncertainty in the material properties and on the level of nonlinearity in the system, it is not clear at this stage that a simple identification of a threshold level is possible. An important outcome of the above discussion, however, is the new perspective on Monte Carlo simulation procedures, and their embedding as a special case of the procedure proposed in this paper. Indeed, this new perspective paves the way for the development of the hybrid stochastic finite element method detailed in the next section.

6 Hybrid Stochastic Finite Element Method

Using an N term polynomial chaos expansion for the solution process, and expanding the residual vector in terms of a basis set $\Phi(\theta)$, results in

$$\mathbf{u} = \sum_{i=0}^{N} \mathbf{u}_i \Psi_i + \sum_{i=0}^{M} \mathbf{u}_i^* \Phi_i. \quad (18)$$

In view of the orthogonality of the polynomial chaoses, the first summation on the right-hand side of the above equation represents the projection of the solution on the subspace spanned by the first N polynomial chaoses, while the second term represents the projection on its orthogonal complement. In the following, the expansion for the residual will be associated with a Monte Carlo simulation procedure. This implies that the set Φ_i is identified with the sequence of delta function, $\delta(\theta - \theta_i)$. Thus, the inner product of any random variable with a function Φ_i is given by

$$\langle f(\theta) \Phi_i(\theta) \rangle = f(\theta_i), \quad (19)$$

and results in the realization $f(\theta_i)$ of f. It should be noted that the random variable Φ_i should be viewed as a generalized function defined through its action on other random variables, this action being specified via Eq. (19). Substituting Eq. (18) into Eq. (10) and taking the inner product with Ψ_k and Φ_k, respectively, results in the following two systems of equations:

$$\sum_{j=0}^{N} \sum_{i=0}^{L} \langle \Psi_i \Psi_j \Psi_k \rangle \mathbf{K}_i \mathbf{u}_j = \langle \Psi_k \mathbf{f} \rangle - \sum_{j=0}^{M} \sum_{i=0}^{L} \mathbf{K}_i \langle \Psi_i \Phi_j \Psi_k \rangle \mathbf{u}_j^* \quad k = 0, \ldots . \quad (20)$$

and

$$\langle \mathbf{K} \mathbf{u}^* \Phi_k \rangle = \langle \mathbf{f}\, \Phi_k \rangle - \sum_{j=0}^{N} \sum_{i=0}^{L} \langle \Psi_i \Psi_j \Phi_k \rangle \mathbf{K}_i \mathbf{u}_j \quad k = 0, \ldots . \quad (21)$$

Equation (20) is associated with Eq. (16), while Eq. (21) can be associated with Eq. (17). A new equation, similar to Eqs. (16) and (17), can be written down that symbolically captures the hybrid stochastic finite element procedure presented herein. Specifically, Eqs. (20) and (21) can be grouped into the following equation:

$$\begin{bmatrix} \mathbf{K}^1 & \cdots & 0 & & & & \\ \mathbf{0} & \cdots & 0 & & [\mathbf{K}^{(\text{hybrid})}] & & \\ \mathbf{0} & \cdots & \mathbf{K}^{(\text{nsimul})} & & & & \\ & & & \mathbf{K}^{(11)} & \cdot & \cdot & \cdot & \mathbf{K}^{(1N)} \\ & & & \mathbf{K}^{(21)} & \cdot & \cdot & \cdot & \mathbf{K}^{(22)} \\ & & & \cdot & \cdot & \cdot & \cdot & \cdot \\ & [\mathbf{K}^{(\text{hybrid})}] & & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & \mathbf{K}^{(N1)} & \cdot & \cdot & \cdot & \mathbf{K}^{(NN)} \end{bmatrix} \times \begin{Bmatrix} \mathbf{u}_1^* \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_{\text{nsimul}}^* \\ \mathbf{u}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_N \end{Bmatrix} = \begin{Bmatrix} \mathbf{f}_1^* \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{f}_{\text{nsimul}}^* \\ \mathbf{f}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{f}_N \end{Bmatrix}, \quad (22)$$

where $\mathbf{K}^{(\text{hybrid})}$ represents a coupling term between the two components of the solution. Equations (20) and (21) are indeed coupled through the terms, $\langle \Psi_i \Psi_j \Phi_k \rangle \mathbf{K}_i \mathbf{u}_j$ and $\langle \Psi_i \Psi_k \rangle \mathbf{K}_i \mathbf{u}_j^*$. This last average is evaluated as the arithmetic mean of the solution of Eq. (21) for different values of k. The double summation on the right-hand side of Eq. (20) can be rewritten as,

$$\sum_{j=0}^{M} \sum_{i=0}^{L} \mathbf{K}_i \langle \Psi_i \Phi_j \Psi_k \rangle \mathbf{u}_j^* = \sum_{i=0}^{L} \mathbf{K}_i \sum_{j=0}^{M} \mathbf{u}_j^* \Psi_k(\theta_j) \Psi_i(\theta_j) = \sum_{j=0}^{M} \mathbf{K}(\theta_j) \mathbf{u}_j^* \Psi_k(\theta_j), \quad (23)$$

where all the terms on the right-hand side of Eq. (23) refer to a particular realization of the corresponding random quantities (associated with the θ_j). Clearly, this equation represents the projection of the second term on the right-hand side of Eq. (18) on the polynomial chaoses of order less than or equal to N. Given the orthogonality of the space spanned by these polynomial chaoses to the subspace spanned by $\{\Phi_i\}$, as indicated at the beginning of this section, this projection is equal to zero. Of course, the validity of the present argument depends on the coordinates of the solution process with respect to the first N polynomial chaoses having been evaluated exactly. In general this is not true, and the value of these coefficients converges to their true values as the order of the polynomial chaos expansion is increased. This discrepancy is generally small for the lower-order terms, and increases for the higher-order ones. Thus, as an approximation, the interaction between Eq. (20) and (21) can be neglected and Eq. (20) can be rewritten as

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\sum_{j=0}^{N}\sum_{i=0}^{L}\langle\Psi_i\Psi_j\Psi_k\rangle\mathbf{K}_i\mathbf{u}_j = \langle\Psi_k\mathbf{f}\rangle \quad k = 0, \ldots. \tag{24}$$

This decoupling eliminates the need for an iterative solution process whereby the values of the polynomial chaos coefficients, $\mathbf{u}_j$, are updated, using Eq. (20), following a certain number of simulations, $\mathbf{u}_j^*$.

From the orthogonality property of the polynomial chaos,

$$\mathbf{R}_{uu} = \sum_{i=1}^{N}\mathbf{u}_i^T\mathbf{u}_i\langle\Psi_i^2\rangle + \mathbf{R}_{u^*u^*}, \tag{25}$$

where $\mathbf{R}_{uu}$ and $\mathbf{R}_{u^*u^*}$ denote the covariance matrices of $\mathbf{u}$ and $\mathbf{u}^*$, respectively. It is therefore clear that the process $\mathbf{u}^*$ has a smaller variance than the original solution process $\mathbf{u}$. This clearly indicates that it requires less realizations to achieve a certain accuracy with specified confidence level. Specifically, for a Gaussian random variable, the number of samples required to achieve accuracy greater than δ in estimating the mean value, given a confidence interval of width Z_α, is given by

$$n = \frac{Z_\alpha^2\sigma^2}{\delta^2}, \tag{26}$$

where Z_α denotes the width of the confidence interval, and δ denotes a specified tolerance. Clearly, although the solution to a stochastic partial differential equation is not a Gaussian process, the foregoing analysis still provides useful guidelines regarding the required number of realizations. Thus if enough terms are taken in the polynomial chaos expansion to reduce the variance of the solution process by a half, it can be expected that four times less samples are required in the simulation of the deviation of the polynomial chaos expansion from the exact process. A compromise can thus be achieved between the number of simulations and the number of terms used in the polynomial chaos expansion of the solution.

The foregoing analysis can also be cast in the framework of variance reduction techniques. Denoting the probability of failure of the system whose response is sought by P_f results in

$$P_f = \int_{\mathbf{x}\in\mathcal{D}_f} f_{\mathbf{x}}(\mathbf{x})d\mathbf{x}, \tag{27}$$

where $\mathcal{D}_f$ denotes the failure domain of the system, and $f_{\mathbf{x}}(\mathbf{x})$ denotes the joint density function of the random variables $\mathbf{x}$. Typical variance reduction techniques are based on rewriting the probability of failure in the form

$$P_f = \int_{\mathbf{x}\in\mathcal{D}_f}\frac{f_{\mathbf{x}}(\mathbf{x})}{f_S(\mathbf{x})}f_S(\mathbf{x})d\mathbf{x}, \tag{28}$$

where $f_S(\mathbf{x})$ is some new sampling density function, thus effectively weighting the original density function in such a way as to favor the occurrence of rare events. The procedure presented in this paper can be viewed, instead, as rewriting the probability of failure in the following form:

$$P_f = \int_{\mathbf{x}_s+\mathbf{x}^*\in\mathcal{D}_f} f_{\mathbf{x}^*}d\mathbf{x}^* = \int_{\mathbf{x}^*\in\mathcal{D}_f-\mathbf{x}_s} f_{\mathbf{x}^*}d\mathbf{x}^*. \tag{29}$$

In the above, the variables $\mathbf{x}_s$ represent the part of the solution corresponding to the spectral expansion, and the integration is over the failure domain $\mathcal{D}_f$ associated with the overall solution, $\mathbf{x}_s + \mathbf{x}^*$. The proposed technique can therefore be viewed as enhancing the Monte Carlo simulation procedure by modifying the domain of simulation instead of weighing the current domain as is typical with variance reduction techniques. Of course, more advanced procedures for enhancing Monte Carlo simulation have recently been developed (Pradlwarter and Schueller, 1997), but a review of their performance is outside the scope of this paper.

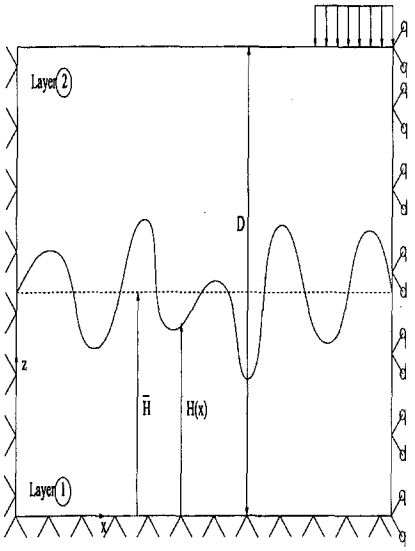


Fig. 1 Randomly layered medium

The procedure for implementing the proposed formulation consists therefore of the following steps:

1 Solve for the first N polynomial chaos coefficients, using Eq. (20) and neglecting the second term on the right-hand side.
2 Adjust the right-hand side of Eq. (21) using the polynomial chaos coefficients $\mathbf{u}_i$.
3 Perform Monte Carlo simulations (i.e., solve Eq. (21)) with the modified right-hand side.
4 After a number of simulations, Eq. (20) could be solved again, with the second term on the right-hand side providing a correction.

The added accuracy provided by step 4 above will reduce the variance of the residual in the approximation, and thus permit its evaluation with fewer Monte Carlo simulations.

7 Numerical Example

The physical model to be treated in the paper is shown in Fig. 1 and involves a linear elastic soil medium consisting of two layers, layer (1) at the bottom, and layer (2) at the top. The medium is subjected to a load applied on a portion of the surface, and a plane-strain condition is assumed to prevail. Each of the two layers has well-defined properties such as a modulus of elasticity $E^{(i)}$ and a Poisson ratio $\mu^{(i)}$, where index i refers to a particular layer. The vertical position of the interface between the two layers, however, is assumed to be the realization of a Gaussian stochastic process $h(\mathbf{x};\theta)$. Here $\mathbf{x}$ denotes the horizontal position within the medium. This random process is assumed to have a mean value $\bar{h}(\mathbf{x})$ and covariance function $C^h(\mathbf{x}, \mathbf{y})$.

The detailed analysis of this problem is developed elsewhere (Ghanem and Brzkala, 1996). The global stiffness matrix for this problem can be assembled from elemental stiffness matrices, each of which is defined over a finite element. The general form of these elemental matrices is given by the following equation:

$$\mathbf{K}^e(\theta) \equiv \mathbf{K}^e[E;\nu] = \int_{D^e} E(\mathbf{x};\theta)\mathbf{B}^{eT}(\mathbf{x})\mathbf{P}^e(\mathbf{x},\nu;\theta)\mathbf{B}^e(\mathbf{x})d\mathbf{x}. \tag{30}$$

In the above equation, $\mathbf{P}^e$ denotes the constitutive matrix of the elastic medium assumed to be in a plane-strain condition, and $\mathbf{B}^e(\mathbf{x})$ denotes a matrix of the derivatives of the finite element interpolation functions; bilinear interpolation is used in this ex-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

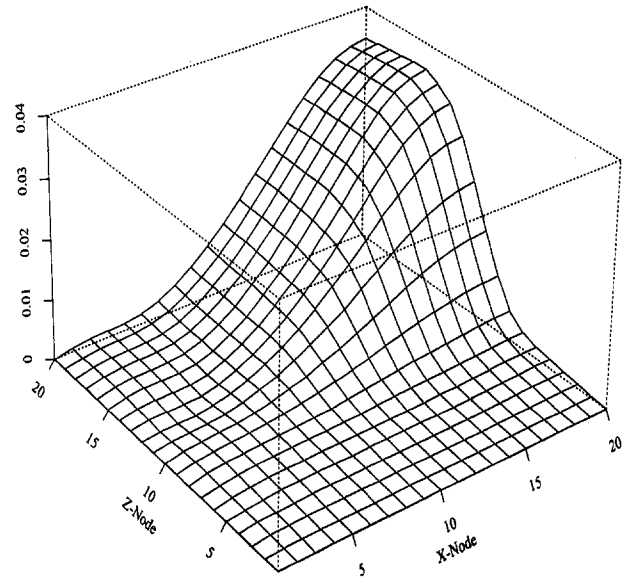


Fig. 2 Variance of the transverse displacement field throughout the domain; 100 simulation. Four terms in K-L expansion; correlation length of exponential covariance function = 10.

amples. Furthermore, the argument θ indicates that the modulus of elasticity and Poisson's ratio are themselves random quantities. The corresponding two random processes are binary since they can only have one of two possible values, depending on whether the associated spatial point is located in the top layer or in the bottom layer. Denoting by $\mathcal{I}[\cdot]$ a step function that equals one if its argument is positive, and zero otherwise, Eq. (30) can be rewritten as

$$\mathbf{K}^e(\theta) = \mathbf{K}^e[E^{(1)}(\mathbf{x})\mathcal{I}[h(\mathbf{x};\theta) - z];\nu^{(1)}] + \mathbf{K}^e[E^{(2)}(\mathbf{x})\mathcal{I}[z - h(\mathbf{x};\theta)];\nu^{(2)}], \quad (31)$$

where superscripts (1) and (2) over E and μ refer to the particular layer of interest, and the variable z denotes the vertical position of a point in element e, measured from the bottom of layer (1). The validity of Eq. (31) hinges on the fact that the functional $\mathbf{K}^e[\cdot;\cdot]$ defined in Eq. (30) is linear in $E^{(i)}$ and therefore $\mathbf{K}^e[0;\cdot] = 0$. Based on this property, it can be seen that only one term in the above equation is nonzero at any location $\mathbf{x}$ in the domain $\mathcal{D}$. The significance of Eq. (13) stems from the fact that it incorporates the geometric randomness of the problems into a coefficient of the associated differential equation. Obviously, the elemental stiffness matrix is random and it depends in a nonlinear manner on the Gaussian process describing the interface position. The polynomial chaos expansion is used to represent this matrix in the form given by Eq. (14). The coefficient matrices $\mathbf{K}_i^e$, appearing in Eq. (14), represent the orthogonal projection of the stiffness matrix $\mathbf{K}^e$ onto successive members of the polynomial chaos basis. Substituting Eq. (31) into the numerator of Eq. (14) results in

$$\langle \mathbf{K}^e(\theta), \Psi_i \rangle = \langle \Psi_i \sum_{k=1}^{2} \mathbf{K}^e[E^{(k)}(\mathbf{x})\mathcal{I}[(-1)^k(z - h(\mathbf{x};\theta))];\nu^{(k)}] \rangle, \quad (32)$$

which has been evaluated analytically elsewhere (Ghanem and Brzkala, 1996). Using these coefficients, matrices $\mathbf{K}_i$ given by Eq. (14) are assembled and Eq. (15) is solved for the coefficients in the spectral expansion for the solution process. At this point, Eq. (21) can be solved for $\mathbf{u}^*$, each time resulting in an independent simulation of this process.

Figure 2 shows a three-dimensional rendition of the variance of the vertical displacement field throughout the domain. One thousand (1000) simulations were used in the simulation. Next, the problem is modeled using the polynomial Chaos theory with a first-order expansion. Thus a linearized version of the solution is computed using an expansion along the polynomial chaos basis followed by a weighted residual minimization process. A residual error results following this procedure, that can be simulated using the concepts presented in this paper. Figure 3 shows the variance of this residual, obtained via simulation. Comparing this figure to the previous figure clearly indicates that the residual process has almost half the maximum variance as the original solution process, and can therefore be simulated, to the same accuracy, using one-fourth of the number of simulations. Of course, the overall computational cost of the procedure described in this section involves in addition to the reduced number of simulations, the cost of performing the first-order polynomial expansion.

8 Conclusion

This paper presents a unifying framework that integrates spectral expansions for the solution of stochastic partial differential equations (pde's with coefficients that are stochastic processes) and Monte Carlo simulation methods. This framework is analogous to the unification of p and h techniques in deterministic finite elements. Lower-order polynomial chaos expansions are used to analytically represent a significant portion of the uncertainty of the solution process. The residual associated with this approximation requires far fewer simulations than the original solution for a similar level of accuracy. A procedure is also presented that relies on this framework in order to efficiently solve for a representation of the solution process. Such procedures are essential for the development of the next generation tools for computational mechanics. Indeed, by developing frameworks that combine the versatility of Monte Carlo simulation with the global convergence properties of spectral expansions, large-scale problems featuring stochastic coefficients can be more readily tackled in a computational setting.

Acknowledgments

This research was supported by NSF Grant No. CMS-9596238.

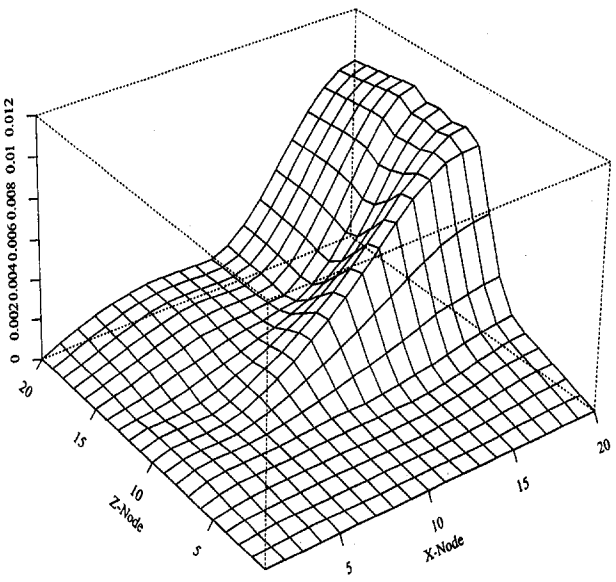


Fig. 3 Variance of the error in the transverse displacement field throughout the domain; first-order polynomial chaos expansion of the solution process; 100 simulation. Four terms in K-L expansion; correlation length of exponential covariance function = 10.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

References

Cameron, R. H., and Martin, W. T., 1947, "The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals," *Ann. Math,* Vol. 48, pp. 385–392.

Deodatis, G., 1989, "Bounds on response variability of stochastic finite element systems," *ASCE Journal of Engineering Mechanics,* Vol. 115, No. 11.

Deodatis, G., 1991, "Weighted integral method, I: Stochastic Stiffness method," *ASCE Journal of Engineering Mechanics,* Vol. 117, No. 8, pp. 1851–1864.

Elishakoff, I., 1979, "Simulation of space-random fields for solution of stochastic boundary-value problems," *Journal of the Acoustical Society of America,* Vol. 65, pp. 399–403.

Ghanem, R., and Spanos, P., 1990, "Polynomial Chaos in Stochastic Finite Elements," ASME JOURNAL OF APPLIED MECHANICS, Vol. 57, pp. 197–202.

Ghanem, R., and Spanos, P., 1991a, "Spectral stochastic finite-element formulation for reliability analysis," *ASCE Journal of Engineering Mechanics,* Vol. 117, No. 10, pp. 2351–2372.

Ghanem, R., and Spanos, P., 1991b, *Stochastic Finite Elements: A Spectral Approach,* Springer-Verlag, New York.

Ghanem, R., and Spanos, P., 1993, "A new computational approach for nonlinear random vibration," *Probabilistic Engineering Mechanics,* Vol. 8, pp. 255–264.

Ghanem, R., and Brzkala, V., 1996, "Stochastic finite element analysis for random layered media," *ASCE Journal of Engineering Mechanics,* Vol. 122, No. 4, pp. 361–369.

Grigoriu, M., 1993, "Simulation of Nonstationary Gaussian Processes by Random Trigonometric Polynomials," Vol. 119, No. 2, pp. 328–343.

Loeve, M., 1948, "Fonctions aleatoires du second ordre," supplement to P. Levy, *Processus Stochastic et Mouvement Brownien,* Gauthier Villars, Paris.

Pradlwarter, H. J., and Schueller, G. I., 1947, "On advanced Monte Carlo simulation procedures in stochastic structural dynamics," *International Journal of Non-linear Mechanics,* to appear.

Shinozuka, M., 1987, "Structural response variability," *ASCE Journal of Engineering Mechanics,* Vol. 113, No. 6, pp. 825–842.

Shinozuka, M., and Deodatis, G., 1991, "Simulation of stochastic processes by spectral representation," ASME *Applied Mechanics Reviews,* Vol. 44, pp. 191–204.

Shinozuka, M., and Lenoe, E., 1976, "A probabilistic model for spatial distribution of material properties," *Eng. Fracture Mechanics,* Vol. 8, No. 1, pp. 217–227.

Spanos, P., and Ghanem, R., 1989, "Stochastic finite element expansion for random media," *ASCE Journal of Engineering Mechanics,* Vol. 115, No. 5, pp. 1035–1053.

Zhang, J., and Ellingwood, B., "Orthogonal series expansion of random fields in first-order reliability analysis," ASME *Journal of Engineering Mechanics,* to appear.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Identification of a Dynamic System Using Ambient Vibration Measurements

M. Q. Feng
Associate Professor,
Department of Civil and Environmental Engineering,
University of California,
Irvine, CA 92697-2175

J.-M. Kim
Assistant Professor,
Department of Ocean Civil Engineering,
Yosu National Fisheries University,
96-1 Dunduck-dong,
Yosu, Chonnam 550-250,
Korea

H. Xue
Senior Lecturer,
Department of Mechanical and Production Engineering,
National University of Singapore,
10 Kent Ridge Crescent,
Singapore 0511

This paper demonstrates how ambient vibration measurements at a limited number of locations can be effectively utilized to estimate parameters of a finite element model of a large-scale structural system involving a large number of elements. System identification using ambient vibration measurements presents a challenge requiring the use of special identification techniques, which can deal with very small magnitudes of ambient vibration contaminated by noise without the knowledge of input forces. In the present study, the modal parameters such as natural frequencies, damping ratios, and mode shapes of the structural system were estimated by means of appropriate system identification techniques including the random decrement method. Moreover, estimation of parameters such as the stiffness matrix of the finite element model from the system response measured by a limited number of sensors is another challenge. In this study, the system stiffness matrix was estimated by using the quadratic optimization involving the computed and measured modal strain energy of the system, with the aid of a sensitivity relationship between each element stiffness and the modal parameters established by the second-order inverse modal perturbation theory. The finite element models thus identified represent the actual structural system very well, as their calculated dynamic characteristics satisfactorily matched the observed ones from the ambient vibration test performed on a large-scale structural system subjected primarily to ambient wind excitations. It is noted that newly developed optical fiber accelerometers were used for this ambient vibration test. The dynamic models identified by this study will be used for design of an active mass damper system to be installed on this structure for suppressing its wind vibration.

1 Introduction

As a U.S.-China collaborative research project, an active mass damper system is currently under development for a 310-m tall TV tower recently constructed in the city of Nanjing, China. Figure 1 shows the main structure of the tower. It consists of three prestressed concrete legs placed in 120 deg with each other, each of which has a box section with four compartments. The legs are incorporated into a space frame with seven coupling ring beams placed along the height. There is a large observation hall at the level of 185 m and a small observation hall at 240 m, connected by a reinforced concrete cylinder tube structure (Cheng, 1992; Ding and Ren, 1994). The objective of the active control is to reduce the wind response of the tower, especially at the small observation hall and the antenna at the top of the tower, in order to improve the human comfort of the tourists and operators as well as to enhance the tower's communication and broadcasting quality. In May of 1995, the first author of this paper and her associates performed an ambient vibration test on this structural system, for the purpose of establishing an accurate and reliable dynamic model of the tower for the use in the design of the active vibration control system.

The sensors used for this test were optical fiber accelerometers recently developed by the first author and her associates. Because of their advantages such as immunity to electromagnetic fields and ease in installation, optical fiber sensors recently attracted great attention in the civil engineering community. Recent developments and applications of optical fiber sensors in civil engineering are summarized in papers by Ansari (1993) and Kersey (1994). The optical fiber accelerometer used for this field testing, however, has an unique configuration totally different from those existing optical fiber sensors summarized in these papers. Each durable sensor head consists of an oscillatory circuit and transforms a measured quantity into a frequency-modulated (FM) signal with proven reliability even over a long-distance transmission. The details of the sensing mechanism are given in (Feng, 1996) based on which a similar sensor was developed earlier. The optical fiber accelerometers were extensively tested and proven to be accurate with a sufficient dynamic range (up to 25 Hz) before being used for this field measurement.

The acceleration response of the structural system measured by the optical fiber sensors under ambient conditions was used to identify the dynamic characteristics of the structural system. System identification using ambient vibration measurements presents a challenge requiring the use of special identification techniques, which can deal with very small magnitudes of ambient vibration contaminated by noise in the situation where input forces usually cannot be measured.

Studies have been conducted on identification of modal parameters including natural frequencies, modal damping ratios, and mode shapes of structural systems using ambient vibration records without input measurements (see Wang and Haldar (1994) for an extensive literature review). Mathematical for-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, July 7, 1997; final revision, Sept. 27, 1997. Associate Technical Editor: M. Shinozuka.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

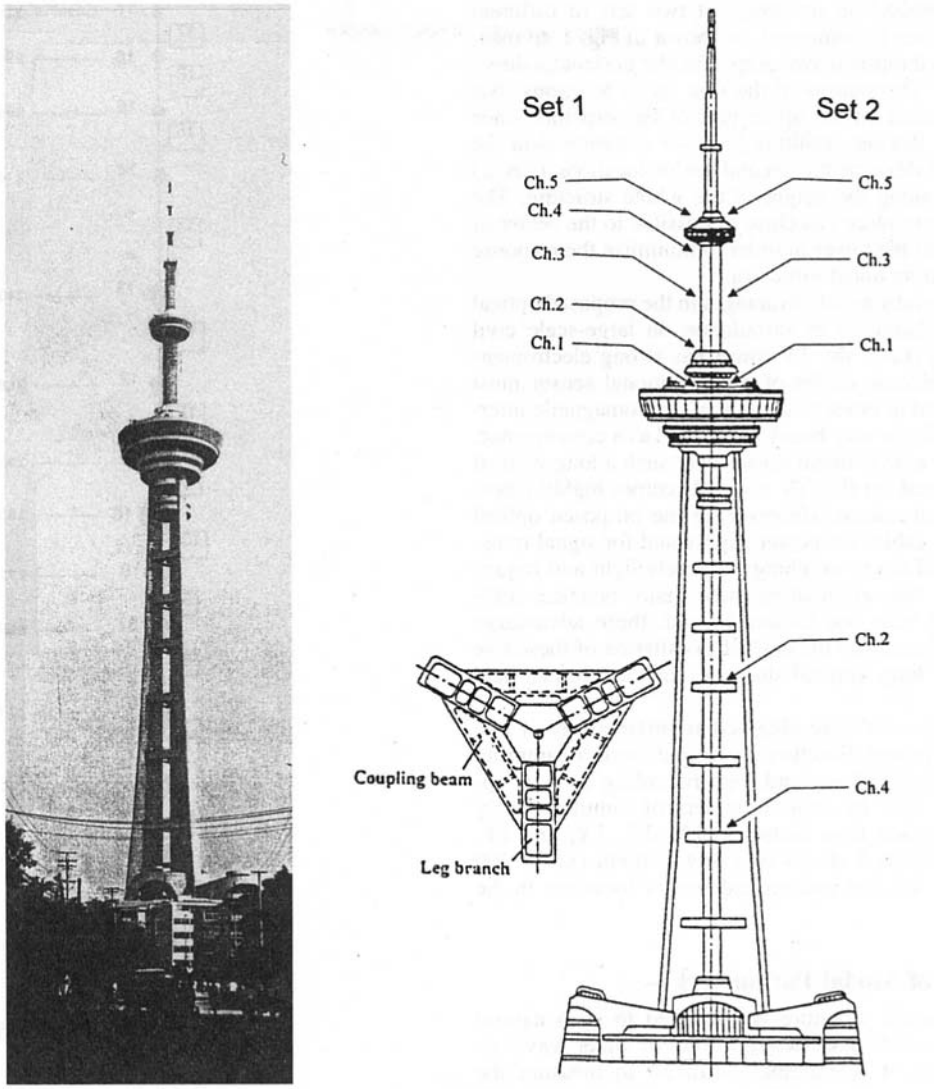


Fig. 1 Elevation, plan, and sensor locations of example structure

mulations for such a identification problem are highly complicated involving too many parameters to be identified. As a result, solutions are usually based on smooth spectral curves obtained by signal processing techniques. Among them, the random decrement technique (Cole, 1973; Ibrahim, 1977; Yang et al., 1984; Tasi and Yang, 1988) appears ideal for near-stationary wind excitations as in the case of the present field testing.

Upon identifying these model parameters, further identification of mass, stiffness, and damping matrices was carried out in order to develop the analytical model of the structural system which is important not only for design of the active mass damper controller of the structure but for evaluation of its dynamic response to wind and seismic excitations. The mass matrix in general can be estimated from the design drawings and the damping matrix constructed utilizing modal damping ratios, together with the mass and stiffness matrices. A technical challenge thus exists in the determination of the dynamic model arising from the difficulties in estimating the stiffness matrix from the ambient vibration measurements at a limited number of sensor locations.

In this study, an initial "baseline" stiffness matrix was estimated from the available design drawings of the structure. Starting from the baseline, an iterative optimization procedure was used to successively update the stiffness matrix by correlating it with the modal parameters including the natural frequencies and mode shapes measured from the field test. Available techniques for this purpose (e.g., Yun and Shinozuka, 1980; Stetson and Harrison, 1981; Tokamani and Ahmadi, 1988; Ting and Ojalvo, 1989; Hosiya and Sutoh, 1993; Okuma, 1993) usually require a sensitivity matrix of modal parameters to each element stiffness of the structural system. In this study, the second-order inverse modal perturbation theory was employed to obtain the sensitivity matrix of the natural frequencies and mode shapes to the flexural stiffness of a finite element. The first-order inverse modal perturbation technique was developed by Stetson and his colleagues for redesign problem (Stetson and Palma, 1975; Stetson and Harrison, 1981), and extended to the second-order theory by Yun and Hong (1992) for the purpose of identifying the extent and location of structural damage. However, the problem of how to utilize the ambient vibration measurements at limited sensor locations in order to develop a stiffness matrix involving a large number of elements for a large-scale structure has not been adequately addressed in the literature, thus posing a major challenge to the present study.

2 Ambient Vibration Measurement

The monitoring system consisted of five optical fiber accelerometers, a single conditioner, and a tape recorder. The acceler-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

ometers were installed on the tower at two sets of different locations (with some in common), as shown in Fig. 1, to measure the ambient vibration in two perpendicular horizontal directions (X and Y). The sensors at the first set of locations (Set 1) were concentrated on the upper part of the structure since this part is more flexible resulting in more vibration than the lower part, while those at the second set of locations (Set 2) were distributed along the height of the whole structure. The accelerometers were placed as close as possible to the center of the cross section of the tower in order to minimize the response component due to torsional vibration.

There are some substantial advantages in the proposed optical fiber sensor in relation to its installation on large-scale civil structures. In this study, the TV tower has strong electromagnetic fields and electric cables of a conventional sensor must be heavily shielded in order to avoid the electromagnetic interference. This results in very heavy cables and as a consequence, installation of the conventional sensor over such a long vertical distance, as required for this TV tower, becomes highly labor-intensive and cumbersome. However, in the proposed optical fiber sensor, both cables for power supply and for signal transmission are optical fibers weighing extremely light and requiring no shielding, thus eliminating these major practical difficulties associated with installation. Indeed, these advantages contributed significantly to the ease of installation of these five sensors over the long vertical distances in the present field experiment.

Acceleration records were obtained simultaneously at five sensor locations in one direction each time, with a sampling time interval of 0.03125 sec and total recording time of approximately 600 sec. In total, four sets of simultaneously measured acceleration time histories, $1X$, $1Y$, $2X$, and $2Y$, were recorded. Figure 3 shows two sets of them (up to 200 sec) measured at the first and second sets of locations in the X-direction.

3 Estimation of Modal Parameters

When a large-scale structure is subjected to such natural and man-made excitations as wind, traffic, or water waves or their combinations, it is extremely difficult to measure the input dynamic forces acting on the structure. As a result, only the structural response is usually measured, and frequency domain approaches (e.g., Ewins, 1984; Hong and Yun, 1992) rather than time domain approaches (e.g., Yun and Shinozuka, 1980; Shinozuka and Ghanem, 1995; Yun et al., 1997) are more appropriate for identifying modal parameters. In the present study, frequency domain identification techniques were used to extract natural frequencies, damping ratios, and mode shapes of the Nanjing TV Tower from the four sets of measured records.

3.1 Randomdec Signals. The response signal $\mathbf{y}(t)$ measured on the structure can be decomposed into free-vibration component $\mathbf{x}_0(t)$ and forced-vibration component $\mathbf{x}_f(t)$:

$$\mathbf{y}(t) = \langle y_1(t), y_2(t), \ldots, y_{N_{ob}}(t)\rangle^T = \mathbf{x}_0(t) + \mathbf{x}_f(\mathbf{f}(t); t) \quad (1)$$

where N_{ob} is the number of measurement points and $\mathbf{f}(t)$ represents wind load vector which is assumed to be a random process with a zero mean. Traffic and other load can be ignored compared to the wind load because the tower structure is tall. The free-vibration component contains an impulse and/or step responses, while the forced-vibration component represents response to the random wind load. Even in the case that the complete information for the random force is not available, the modal parameters such as natural frequencies and damping ratios can be estimated accurately using signals of the free-response component. The free-vibration component can easily be

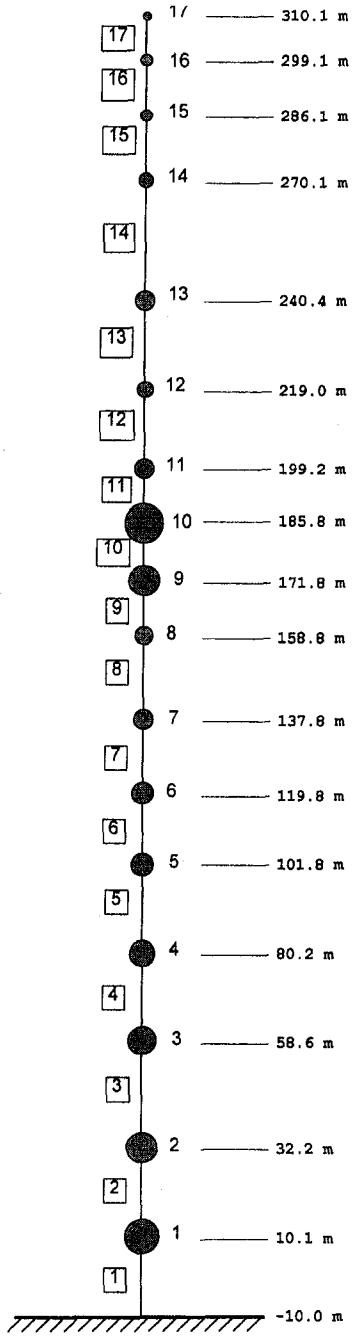


Fig. 2 Finite element model of example structure

generated by a special averaging procedure called the random decrement technique as long as measurements are carried out simultaneously at sensor locations (Ibrahim, 1977).

The randomdec signal $\mathbf{y}_r(\tau)$ of the measured signal is defined by Eq. (2a) and similarly corresponding force vector $\mathbf{f}_r(\tau)$ is defined by Eq. (2b)

$$\mathbf{y}_r(\tau) \equiv \frac{1}{N_r}\sum_{m=1}^{N_r} \mathbf{y}(t_m + \tau) \quad (2a)$$

$$\mathbf{f}_r(\tau) \equiv \frac{1}{N_r}\sum_{m=1}^{N_r} \mathbf{f}(t_m + \tau) \quad (2b)$$

where t_m is determined from original data as all the time instants satisfying initial conditions specified for the randomdec signal, e.g., $y_1(t_m) = y_0$ and $\dot{y}_1(t_m) \geq 0$; and N_r is the number of time

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

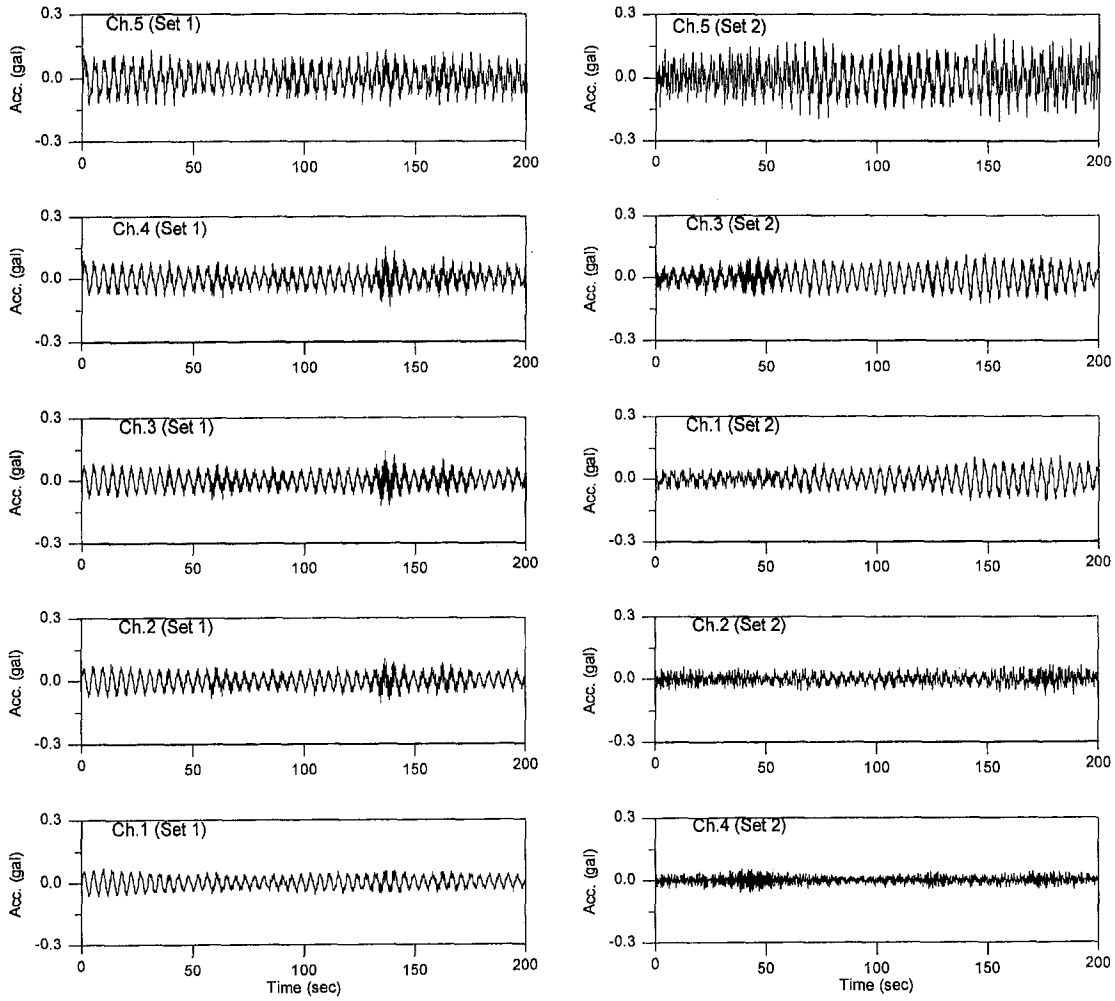


Fig. 3 Time histories of horizontal (X-direction) acceleration records

points with respect to which the average is taken. If the number of data sets N_r is sufficiently large and $\mathbf{f}(t)$ is a weakly stationary random process, the randomdec force vector $\mathbf{f}_r(\tau)$ will be averaged out owing to the zero-mean property of $\mathbf{f}(t)$. Thus the randomdec signal is expected to be free from the forced-vibration component under those conditions.

$$\lim_{N_r \to \infty} \mathbf{f}_r(\tau) = \mathbf{0} \tag{3}$$

In this study, t_m's are taken as those of zero crossing time with positive slope on the acceleration time history recorded at the small observation hall (Ch. 5 sensor location). The total length of the randomdec signal is 4096 points. Based on these conditions, four sets of randomdec signals were obtained from the original data sets of $1X$, $2X$, $1Y$, and $2Y$. Figure 4 shows the randomdec time histories, and Fig. 5 displays Fourier amplitudes of typical randomdec signals compared with those of the measured signals which have been averaged using the conventional approach with data length of 8192 points and 30 percent overlap. Apparently, the time histories of the randomdec signals look like free-vibration responses, and their FFT plots are much smoother than those of the original. Now, various modal parameter estimation techniques using free-vibration measurements can be employed. It is more accurate to identify the natural frequencies and damping ratios from the randomdec signals, than from the original signals.

3.2 Natural Frequencies. The natural frequencies of the structural system were estimated by reading directly the frequencies at the peaks of the FFT plots of the measured acceleration averaged using the conventional approach with the total data length of 8192 points and 30 percent overlap. Because the observed frequencies might result from the dominant frequencies of the input excitations as well as structural resonance frequencies, the direct estimates were verified by checking coherence functions between two response records measured at different locations. Meanwhile, the natural frequencies were also estimated by applying the circle-fit method on the randomdec signals taking advantage of circle characteristics of frequency response function around natural frequencies in the complex plane (Ewins, 1984). The first seven natural frequencies of the structure were estimated averaging the results from the two aforementioned methods. The results are shown in Table 1 together with the estimates from the previous measurement using conventional servo-type accelerometers performed by Tang et al. (1995). It is demonstrated that the estimated frequencies based on this measurement using optical fiber accelerometers are very close to those using the conventional accelerometers. The first five identified natural frequencies were 0.237 Hz, 0.727 Hz, 1.271 Hz, 1.588 Hz, and 2.721 Hz, and they are judged to be accurate since the deviation of the estimates obtained by different methods and from different data sets was very small.

The small observation hall (Ch. 5 sensor location) is an important location whose vibration is the major control target

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

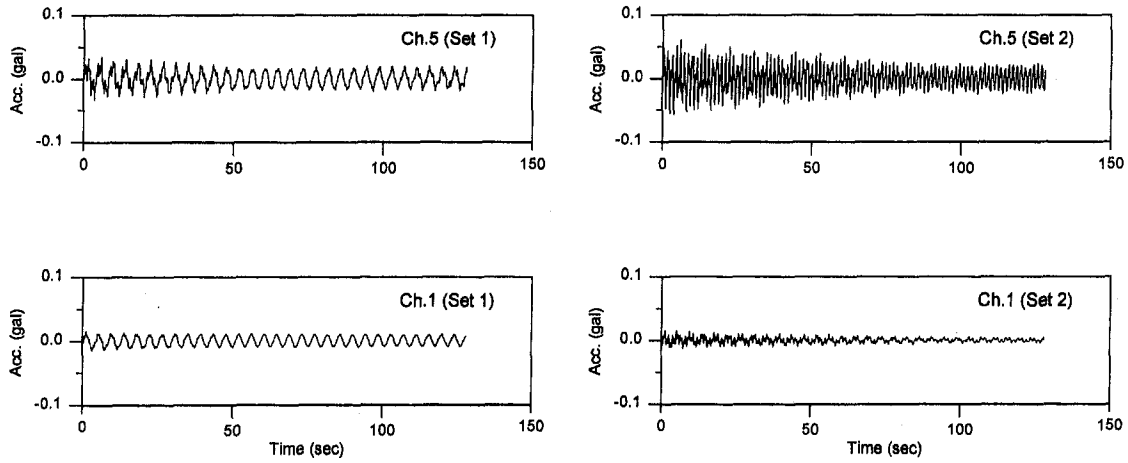


Fig. 4 Typical time histories of randomdec acceleration signals

of the active mass damper system to be installed. As seen in Fig. 6, the dominant modes associated with the response of the small observation hall are the first, second, and fourth modes. Therefore, the analytical model of the structural system to be developed from this study should well match the measured modal properties especially at these dominant modes.

A mode with a frequency around 2.5 Hz was identified from the acceleration recorded at the locations below the large observation hall. It appears to be a torsional mode since the sensors at those locations were placed at a distance from the vertical axis of the structure. In this study, this mode was neglected in development of a two-dimensional dynamic model for the structural system.

3.3 Damping Ratios. In general, the damping values estimated from the wind-induced structural response contains aerodynamic effects. For this reason, the direct estimation of damping ratios using the original acceleration records may result in an overestimation, and the randomdec signals free from the wind loads are ideal to use. On the other hand, the Fourier transformed response functions obtained from the randomdec signals are generally much smoother than those using the original records, as shown in Fig. 5. Therefore, damping

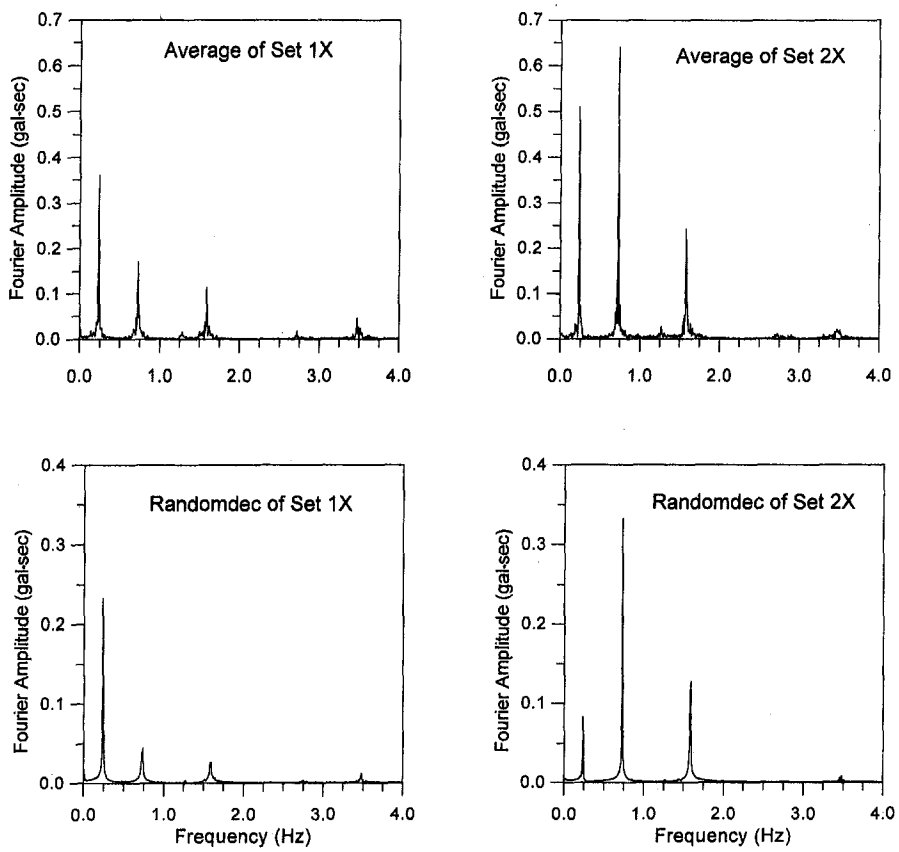


Fig. 5 Fourier amplitudes of averaged and randomdec accelerations at small hall (Ch. 5)

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 1 Natural frequencies of Nanjing TV Tower estimated using measured vibration data (unit: HZ)

Mode Number	This Study						Tang (1995)
	Signal Used	1X	1Y	2X	2Y	Average	
1	Original	0.236	0.238	0.234	0.238	0.237	0.234
	Randomdec	0.236	0.237	0.235	0.238		
2	Original	0.729	0.727	0.727	0.729	0.727	0.730
	Randomdec	0.727	0.727	0.727	0.726		
3	Original	1.272	1.268	1.268	1.277	1.271	1.266
	Randomdec	1.271	1.270	1.266	1.277		
4	Original	1.588	1.588	1.582	1.596	1.588	1.596
	Randomdec	1.586	1.589	1.584	1.592		
5	Original	2.709	2.712	2.723	2.719	2.721	-
	Randomdec	2.727	2.722	2.719	2.738		
6	Original	3.471	3.481	3.453	3.467	3.468	-
	Randomdec	--	--	--	--		
7	Original	5.432	5.422	5.434	5.424	5.428	-
	Randomdec	--	--	--	--		

Original: It is read from FFT plot of averaged acceleration (number of data points = 8192)
Randomdec: It is estimated by applying the circle fitting method on randomdec signals.

ratios, which are very sensitive to the smoothness of the Fourier transformed response functions near the resonant frequencies, can be estimated more accurately utilizing the randomdec signals. In this study, the damping ratios for the first four modes of the structural system were estimated from the randomdec signals by the half-power bandwidth method. The estimated values are summarized in Table 2, along with other experimental results from Tang et al. (1995). The estimated

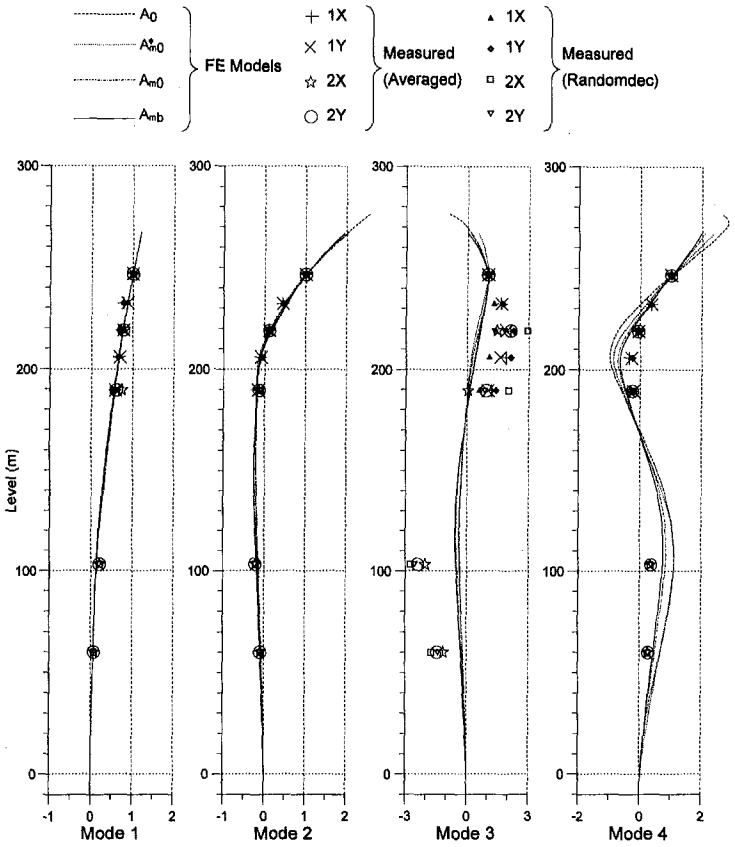


Fig. 6 Calculated and measured mode shapes

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 2 Estimated damping ratios of the Nanjing TV Tower utilizing randomdec signals

Used Record		Estimated Damping Ratio (%)			
		mode 1	mode 2	mode 3	mode 4
1X	5	1.298	1.585	--	--
	4	1.302	1.591	0.544	--
	3	1.326	0.746	0.584	0.680
	2	1.330	1.420	0.596	--
	1	1.355	1.661	0.596	--
1Y	5	1.700	0.396	--	--
	4	1.715	0.396	--	--
	3	1.725	0.397	--	--
	2	1.731	0.394	--	--
	1	1.787	0.390	--	--
2X	5	1.273	0.385	0.748	0.431
	3	1.285	0.389	--	0.433
	1	1.308	0.389	--	0.410
	2	1.268	0.387	--	0.434
	4	1.396	0.437	--	0.392
2Y	5	2.036	0.504	0.761	0.607
	3	2.518	0.502	0.623	0.534
	1	2.178	0.512	0.624	0.558
	2	2.490	0.500	0.601	0.515
	4	2.231	0.473	--	0.534
This Study (Average)		1.663	0.677	0.631	0.503
Tang		1.480	1.348	0.887	0.760

damping rations were 1.66 percent for the first mode and less than one percent for the other modes, which are comparable with typical values observed in similar structures (Bao and Ko, 1991).

3.4 Mode Shapes. If the estimates of natural frequencies are accurate, corresponding mode shapes at the measured points can be easily obtained using power spectral density functions, as follows (Bao and Ko, 1991; Gavin et al., 1992):

$$\hat{\Phi}_{pk}^{(I)} = \frac{S_{y_p y_q}(f_k)}{S_{y_q y_q}(f_k)} \tag{4a}$$

$$\hat{\Phi}_{pk}^{(II)} = \frac{S_{y_p y_p}(f_k)}{S_{y_p y_q}(f_k)} \tag{4b}$$

$$\hat{\Phi}_{pk}^{(III)} = \text{sign}\,(\hat{\Phi}_{pk}^{(I)})\sqrt{\frac{S_{y_p y_p}(f_k)}{S_{y_q y_q}(f_k)}} \tag{4c}$$

in which $\hat{\Phi}_{pk}^{(I)}$, $\hat{\Phi}_{pk}^{(II)}$, and $\hat{\Phi}_{pk}^{(III)}$ are estimates of the kth mode shape at point p normalized to the amplitude at reference point q; $S_{y_p y_p}$ and $S_{y_q y_q}$ are respectively auto and cross-spectral density functions; f_k is the kth natural frequency; and sign $(\cdot)$ is a sign function which equals to 1 or -1. Assuming the measured response $y_p(t)$ at point p can be separated into uncorrelated pure response $\hat{y}_p(t)$ and noise $n_p(t)$, Eq. (4) can be rewritten as

$$\hat{\Phi}_{pk}^{(I)} = \hat{\Phi}_{pk}\frac{1}{1 + S_{n_q n_q}(f_k)/S_{\hat{y}_q \hat{y}_q}(f_k)} \tag{5a}$$

$$\hat{\Phi}_{pk}^{(II)} = \hat{\Phi}_{pk}\left(1 + \frac{S_{n_p n_p}(f_k)}{S_{\hat{y}_p \hat{y}_p}(f_k)}\right) \tag{5b}$$

$$\hat{\Phi}_{pk}^{(III)} = \hat{\Phi}_{pk}\sqrt{\frac{1 + S_{n_p n_p}(f_k)/S_{\hat{y}_p \hat{y}_p}(f_k)}{1 + S_{n_q n_q}(f_k)/S_{\hat{y}_q \hat{y}_q}(f_k)}} \tag{5c}$$

where $\hat{\Phi}_{pk} = S_{\hat{y}_p \hat{y}_q}(f_k)/S_{\hat{y}_q \hat{y}_q}(f_k)$. Equation (5$c$) shows that the mode shape can be identified accurately regardless of noise level, if the noise levels at all the measured locations are proportional to the corresponding intensity levels of records. Because $|\hat{\Phi}_{pk}^{(I)}| \le |\hat{\Phi}_{pk}| \le |\hat{\Phi}_{pk}^{(II)}|$ and $|\hat{\Phi}_{pk}^{(I)}| \le |\hat{\Phi}_{pk}^{(III)}| \le |\hat{\Phi}_{pk}^{(II)}|$, the effects of noise on the mode shape estimation can also be examined by evaluating the difference between the lower bound $\hat{\Phi}_{pk}^{(I)}$ and the upper bound $\hat{\Phi}_{pk}^{(II)}$ of the mode shape estimates.

The first four mode shapes were identified from the power spectral density functions of the randomdec signals and acceleration records averaged using the conventional approach with 4096 points segments and 30 percent overlap. The Ch. 5 location (small observation hall) was chosen as the reference point q. The estimated results by Eq. (5c) are shown in Fig. 6 (compared with those calculated by the identified mass and stiffness matrices to be described in the next section). Despite of the fact that the input loads on the structural system were not possible to measure, the estimated mode shapes appear reasonable. In addition, the differences between the lower bound and the upper

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 3 Nodal coordinates, lumped masses, and stiffness parameters of baseline finite element model A_0

Node Number	Lumped Mass (Mg)	Element Number	Cross Sectional Area (m^2)	Moment of Inertia (m^4)
0	--	1	53.22	3695.5
1	3992.90			
2	3186.74	2	47.95	2707.4
3	2820.10	3	42.69	1938.7
4	2319.79	4	38.89	1457.4
5	1917.95	5	32.77	1114.2
6	1624.51	6	30.79	883.0
7	1628.07	7	29.57	792.6
8	1322.34	8	27.90	673.0
9	3395.32	9	25.31	568.5
10	5678.60	10	47.26	1134.7
11	1325.79	11	47.26	465.9
12	359.10	12	7.386	27.78
13	1081.47	13	7.386	27.78
14	165.13	14	4.96	8.08
15	18.70	15	0.2124	1.10930e-01
16	12.00	16	0.1075	2.07160e-02
17	4.00	17	0.0584	5.19079e-03

bound of the mode shape estimates were less than one percent in the present ambient vibration records. Therefore, the noise effects can be neglected in this analysis.

4 Finite Element Model Correlated With Measurements

4.1 Baseline Finite Element Model. A two-dimensional baseline finite element model for the tower structural system, namely A_0, is constructed based on design drawings of the structure and the finite element model developed by Cheng (1992). The model consists of 17 beam elements and lumped masses. The elevation of the finite element model is shown in Fig. 2. The element properties of the baseline model are shown in Table 3. In this model, tapered beams were replaced by equivalent piece-wise uniform prismatic beams for the analysis purpose. The cross-sectional area and moment of inertia were taken at the midsection of each element. Lumped masses were assumed to be the same as Cheng's model. The Young's moduli for the finite elements used in analysis are: $E_c = 3.237 \times 10^7$ KPa for concrete members, and $E_s = 2.060 \times 10^8$ KPa for steel members.

4.2 Method of Updating Baseline Model.

Definition of Updated Model. Updated mass and stiffness matrices can be written in the form

$$\hat{\mathbf{M}} = \mathbf{M} + \Delta\mathbf{M} \tag{6}$$

$$\hat{\mathbf{K}} = \mathbf{K} + \Delta\mathbf{K} \tag{7}$$

where $\mathbf{M}$ and $\mathbf{K}$ are the baseline mass and stiffness matrices; $\Delta\mathbf{M}$ and $\Delta\mathbf{K}$ are correction mass and stiffness matrices. It is assumed that the updated mass matrix is the same as the baseline mass matrix, and the correction stiffness matrix is represented by a linear combination of the element stiffness matrices in the baseline model

$$\Delta\mathbf{M} = \mathbf{0} \tag{8}$$

$$\Delta\mathbf{K} = \sum_{e=1}^{N_e} a_e\mathbf{K}_e \tag{9}$$

in which a_e is correction coefficient of flexural stiffness EI for the eth element; N_e denotes the number of finite elements; and $\mathbf{K}_e$ is the eth element stiffness matrix of the baseline.

Inverse Modal Perturbation Theory. The sensitivities of the natural frequencies and mode shapes to the stiffness correction coefficients were obtained from the second-order inverse modal perturbation theory (Yun and Hong, 1992), as follows:

$$\hat{\lambda}_k = \lambda_k + \frac{\mathbf{s}_{kk}\mathbf{a}}{\mu_k(1 + D_{kk})} \tag{10}$$

$$\{\hat{\Psi}_k\} = \mathbf{T}_k\mathbf{S}_k\mathbf{a} \tag{11}$$

where

$$S_k(j, e) = \{\Psi_j\}^T\mathbf{K}_e\{\hat{\Psi}_k\} \tag{12}$$

$$T_k(i, j) = \frac{\Psi_{ij}}{\mu_j(\hat{\lambda}_k - \lambda_j)} \tag{13}$$

$$1 + D_{kk} = \sqrt{\hat{\mu}_k - \sum_{j=1, j\neq k}^{N_{\text{mode}}} \mu_j D_{kj}^2} \tag{14}$$

$$D_{kj} = \frac{\mathbf{s}_{kj}\mathbf{a}}{\mu_j(\hat{\lambda}_k - \lambda_j)} \tag{15}$$

$$\mu_k = \{\Psi_k\}^T\mathbf{M}\{\Psi_k\}, \quad \hat{\mu}_k = \{\hat{\Psi}_k\}^T\mathbf{M}\{\hat{\Psi}_k\}. \tag{16}$$

λ_k and $\hat{\lambda}_k$ are, respectively, the square of the kth natural frequencies from baseline and updated models; N_{mode} is the number of modes used; $\{\Psi_k\}$ and $\{\hat{\Psi}_k\}$ are, respectively, the kth mode shape vectors of baseline and updated models; $\mathbf{s}_{kj}$ is the jth *row vector* of matrix $\mathbf{S}_k$; and $\mathbf{a}$ is the stiffness correction coefficient vector.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Quadratic Optimization. From Eq. (10), an expected difference in the normalized strain energy of the kth mode between an updated model and the measurement can be written as

$$e_k(\mathbf{a}) = \frac{1}{\tilde{\lambda}_k}\{\mathbf{s}_{kk}\mathbf{a} - \mu_k(1 + D_{kk})\Delta\lambda_k\} \tag{17}$$

where $\Delta\lambda_k$ is a difference in the square of kth natural frequency between the baseline and measurement ($=\tilde{\lambda}_k - \lambda_k$); and $\tilde{\lambda}_k$ denotes the square of measured natural frequency of the kth mode. Now, the following objective function is established and minimized to obtain the optimal value of $\mathbf{a}$,

$$\min_{\mathbf{a}} J = \sum_{k=1}^{N_k}\{w_k e_k(\mathbf{a})\}^2, \tag{18}$$

in which N_k is the number of measured natural frequencies, and w_k is a weighting factor for the kth mode that may be assumed as the amplitude of Fourier transformed response function at the kth natural frequency.

Measured mode shapes are considered in constraining equations with proper error bounds, rather than being involved in the objective function, since their accuracy is worse than that of natural frequencies. Moreover, since it is impossible to measure the ambient vibration at all nodal points in practice for a large-scale structural system, it is very difficult to define an objective function in conjunction with the natural frequencies systematically. Thus the relative values of mode shapes were incorporated into constraining equations in this study. Eventually, the constraining equations related to the kth mode shape is expressed as

$$\left|\frac{\hat{\Psi}_{pk}(\mathbf{a})}{\hat{\Phi}_{qk}} - \frac{\hat{\Psi}_{qk}(\mathbf{a})}{\hat{\Phi}_{qk}}\right| \le \epsilon_k$$

$$(p = 1, 2, \ldots, N_{ob};\ \text{but } p \neq q) \tag{19}$$

where $\hat{\Phi}_{pk}$ is the kth measured mode shape at point p; and ϵ_k is an error bound for the kth mode shape. In addition to the constraints, upper and lower bounds for the stiffness correction coefficient vector $\mathbf{a}$, which can be determined by engineering judgment, are added to the constraints to reflect real conditions of the structure. Because the measurement points did not coincide with the nodal points of the finite element model in this study, the constraining equations were applied after representing the mode shape data by the corresponding nodal responses and shape functions of the beam element.

In summary, the system identification problem to establish an appropriate dynamic model correlating with measured modal parameters can be represented as an optimization problem to determine the stiffness correction coefficients under constraining equations, as follows:

Find out vector $\mathbf{a}$ which minimizes the following objective function:

$$J = \tfrac{1}{2}\mathbf{a}^T\mathbf{H}\mathbf{a} + \mathbf{f}^T\mathbf{a} \tag{20}$$

subject to conditions as

$$\mathbf{G}\mathbf{a} \le \mathbf{b} \quad \text{and} \quad \mathbf{a}_{\min} \le \mathbf{a} \le \mathbf{a}_{\max} \tag{21}$$

where $\mathbf{H}$ and $\mathbf{f}$ are derived from Eqs. (17) and (18), and $\mathbf{G}$ and $\mathbf{b}$ are from Eqs. (11) and (19).

Because the matrices $\mathbf{H}$, $\mathbf{f}$, and $\mathbf{G}$ are dependent on parameter $\mathbf{a}$, the solution should be found by iteration.

4.3 Updated Finite Element Model of Nanjing TV Tower.

Better Baseline Finite Element Model. It is well known that the convergence and accuracy of an optimization solution depend on the initial baseline estimates. In this study, an analytical model, namely A_b, was developed as a better baseline (compared with model A_0 mentioned in Section 4.1). The normalized element modal strain energy diagram ($\mathbf{s}_{kk}$/max ($\mathbf{s}_{kk}$)), which represents the contribution of an element stiffness to the kth natural frequency, was produced in order to get insight on how to modify the element stiffness to reduce the difference in natural frequencies between the calculated and the measured. The diagram for the first four modes of the initial baseline model A_0 is shown in Fig. 7. From the figure, it can be seen that stiffness correction coefficients for the lower part of the structure, say elements below level of 100 m, shall be positive to increase natural frequency for the first mode of updated model, while those for the upper part of the structure for the second and third modes. Therefore, the stiffness correction coefficient vector can be determined by trial and error referring to the diagram and comparing to the recalculated and measured natural frequencies. Natural frequencies by this better baseline are listed in Table 6, which show a much better agreement with the measured ones than those by the baseline.

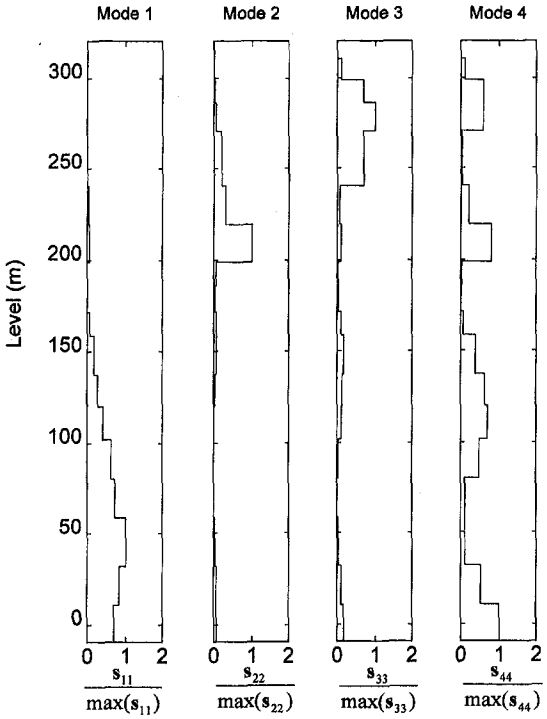


Fig. 7 Normalized element modal strain energy diagram

Updated Model. Three updated models A^*_{m0}, A_{m0}, and A_{mb} were obtained starting from baseline models A_0 and A_b. The model A^*_{m0} has been updated by using the measured natural frequencies only, while models A_{m0}, and A_{mb} were identified utilizing the measured mode shape data as well as the natural frequencies. The brief descriptions regarding the models are given in Table 4. Seven natural frequencies and four mode shapes identified on the basis of the measurement data were used in updating the finite element models.

In the optimization process, the degree of convergence and accuracy of solutions are fairly sensitive to the upper and lower bounds of the stiffness correction factor ($\mathbf{a}_{\min}$ and $\mathbf{a}_{\max}$), the error bounds for the mode shape data (ϵ_k) and weighting factors (w_k) in the objective function. The various parameters used in the optimization process are given in Table 5. Convergence of the natural frequencies is demonstrated in Fig. 8. The converged stiffness correction coefficients are summarized in Table 5 and Fig. 9. As shown in Fig. 9, the stiffness scale factors of different updated models are very similar to each other, independent of the baseline model and information used. In particular, models

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 4 Description on baseline and updated finite element models

Model Name	Base Model	Construction Method	Information Used
A_0	Cheng's 3D model		Design drawings, 3D model
A^*_{mo}	A_0	IMP*	7 measured natural frequencies
A_{mo}	A_0	IMP*	7 measured natural frequencies and 4 mode shapes
A_b	A_0	trial and error	Element modal strain energy diagram (Fig. 7)
A_{mb}	A_b	IMP*	7 measured natural frequencies and 4 mode shapes

Note: * Inverse Modal Perturbation

A_{m0} and A_{mb} are nearly identical despite of their quite different baseline models. The updated stiffness models appear stiffer than the baseline model, especially in the location of small and large observation hall structures and lower part of the main three-leg structure. These may result from difference of concrete stiffness between the baseline and the real structure, and additional stiffening effects by prestressing forces, reinforced steel, and nonstructural components.

The natural frequencies and mode shapes calculated from the updated models are shown in Table 6 and Fig. 6, respectively. The first five natural frequencies show excellent agreement with the measured results. The corresponding mode shapes also match very well with the measured ones except the third mode which is, however, not a critical mode for the response at the small observation hall.

Although a better baseline model, such as A_b in this study, may improve the convergence, a conventional model made from the design drawings of the structure, such as A_0 in this study, appears to be a reasonable baseline for establishing the updated model, if sufficient information for the real structure can be incorporated in the optimization process. In general, the updated models A_{m0} and A_{mb} constructed in conjunction with both the

Table 5 Stiffness correction coefficients of better baseline and updated finite element models

Element Number	Stiffness Correction Coefficient (a_e)*			
	A_b	A^*_{m0}	A_{m0}	A_{mb}
1	0.1	-0.300	-0.262	-0.300
2	2.0	3.000	3.000	2.349
3	1.5	3.000	3.000	3.000
4	1.5	3.000	3.000	3.000
5	-0.2	0.047	0.757	0.779
6	-0.3	-0.400	-0.400	-0.400
7	-0.3	-0.400	-0.400	-0.400
8	-0.2	3.000	-0.400	-0.400
9	0.1	3.000	3.000	3.000
10	1.5	4.000	4.000	4.000
11	1.1	3.000	-0.400	-0.400
12	0.2	0.016	0.116	0.174
13	0.5	4.000	4.000	4.000
14	0.4	0.162	0.191	0.154
15	0.0	0.085	0.150	0.150
16	0.0	-0.150	-0.088	-0.031
17	0.0	-0.150	-0.150	-0.150

Note: * The updated element stiffness matrix shall be $(1+a_e)\,K_e$, in which K_e is the e-th element stiffness matrix of baseline model. The coefficients used for optimization are as follows: w_1=w_2=100, w_3=w_4=4, w_5=1, w_6=w_7=0.001; $-0.3 \le a_1 \le 3.0$, $-0.3 \le a_{10} \le 4.0$, $-0.4 \le a_{13} \le 4.0$, $-0.3 \le a_e \le 3.0$ for other concrete member, and $-0.15 \le a_e \le 0.15$ for steel members; $\varepsilon_1 = 0.0075$ and $\varepsilon_k = 0.01$ for the other modes.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

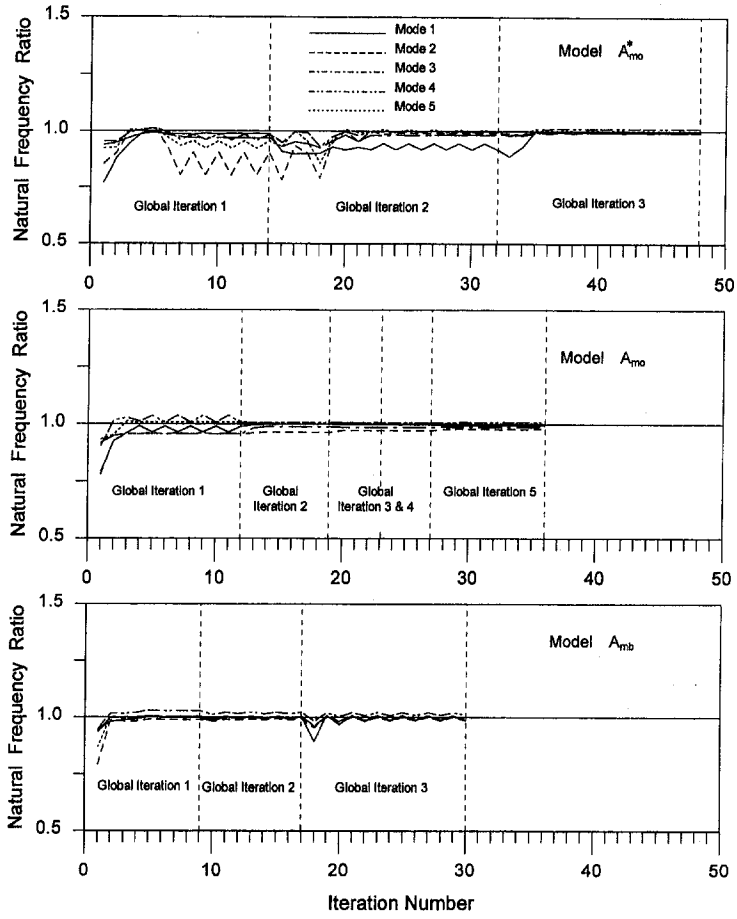


Fig. 8 Convergence of natural frequencies of updated finite element models

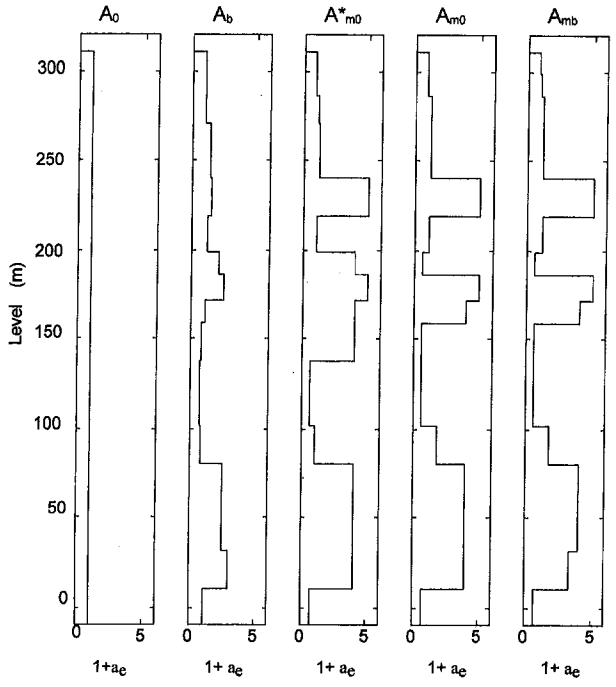


Fig. 9 Element stiffness scale factor (1 + a_e) of baseline and updated finite element models

measured natural frequencies and mode shapes show better agreement with the measured mode shapes than the updated model A^*_{m0} using natural frequency information only.

5 Conclusion

Ambient vibration tests were performed on a large-scale structural system using the newly developed optical fiber accelerometers, in order to establish an analytical model of the structure for designing an active mass damper system to control the structural vibration under wind excitations. Modal parameters including natural frequencies, damping ratios, and mode shapes were estimated from the measured acceleration records by using various system identification techniques, among which the random decrement technique was found to be particularly effective in identifying the damping ratios due to the removal of random wind effects from the measured signals. More importantly, several finite element stiffness matrices involving a large number of elements were successfully estimated taking advantage of the very limited ambient vibration information obtained from a limited number of sensors. The second-order inverse modal perturbation theory was effective in establishing the sensitivity relationship between the element stiffness and the modal parameters, which was convenient used in the quadratic optimization to derive the system stiffness matrix producing the minimum difference between the calculated and measured modal strain energy. In the process of this optimization, it was found reasonable to consider the mode shape information as a constraint, rather than embedding it together with the natural frequency information in the objective function. The finite element models thus identified represent the actual structural system very well, as their calculated dynamic characteristics satisfactorily

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 6 Comparison of natural frequencies calculated using finite element models with measured results (unit: HZ)

Mode Number	Calculated using Finite Element Model					Measured
	A_o	A_b	A^*_{mo}	A_{mo}	A_{mb}	
1	0.199	0.228	0.235	0.236	0.233	0.237
2	0.628	0.716	0.727	0.711	0.716	0.727
3	1.228	1.285	1.264	1.257	1.256	1.271
4	1.625	1.681	1.605	1.600	1.602	1.588
5	2.489	2.798	2.726	2.730	2.721	2.721
6	4.153	4.385	4.622	4.605	4.611	3.468
7	4.705	4.822	5.461	4.868	4.864	5.428

matched the observed ones from the ambient vibration test actually performed on the large-scale structural system.

Acknowledgments

The authors wish to thank Prof. C.-B. Yun and Dr. H.-J. Lee at the Korea Advanced Institute of Science and Technology for their valuable comments on the system identification results. This work was supported by the National Science Foundation under the CAREER Award and monitored by Dr. N. Sabadell, Program Director. The second author thanks, for the partial financial support, the Korea Science and Engineering Foundation on his postdoctoral research related to this work.

References

Ansari, F., ed., 1993, *Applications of Fiber Optic Sensors in Engineering Mechanics,* ASCE, Reston, VA.

Bao, Z. W., and Ko, J. M., 1991, "Determination of modal parameters of tall buildings with ambient vibration measurements," *The International Journal of Analytical and Experimental Modal Analysis,* Vol. 6, No. 1, pp. 57–68.

Cheng, W. R., 1992, "The analysis of dynamic characteristics of Nanjing TV Tower," Technical Report of Southeast University, in Chinese.

Cole, H. J., 1973, "On-line failure detection and damping measurement of aerospace structures by random decrement signatures," NASA Cr-2205.

Ding, Dajun, and Ren, Zuenhua, 1994, "Design testing and analysis of the Nanjing TV Tower," *Concrete International,* Nov., pp. 42–44.

Ewins, D. J., 1984, *Modal Testing: Theory and Practice,* John Wiley and Sons, New York.

Feng, M. Q., 1996, "An experimental study on electro-optical displacement sensors," *Nondestructive Test and Evaluation,* Vol. 13, pp. 5–14.

Gavin, H., Suan, S., Grossman, J., Pekelis, E., and Jacob, K., 1992, "Low-level dynamic characteristics of four tall flat-plate buildings in New York City," NCEER Technical Report, NCEER-92-0034.

Hong, K.-S., and Yun, C.-B., 1992, "Improved Method for Frequency Domain Identification of Structures," *Engineering Structures,* Vol. 15, No. 3, pp. 179–188.

Hoshiya, M., and Sutoh, A., 1993, "Kalman filter-finite element method in identification," *Journal of Engineering Mechanics, ASCE,* Vol. 119, No. 2, pp. 197–210.

Ibrahim, S. R., 1977, "Random decrement technique for modal identification of structures," *Journal of Spacecraft and Rocket, AIAA,* Vol. 14, No. 11, pp. 696–700.

Kersey, A. D., 1994, "Role of fiber optic sensors in intelligent infrastructure systems," *Proc. the First World Conference on Structural Control,* Los Angeles, Aug., Plenary-3-9.

Okuma, M., 1993, "Correction of finite element models using experimental modal data for vibration analysis," *Finite Elements in Analysis and Design,* Vol. 14, pp. 153–162.

Shinozuka, M., and Ghanem, R., 1995, "Structural system identification I: Theory," *Journal of Engineering Mechanics, ASCE,* Vol. 121, No. 2, pp. 255–264.

Stetson, K. A., and Harrison, I. R., 1981, "Redesign of structural vibration modes by finite-element inverse perturbation," ASME *Journal of Engineering for Power,* Vol. 103, pp. 319–324.

Stetson, K. A., and Palma, G. E., 1975, "Inversion of first-order perturbation theory and its application to structural design," *AIAA Journal,* Vol. 14, No. 4, pp. 454–460.

Tang, J. X., Li, L., and Wu, C. Y., 1995, "Report on field measurement of Nanjing TV Tower," Technical Report of Southeast University, in Chinese.

Ting, T., and Ojalvo, I. U., 1989, "Dynamic structural correction via nonlinear programming technique," *Finite Elements in Analysis and Design,* Vol. 5, pp. 247–256.

Torkamani, M. A. M., and Ahmadi, A. K., 1988, "Stiffness identification of a tall building during construction period using ambient tests," *Earthquake Engineering and Structural Dynamics,* Vol. 16, pp. 1177–1188.

Tsai, W. H., and Yang, J. C. S., 1988, "Non-destructive Evaluation of Composite Structures Using System Identification Technique," ASME *Journal of Engineering Materials and Technology,* Vol. 110, pp. 134–139.

Vries, M. D., Murphy, K. A., Poland, S., Arya, V., Claus, R. O., and Masri, S. F., 1994, "Optical fiber sensor for civil structure applications: a comparative evaluation," *Proc. the First World Conference on Structural Control,* Los Angeles, CA, Aug., WA3-23-29.

Wang, D., and Haldar, A., 1994, "Element-level system identification with unknown input," *Journal of Engineering Mechanics, ASCE,* Vol. 120, No. 1, pp. 159–176.

Yang, J. C. S., Chen, J., and Dagalakis, N. G., 1984, "Damage Detection in Offshore Structures by the Random Decrement Technique," ASME *Journal of Energy Resources Technology,* Vol. 106, pp. 38–42.

Yun, C.-B., and Hong, K.-S., 1992, "Damage assessment of structures by inverse modal perturbation method," *Proc. 4th East Asia-Pacific Conf. Structural Engineering and Construction,* Seoul, Korea.

Yun, C.-B., Lee, H.-J., and Lee, C.-G., 1997, "Sequential prediction error method for structural identification," *Journal of Engineering Mechanics, ASCE,* Vol. 123, No. 2, pp. 115–122.

Yun, C.-B., and Shinozuka, M., 1980, "Identification of Nonlinear Structural Dynamic Systems," *Journal of Structural Mechanics,* Vol. 1, pp. 187–203.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Almost Classically Damped Continuous Linear Systems

S. Natsiavas
Department of Mechanical Engineering,
Aristotle University of Thessaloniki,
54006 Thessaloniki, Greece

J. L. Beck
Division of Engineering
and Applied Science,
California Institute of Technology,
Pasadena, CA 91125

The dynamic response of a general class of continuous linear vibrating systems is analyzed which possess damping properties close to those resulting in classical (uncoupled) normal modes. First, conditions are given for the existence of classical modes of vibration in a continuous linear system, with special attention being paid to the boundary conditions. Regular perturbation expansions in terms of undamped modeshapes are then utilized for analyzing the eigenproblem as well as the vibration response of almost classically damped systems. The analysis is based on a proper splitting of the damping operators in both the field equations and the boundary conditions. The main advantage of this approach is that it allows application of standard modal analysis methodologies so that the problem is reduced to that of finding the frequencies and mode shapes of the corresponding undamped system. The approach is illustrated by two simple examples involving rod and beam vibrations.

Introduction

In many instances, the dynamic response of a vibrating system can be well predicted by a linear model. When damping effects are neglected, the equations of motion of these systems can be reduced to a set of uncoupled ordinary differential equations, each representing a single-degree-of-freedom linear oscillator. However, introduction of linear viscous damping allows this uncoupling only if the damping distribution bears a special relationship with the mass and stiffness distributions of the system (Caughey, 1960; Caughey and O'Kelly, 1965). If it does, the system is often said to be "classically damped" and to possess "classical normal modes." The terminology "proportional damping" is also used but this is easily confused with the very special case of Rayleigh damping (Rayleigh, 1945), where the damping distribution is a linear combination of the mass and stiffness distributions.

The existence of classical normal modes provides a simple physical interpretation of the behavior of a vibrating system as a linear superposition of characteristic modes of vibration of the system. It also leads to computational advantages when only a few modes of vibration contribute to the response of the system because of a limited bandwidth excitation. When the conditions for classical modes are not met exactly (e.g., Natsiavas, 1993), the system response may be obtained by one of several different methods: by direct numerical integration of the equations of motion; or by properly converting these equations to first-order state-space form (e.g., Foss, 1958; Meirovitch, 1967); or by approximating the damping in some way so that classical modes exist (Bellos and Inman, 1989).

The present work focuses on the response of linear continuous systems which do not possess classical normal modes, but the damping properties are close to those meeting the conditions leading to normal modes. These systems are called "almost classically damped" and are of practical significance. As a result, various aspects of their response have already been examined by others, but with emphasis placed on discrete systems. For example, Chung and Lee (1986) extended a perturbation methodology developed earlier by Meirovitch and Ryland (1979) in order to obtain the eigensolution of these discrete systems. Also, Cronin (1976) presented a perturbation analysis for the response of such systems under harmonic excitation, while Udwadia and Esfandiari (1990) and Udwadia and Kumar (1994) presented an iterative approach for general forcing functions. Other relevant work in discrete systems include the contributions of Knowles (1985), Nicholson (1987), Shahruz and Ma (1988), and Natsiavas and Beck (1994). Various aspects of damped continuous linear systems have also been examined (e.g., Caughey and O'Kelly, 1965; Pan, 1966; Plaut and Infante, 1972; Inman and Andry, 1982; Bergman and Nicholson, 1985; Yang, 1996). More references and information about damping effects can also be found in the books by Snowdon (1968) and Nashif et al. (1985).

The objectives of the present work are, first, to present conditions for the existence of classical modes for continuous linear vibrating systems and, second, to present a perturbation approach for the approximate solution of almost classically damped continuous linear systems under arbitrary forcing functions. For the first objective, the work of Caughey and O'Kelly (1965) is extended by being specific about the concept of "compatible" internal and boundary operators for both the stiffness and damping characteristics. For the second objective, an analysis is developed, based on a proper decomposition of the internal and boundary damping operators. The approach uses classical modal analysis so that the solution is obtained by utilizing only the real modes and frequencies of the corresponding undamped system. This is especially efficient in cases where the solution of a classically damped system has been obtained and the effect of small changes in the damping distribution is sought. Finally, the approach is illustrated by two simple examples.

Problem Statement and Definitions

Boundary Value Problem for Forced Vibrations. The forced vibration behavior of a linear elastic viscously damped continuous body, such as a stretched rod, a torsional shaft, a bending beam, a membrane or a plate, is commonly governed by an equation of motion of the form

$$m(\mathbf{r})\ddot{u} + L_1\dot{u} + L_2u = f(\mathbf{r}, t) \quad \forall \mathbf{r} \in S \tag{1}$$

where the body occupies $\bar{S} = S \cup \partial S$. Here S is a bounded domain in $\mathbb{R}^{n_s}$ ($n_s = 1$, 2 or 3) with boundary ∂S. Typically, $u(\mathbf{r}, t)$ is a (generalized) displacement at position $\mathbf{r}$ in the body

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, July 22, 1997; final revision, Apr. 3, 1998. Associate Technical Editor: N. C. Perkins.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

at time t, while $\dot{u}(\mathbf{r}, t)$ and $\ddot{u}(\mathbf{r}, t)$ denote the corresponding velocity and acceleration; $m(\mathbf{r})$ is the mass density of the body; $f(\mathbf{r}, t)$ is an external body force and L_1 and L_2 are time-invariant linear differential spatial operators of orders n_1 and n_2, respectively (typically $n_1 \leq n_2$), which model the internal damping forces and elastic restoring forces, respectively, within the body.

The appropriate boundary conditions for (1) are of the form

$$D_N^{(i-1)}u = O \quad \forall \mathbf{r} \in \partial S_i$$

$$\text{and} \quad B_i u + C_i \dot{u} = g_i(\mathbf{r}, t) \quad \forall \mathbf{r} \in \partial S_i' \tag{2}$$

for $i = 1, \ldots, n_b$ ($=1/2n_2$, typically), expressing the fact that for each i, on some parts of the boundary (∂S_i), a geometric condition will hold, while on the other parts ($\partial S_i'$), a generalized force condition will hold (so $\partial S_i \cup \partial S_i' = \partial S$, $\partial S_i \cap \partial S_i' = \phi$). $D_N^{(i)}$ denotes the derivative of order i normal to the boundary surface at $\mathbf{r} \in \partial S$, with $D_N^{(0)}u$ being simply u itself. Each B_i and C_i is a time-invariant linear differential spatial operator of order $(n_2 - i)$ and $(n_1 - i)$, respectively, which models the deformation-induced internal and external generalized restoring forces and damping forces, respectively, at the boundary. Each $g_i(\mathbf{r}, t)$ represents the resultant of any external generalized forces applied at the boundary of the body at time t.

An important additional assumption is that the B_i and C_i are *compatible* boundary operators for L_2 and L_1, respectively, in the sense that they are related through "integration by parts" (or a generalized Green's formula) by

$$\langle L_1\psi, \phi\rangle - \langle \psi, L_1\phi\rangle = \sum_{i=1}^{n_b} [(D_N^{(i-1)}\psi, C_i\phi) - (C_i\psi, D_N^{(i-1)}\phi)] \tag{3}$$

and

$$\langle L_2\psi, \phi\rangle - \langle \psi, L_2\phi\rangle = \sum_{i=1}^{n_b} [(D_N^{(i-1)}\psi, B_i\phi) - (B_i\psi, D_N^{(i-1)}\phi)] \tag{4}$$

for all $\psi, \phi \in V$, a linear space of suitably smooth functions on $\bar{S}$, such as $C^{n_3}(S) \cap C^{n_3-1}(\bar{S})$, where $n_3 = \max(n_1, n_2)$, and which satisfy the geometric boundary conditions in (2). Here $C^{n_3}(S)$ denotes the set of functions on S which are continuously differentiable up to order n_3. The body and boundary inner products in (3) and (4) are defined by

$$\langle \psi, \phi\rangle = \int_S \psi(\mathbf{r})\phi(\mathbf{r})d\mathbf{r} \tag{5}$$

and

$$(\psi, \phi) = \int_{\partial S} \psi(\mathbf{r})\phi(\mathbf{r})dS(\mathbf{r}). \tag{6}$$

The compatibility conditions (3) and (4) are types of reciprocity principles which are to hold in the absence of external applied body and boundary forces. Condition (3) can be interpreted as stating that the virtual power from the virtual velocity ϕ together with the body and boundary damping forces set up by the virtual velocity ψ must be equal to the virtual power with the roles of ϕ and ψ swapped. Similarly, condition (4) can be interpreted as stating that the virtual work from the virtual displacement ϕ together with the body and boundary restoring forces set up by the virtual displacement ψ must be equal to the virtual work with the roles of ϕ and ψ swapped.

Equations (1) and (2), together with specified initial conditions $u(\mathbf{r}, 0)$ and $\dot{u}(\mathbf{r}, 0)$ on S, are assumed to give a well-posed boundary value problem on $S \times [0, \infty)$ with solution $u \in V$, $\forall t \in [0, \infty)$. The goal is to find an approximate solution when the system is almost classically damped.

Undamped System. We start with the *undamped homogeneous* version of Eqs. (1) and (2) where L_1 and f are zero on S and each C_i and g_i is zero on $\partial S_i'$. We define a linear subspace U_0 of V consisting of all those functions in V which satisfy this simplified version of the boundary conditions (2). From Eq. (4),

$$\langle L_2\psi, \phi\rangle = \langle \psi, L_2\phi\rangle \quad \forall \psi, \phi \in U_0$$

so that L_2 is a self-adjoint linear operator on U_0 with respect to the inner product defined by (5). We assume, therefore, as is typical for vibration problems, that $m^{-1}L_2$ has a discrete spectrum of distinct positive eigenvalues $\{\omega_n^2: n \in \mathbf{Z}^+\}$ and a corresponding complete set of orthonormal real eigenfunctions $\{\phi_n(\mathbf{r}): n \in \mathbf{Z}^+\}$ in U_0 so that $\forall \psi \in V$:

$$\psi(\mathbf{r}) = \sum_{n=1}^{\infty} \langle \psi, m\phi_n\rangle \phi_n(\mathbf{r}) \quad \forall \mathbf{r} \in S \tag{7}$$

where the orthonormality guaranteed by the self-adjointness of L_2 gives.

$$\langle \phi_n, m\phi_p\rangle = \delta_{np}, \tag{8}$$

the Kronecker delta, and

$$L_2\phi_n = \omega_n^2 m\phi_n \quad \text{on } S, \tag{9}$$

$$D_N^{(i-1)}\phi_n = 0 \quad \text{on } \partial S_i \quad \text{and} \quad B_i\phi_n = 0 \quad \text{on } \partial S_i' \quad \forall i = 1, \ldots, n_b. \tag{10}$$

In general, the eigenfunction expansion in (7) will only converge to ψ on the boundary ∂S if ψ and the ϕ_n satisfy the same boundary conditions (e.g., Courant and Hilbert, 1989, Chapter 5). In this case, it is typical to have uniform convergence on $\bar{S}$ and no Gibbs phenomenon appears near the boundary ∂S. This behavior is advantageous in applications because then the eigenfunction series converges faster, that is, a smaller number of terms is required in (7) to give an approximation of the function ψ to within a specified accuracy.

Almost Classically Damped System. We now consider two special cases of the full damped linear system given by Eqs. (1) and (2):

Definition: The damped system is *classically damped* if the body and boundary damping operators satisfy

$$L_1\phi_n = \beta_n m\phi_n \quad \text{on } S \tag{11}$$

for some constants $\beta_n \geq 0$, and

$$C_i\phi_n = 0 \quad \text{on } \partial S_i' \quad \forall i = 1, \ldots, n_b. \tag{12}$$

This means that the eigenfunctions ϕ_n, $n \in \mathbf{Z}^+$, of L_2 on U_0 are also eigenfunctions of L_1 on U_0. For this to be possible, it is necessary that L_1 be a positive semidefinite self-adjoint linear operator on U_0 and that $m^{-1}L_1$ and $m^{-1}L_2$ commute on U_0. These conditions are also sufficient for (11) to hold (Caughey and O'Kelly, 1965). We shall hereafter denote the body and boundary damping operators for a classically damped system by L_{10} and C_{i0}, $i = 1, \ldots, n_b$, respectively.

The eigenfunctions of a classically damped system are identical to the undamped eigenfunctions, $\phi_n(\mathbf{r})$, $n \in \mathbf{Z}^+$. To show this, let $\hat{\phi}_n$ denote an eigenfunction corresponding to eigenvalue ρ_n, then $\hat{\phi}_n(\mathbf{r})e^{\rho_n t}$ is a solution of Eqs. (1) and (2) with $f = 0$ on S and each $g_i = 0$ on $\partial S_i'$, so

$$(\rho_n^2 m + \rho_n L_{10} + L_2)\hat{\phi}_n = 0 \quad \text{on } S$$

and $\forall i = 1, \ldots n_b$,

$$D_N^{(i-1)}\hat{\phi}_n = 0 \text{ on } \partial S_i, \quad (B_i + \rho_n C_{i0})\hat{\phi}_n = 0 \quad \text{on } \partial S_i'.$$

In view of Eqs. (9)–(12), $\hat{\phi}_n = \phi_n$ is the solution if ρ_n satisfies

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\rho_n^2 + \beta_n \rho_n + \omega_n^2 = 0. \quad (13)$$

The two roots of (13) corresponding to a given ϕ_n are a complex conjugate pair in the underdamped case, that is, when $\beta_n < 2\omega_n$.

Definition. The damped system defined by equations (1) and (2) is *almost classically damped* if

$$L_1 = L_{10} + \epsilon L_{11}, \quad C_i = C_{i0} + \epsilon C_{i1} \quad \forall i = 1, \ldots, n_b \quad (14)$$

where ϵ is a small parameter and L_{10} and the C_{i0} are the classically damped operators defined by (11) and (12), respectively.

In this case, the damping operators are perturbed slightly from the classically damped case, so it is possible to analyze the system using an approximate perturbation approach, as shown in the subsequent sections. Although ϵ is assumed to be small, the underlying classical damping need not be small. It is assumed that the compatibility condition (3) is satisfied by the classical-damping operators L_{10} and C_{i0}, $\forall i$, and so by linearity, the damping operators L_{11} and C_{i1}, $\forall i$, also satisfy this condition.

Approximate Solution of the Eigenproblem

Let $\hat{\phi}_p$, $p \in \mathbf{Z}^+$, be an eigenfunction of the almost classically damped system, with corresponding eigenvalue s_p, then

$$u_p(\mathbf{r}, t) = \hat{\phi}_p(\mathbf{r}) e^{s_p t}$$

is a solution of Eqs. (1) and (2) with $f = 0$ on S and each $g_i = 0$ on $\partial S_i'$, so

$$(s_p^2 m + s_p L_1 + L_2)\hat{\phi}_p = 0 \quad \text{on } S \quad (15)$$

and $\forall i = 1, \ldots n_b$

$$D_N^{(i-1)}\hat{\phi}_p = 0 \text{ on } \partial S_i \quad \text{and}$$

$$(B_i + s_p C_i)\hat{\phi}_p = 0 \quad \text{on } \partial S_i'. \quad (16)$$

Since ϵ is small in (14), we expect that $\hat{\phi}_p$ and s_p will remain close to the eigenquantities ϕ_p and ρ_p of the classically damped system, which satisfy Eqs. (9) through (13) when L_1 and C_i are replaced by L_{10} and C_{i0} in Eqs. (11) and (12). We therefore use the asymptotic expansions (Erdelyi, 1956)

$$\hat{\phi}_p(\mathbf{r}) \sim \phi_p(\mathbf{r}) + \epsilon\chi_p(\mathbf{r}) + \epsilon^2\psi_p + \ldots \quad (17)$$

$$s_p \sim \rho_p + \epsilon\sigma_p + \epsilon^2\tau_p + \ldots. \quad (18)$$

In the underdamped case, each s_p is complex-valued, and so, therefore, are the quantities in these expansions, except ϕ_p. Corresponding to each complex conjugate eigenvalue pair, ρ_p and $\bar{\rho}_p$, for the underlying classically damped system, there are perturbed eigenvalues s_p and $\bar{s}_p$, respectively, with associated eigenfunctions $\hat{\phi}_p(\mathbf{r})$ and $\bar{\hat{\phi}}_p(\mathbf{r})$. It is therefore only necessary to determine the perturbed eigenvalue and eigenfunction for one of the two complex roots of Eq. (13). In the overdamped case, however, it is necessary to determine separately a perturbed eigenvalue and eigenfunction for each of the real-valued roots of Eq. (13).

For suitably small ϵ, the second-order expansions are expected to be sufficient to capture the essential features of the eigenquantities. By substituting (17) and (18) into (15) and (16), taking Eqs. (9)–(14) into account, we get from the *first-order* terms in ϵ

$$\mathcal{L}_p\chi_p = \zeta_p \equiv -m\sigma_p(2\rho_p + \beta_p)\phi_p - \rho_p L_{11}\phi_p \quad \text{on } S \quad (19)$$

and $\forall i = 1, \ldots, n_b$,

$$D_N^{(i-1)}\chi_p = 0 \quad \text{on } \partial S_i,$$

$$\mathcal{B}_{ip}\chi_p = \zeta_{ip} \equiv -\rho_p C_{i1}\phi_p \quad \text{on } \partial S_i'. \quad (20)$$

Likewise, the *second-order* terms in ϵ give

$$\mathcal{L}_p\psi_p = \eta_p \equiv -(2m\rho_p\sigma_p + \sigma_p L_{10} + \rho_p L_{11})\chi_p - m(\sigma_p^2 + 2\rho_p\tau_p + \beta_p\tau_p)\phi_p - \sigma_p L_{11}\phi_p \quad \text{on } S \quad (21)$$

and $\forall i = 1, \ldots, n_b$,

$$D_N^{(i-1)}\psi_p = 0 \text{ on } \partial S_i,$$

$$\mathcal{B}_{ip}\psi_p = \eta_{ip} \equiv -(\sigma_p C_{i0} + \rho_p C_{i1})\chi_p - \sigma_p C_{i1}\phi_p \quad \text{on } \partial S_i'. \quad (22)$$

In these equations, the body and boundary operators $\mathcal{L}_p$ and $\mathcal{B}_{ip}$ are defined by

$$\mathcal{L}_p \equiv \rho_p^2 m + \rho_p L_{10} + L_2 \quad (23)$$

$$\mathcal{B}_{ip} \equiv B_i + \rho_p C_{i0}, \quad \forall i = 1, \ldots, n_b. \quad (24)$$

Using Eqs. (9)–(13), we find

$$\mathcal{L}_p\phi_q = m\pi_{pq}\phi_q \quad \text{on } S$$

$$\mathcal{B}_{ip}\phi_q = 0 \text{ on } \partial S_i', \quad \forall i = 1, \ldots, n_b \quad (25)$$

where π_{pq} is defined by

$$\pi_{pq} = \rho_p^2 + \rho_p\beta_q + \omega_q^2 \quad (26)$$

so that $\pi_{pp} = 0$, from Eq. (13). Using the compatibility Eqs. (3) and (4), we can show that $\forall\psi, \phi \in V$,

$$\langle \mathcal{L}_p\psi, \phi\rangle - \langle \psi, \mathcal{L}_p\phi\rangle = \sum_{i=1}^{n_b} [(D_N^{(i-1)}\psi, \mathcal{B}_{ip}\phi) - (\mathcal{B}_{ip}\psi, D_N^{(i-1)}\phi)]. \quad (27)$$

To determine the first-order terms χ_p and σ_p in (17) and (18) from Eqs. (19) and (20) in a way that avoids convergence problems on the boundary, we first employ the decomposition

$$\chi_p(\mathbf{r}) = \hat{\chi}_p(\mathbf{r}) + \mu_p(\mathbf{r}) \quad \text{on } \bar{S} \quad (28)$$

where μ_p is chosen as any function in V which satisfies the following boundary conditions for each $i = 1, \ldots, n_b$,

$$D_N^{(i-1)}\mu_p = 0 \quad \text{on } \partial S_i, \quad \mathcal{B}_{ip}\mu_p = \zeta_{ip} \quad \text{on } \partial S_i'. \quad (29)$$

Then, from Eq. (19), $\hat{\chi}_p \in V$ is found to satisfy the transformed equations

$$\mathcal{L}_p\hat{\chi}_p = \zeta_p - \mathcal{L}_p\mu_p \quad \text{on } S \quad (30)$$

and $\forall i = 1, \ldots, n_b$,

$$D_N^{(i-1)}\hat{\chi}_p = 0 \quad \text{on } \partial S_i, \quad \mathcal{B}_{ip}\hat{\chi}_p = 0 \quad \text{on } \partial S_i'. \quad (31)$$

Since $\hat{\chi}_p$ satisfies the same boundary conditions as the complete orthonormal set $\{\phi_q(\mathbf{r}), q \in \mathbf{Z}^+\}$ (see Eqs. (10) and (25)), its eigenfunction expansion

$$\hat{\chi}_p(\mathbf{r}) = \sum_{q=1}^{\infty} \alpha_{pq}\phi_q(\mathbf{r}) \quad \text{on } \bar{S} \quad (32)$$

where

$$\alpha_{pq} = \langle \hat{\chi}_p, m\phi_q\rangle \quad (33)$$

can also be expected to converge on the boundary ∂S (e.g., Courant and Hilbert, 1989, Chapter 5). To determine χ_p, it remains to determine the α_{pq}, $q \in \mathbf{Z}^+$.

Using Eqs. (25)–(27) and (29)–(31), we get

$$\langle \mathcal{L}_p\hat{\chi}_p, \phi_q\rangle = \pi_{pq}\alpha_{pq} \quad (34)$$

$$\langle \mathcal{L}_p\mu_p, \phi_q\rangle = \pi_{pq}\mu_{pq} + \rho_p\hat{\zeta}_{pq} \quad (35)$$

if we define

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\mu_{pq} = \langle \mu_p, m\phi_q \rangle \tag{36}$$

$$\zeta_{pq} = \sum_{i=1}^{n_b} (C_{i1}\phi_p, D_N^{(i-1)}\phi_q)_{\partial S_i'}. \tag{37}$$

But from Eq. (19)

$$\langle \mathcal{L}_p \chi_p, \phi_q \rangle = \langle \zeta_p, \phi_q \rangle = -\sigma_p(2\rho_p + \beta_p)\delta_{pq} - \rho_p \phi_{pq} \tag{38}$$

if we define

$$\phi_{pq} = \langle L_{11}\phi_p, \phi_q \rangle. \tag{39}$$

Using Eq. (30) to relate Eqs. (34), (35), and (38),

$$\pi_{pq}\alpha_{pq} = -\sigma_p(2\rho_p + \beta_p)\delta_{pq} - \rho_p\phi_{pq} - \pi_{pq}\mu_{pq} - \rho_p\zeta_{pq}. \tag{40}$$

When $p = q$, this equation determines the first-order eigenvalue correction constant

$$\sigma_p = -\rho_p \frac{\zeta_{pp} + \phi_{pp}}{2\rho_p + \beta_p}, \tag{41}$$

while for $p \neq q$, it determines the Fourier coefficients

$$\alpha_{pq} = -\rho_p(\zeta_{pq} + \phi_{pq})/\pi_{pq} - \mu_{pq}. \tag{42}$$

The coefficients α_{pp} will be determined later.

Next, in order to determine the second-order terms ψ_p and τ_p in (17) and (18) from Eqs. (21) and (22), we again employ a decomposition

$$\psi_p(\mathbf{r}) = \hat{\psi}_p(\mathbf{r}) + \upsilon_p(\mathbf{r}) \text{ on } \bar{S} \tag{43}$$

where υ_p is chosen as any function in V which satisfies for each $i = 1, \ldots, n_b$, the conditions

$$D_N^{(i-1)}\upsilon_p = 0 \quad \text{on } \partial S_i, \quad \mathcal{B}_{ip}\upsilon_p = \eta_{ip} \quad \text{on } \partial S_i'. \tag{44}$$

Then, from Eq. (21), $\hat{\psi}_p \in V$ satisfies

$$\mathcal{L}_p\hat{\psi}_p = \eta_p - \mathcal{L}_p\upsilon_p \quad \text{on } S \tag{45}$$

and $\forall i = 1, \ldots, n_b$,

$$D_N^{(i-1)}\hat{\psi}_p = 0 \quad \text{on } \partial S_i, \quad \mathcal{B}_{ip}\hat{\psi}_p = 0 \quad \text{on } \partial S_i'. \tag{46}$$

Therefore

$$\hat{\psi}_p(\mathbf{r}) = \sum_{q=1}^{\infty} \beta_{pq}\phi_q(\mathbf{r}) \quad \text{on } \bar{S} \tag{47}$$

where

$$\beta_{pq} = \langle \hat{\psi}_p, m\phi_q \rangle. \tag{48}$$

Using Eqs. (25)–(27) again and following a similar argument as before, we find

$$\langle \mathcal{L}_p\hat{\psi}_p, \phi_q \rangle = \pi_{pq}\beta_{pq} \tag{49}$$

$$\langle \mathcal{L}_p\upsilon_p, \phi_q \rangle = \pi_{pq}\upsilon_{pq} - \hat{\eta}_{pq}, \tag{50}$$

if we define

$$\upsilon_{pq} = \langle \upsilon_p, m\phi_q \rangle \tag{51}$$

$$\hat{\eta}_{pq} = \sum_{i=1}^{n_b} (\eta_{ip}, D_N^{(i-1)}\phi_q)_{\partial S_i'}. \tag{52}$$

But from Eq. (21),

$$\langle \mathcal{L}_p\psi_p, \phi_q \rangle = \langle \eta_p, \phi_q \rangle = -(\sigma_p^2 + 2\rho_p\tau_p + \beta_p\tau_p)\delta_{pq} - \theta_{pq}, \tag{53}$$

where we define

$$\theta_{pq} = \sigma_p\phi_{pq} + \langle (2m\rho_p\sigma_p + \sigma_p L_{10} + \rho_p L_{11})\chi_p, \phi_q \rangle. \tag{54}$$

Simplified expressions for $\hat{\eta}_{pq}$ and θ_{pq} defined in Eqs. (52) and (54) are given in Appendix A.

Using Eq. (45) to relate Eqs. (49), (50), and (53),

$$\pi_{pq}\beta_{pq} = -(\sigma_p^2 + 2\rho_p\tau_p + \beta_p\tau_p)\delta_{pq} - \theta_{pq} - \pi_{pq}\upsilon_{pq} + \hat{\eta}_{pq}. \tag{55}$$

When $p = q$, this equation determines the second-order eigenvalue correction constant

$$\tau_p = \frac{\hat{\eta}_{pp} - \theta_{pp} - \sigma_p^2}{2\rho_p + \beta_p}, \tag{56}$$

while for $p \neq q$, it determines the Fourier coefficients

$$\beta_{pq} = (\hat{\eta}_{pq} - \theta_{pq})/\pi_{pq} - \upsilon_{pq}. \tag{57}$$

Finally, α_{pp} and β_{pp} can be determined by normalizing the eigenfunction $\hat{\phi}_p$ so that

$$\langle \hat{\phi}_p, m\hat{\phi}_p \rangle = 1.$$

Employing Eqs. (8), (17), (28), (32), (33), and (48), the last condition yields

$$\alpha_{pp} = -\mu_{pp} \tag{58}$$

$$\beta_{pp} = -\upsilon_{pp} + \tfrac{1}{2}\sum_{q=1}^{\infty} (\mu_{pq}^2 - \epsilon_{pq}^2) \tag{59}$$

where

$$\epsilon_{pq} = \begin{cases} 0, & p = q \\ (\zeta_{pq} - \rho_p\phi_{pq})/\pi_{pq}, & p \neq q. \end{cases}$$

In summary, the above analysis gives an asymptotic approximation for the solution of the eigenproblem of almost classically damped continuous linear systems to second order in the small damping parameter ϵ in Eq. (14). The analysis breaks down when multiple eigenvalues occur in the underlying classically damped system, since then the quantity π_{pp} appearing in Eqs. (42) and (57) can be zero for $p \neq q$. A general procedure for treating this case can be found in Courant and Hilbert (1989). The analysis also breaks down at Eqs. (41) and (56), when $2\rho_p + \beta_p = 0$ for some $p \in \mathbf{Z}^+$, which corresponds to a critically damped system, that is, when $\beta_p = 2\omega_p$.

Approximate Solution for Forced Vibrations

Let $u(\mathbf{r}, t)$ be the general solution of (1) and (2) so that $u \in V$ for each time $t \in [0, \infty)$. We again use an asymptotic expansion in terms of the small damping parameter ϵ in Eq. (14),

$$u(\mathbf{r}, t) \sim \sum_{p=0}^{\infty} u_p(\mathbf{r}, t)\epsilon^p \quad \text{on } \bar{S} \times [0, \infty) \tag{60}$$

where $u_0(\mathbf{r}, t)$ is the solution for the classically damped case ($\epsilon = 0$). This gives a regular perturbation problem which will give a good approximation for the vibration response over some finite time interval, but not as $t \to \infty$ (Kevorkian and Cole, 1981). Substitution of this expression and Eq. (14) into the field Eq. (1) yields

$$\mathcal{L}u_p \equiv m\ddot{u}_p + L_{10}\dot{u}_p + L_2 u_p = f_p(\mathbf{r}, t) \quad \text{on } S \times (0, \infty) \tag{61}$$

with

$$f_p = \begin{cases} f(\mathbf{r}, t), & p = 0 \\ -L_{11}\dot{u}_{p-1}(\mathbf{r}, t), & p \in \mathbf{Z}^+. \end{cases} \tag{62}$$

Likewise, substitution of (60) and (14) into the boundary conditions (2) gives $\forall t \in (0, \infty)$ and $\forall i = 1, \ldots, n_b$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$D_N^{(i-1)}u_p = 0 \quad \text{on } \partial S_i,$$

$$\mathcal{B}_i u_p \equiv B_i u_p + C_{i0}\dot{u}_p = g_{ip}(\mathbf{r}, t) \quad \text{on } \partial S_i' \tag{63}$$

with

$$g_{ip} = \begin{cases} g_i(\mathbf{r}, t), & p = 0 \\ -C_{i1}\dot{u}_{p-1}(\mathbf{r}, t), & p \in \mathbf{Z}^+. \end{cases} \tag{64}$$

In order for the eigenfunction expansion (7) to be valid over the closure $\bar{S}$ of the domain S, the boundary conditions (63) must be put in homogeneous form. For this, we first employ the decomposition

$$u_p(\mathbf{r}, t) = v_p(\mathbf{r}, t) + h_p(\mathbf{r}, t) \quad \text{on } \bar{S} \times [0, \infty) \tag{65}$$

where h_p is chosen as any real-valued function in V for each t, which is twice-differentiable with respect to t almost everywhere on $(0, \infty)$ and satisfies the boundary conditions

$$D_N^{(i-1)}h_p = 0 \quad \text{on } \partial S_i,$$

$$\mathcal{B}_i h_p = g_{ip} \quad \text{on } \partial S_i', \quad \forall i = 1, \ldots, n_b. \tag{66}$$

Hence, $v_p \in V$ for each t, it is twice-differentiable with respect to t almost everywhere on $(0, \infty)$, and it satisfies

$$\mathcal{L}v_p = f_p - \mathcal{L}h_p \quad \text{on } S \tag{67}$$

$$D_N^{(i-1)}v_p = 0 \quad \text{on } \partial S_i, \quad \mathcal{B}_i v_p = 0 \quad \text{on } \partial S_i',$$
$$\forall i = 1, \ldots, n_b. \tag{68}$$

The functions v_p can be expressed in the form

$$v_p(\mathbf{r}, t) = \sum_{n=1}^{\infty} q_{pn}(t)\phi_n(\mathbf{r}) \quad \text{on } \bar{S} \tag{69}$$

where

$$q_{pn} = \langle v_p, m\phi_n \rangle. \tag{70}$$

We can obtain a differential equation for the $q_{pn}(t)$ by first employing the compatibility conditions (3) and (4):

$$\begin{aligned} \langle \mathcal{L}v_p, \phi_n \rangle &= \frac{d^2}{dt^2}\langle mv_p, \phi_n \rangle + \frac{d}{dt}\langle L_{10}v_p, \phi_n \rangle + \langle L_2 v_p, \phi_n \rangle \\ &= \ddot{q}_{pn} + \frac{d}{dt}\left\{ \langle v_p, L_{10}\phi_n \rangle + \sum_{i=1}^{n_b} [(D_N^{(i-1)}v_p, C_{i0}\phi_n) \right. \\ &\quad \left. - (C_{i0}v_p, D_N^{(i-1)}\phi_n)] \right\} \\ &\quad + \left\{ \langle v_p, L_2\phi_n \rangle + \sum_{i=1}^{n_b} (D_N^{(i-1)}v_p, B_i\phi_n) - (B_i v_p, D_N^{(i-1)}\phi_n) \right\}. \end{aligned} \tag{71}$$

By using Eqs. (9) to (12), (68), and (70), this expression can be reduced to

$$\langle \mathcal{L}v_p, \phi_n \rangle = \ddot{q}_{pn} + \beta_n \dot{q}_{pn} + \omega_n^2 q_{pn}. \tag{72}$$

In a similar fashion, except that Eq. (66) is used in place of (68), it can be shown that

$$\langle \mathcal{L}h_p, \phi_n \rangle = \ddot{h}_{pn} + \beta_n \dot{h}_{pn} + \omega_n^2 h_{pn} - \hat{g}_{pn}(t) \tag{73}$$

where we define

$$h_{pn} = \langle h_p, m\phi_n \rangle \tag{74}$$

$$\hat{g}_{pn} = \sum_{i=1}^{n_b} (g_{ip}, D_N^{(i-1)}\phi_n)_{\partial S_i'}. \tag{75}$$

Finally, define

$$f_{pn}(t) = \langle f_p, \phi_n \rangle. \tag{76}$$

A differential equation for each $q_{pn}(t)$ can then be obtained from (67), (72), (73), and (76),

$$\ddot{q}_{pn} + \beta_n \dot{q}_{pn} + \omega_n^2 q_{pn} = Q_{pn}(t) \tag{77}$$

where we define

$$Q_{pn}(t) = f_{pn} + \hat{g}_{pn} - (\ddot{h}_{pn} + \beta_n \dot{h}_{pn} + \omega_n^2 h_{pn}). \tag{78}$$

Here f_{pn} and $\hat{g}_{pn}$ for $p \in \mathbf{Z}^+$ depend on $\dot{u}_{p-1}$ through Eqs. (62) and (64), but this dependence can be simplified as shown in Appendix B.

To complete the specification of the coefficients $q_{pn}(t)$, the appropriate initial conditions must be given. From Eqs. (60), (65), and (70),

$$\begin{aligned} q_{pn}(0) &= \langle u_p(\mathbf{r}, 0), m\phi_n(\mathbf{r}) \rangle \\ &= \begin{cases} \langle u(\mathbf{r}, 0), m\phi_n(\mathbf{r}) \rangle - \langle h_0(\mathbf{r}, 0), m\phi_n(\mathbf{r}) \rangle, & p = 0 \\ -\langle h_p(\mathbf{r}, 0), m\phi_n(\mathbf{r}) \rangle, & p \in \mathbf{Z}^+ \end{cases} \end{aligned} \tag{79}$$

$$\begin{aligned} \dot{q}_{pn}(0) &= \langle \dot{u}_p(\mathbf{r}, 0), m\phi_n(\mathbf{r}) \rangle \\ &= \begin{cases} \langle \dot{u}(\mathbf{r}, 0), m\phi_n(\mathbf{r}) \rangle - \langle \dot{h}_0(\mathbf{r}, 0), m\phi_n(\mathbf{r}) \rangle, \\ \quad p = 0 \\ -\langle \dot{h}_p(\mathbf{r}, 0), m\phi_n(\mathbf{r}) \rangle, \\ \quad p \in \mathbf{Z}^+ \end{cases} \end{aligned} \tag{80}$$

where, without loss of generality, we can take $h_p(\mathbf{r}, 0)$ and $\dot{h}_p(\mathbf{r}, 0)$ as zero for $p = 0$ and $p \in \mathbf{Z}^+$.

In summary, an asymptotic approximation of the solution of the forced vibration problem to, say, second order in ϵ, may be obtained by using the eigenfunction expansion (69) for $p = 0$, 1, 2 and by taking a sufficient number of terms, N, for the desired accuracy. The coefficients in these expansions for the $v_p(\mathbf{r}, t)$ are determined by solving the initial-value problem (77)–(80) for $q_{pn}(t)$ for $p = 0$, 1, 2 and $n = 1, 2, \ldots, N$.

The functions $h_p(\mathbf{r}, t)$, $p = 0$, 1, 2, involved in Eqs. (78) and in the construction of the $u_p(\mathbf{r}, t)$ from Eq. (65), must be chosen to satisfy the conditions in Eqs. (66). A typical form for these functions can be obtained by generalizing an expression given in Meirovitch (1967) for the undamped problem, namely,

$$h_p(\mathbf{r}, t) = \sum_{i=1}^{n_b} [a_{ip}(\mathbf{r})\lambda_{ip}(t) + b_{ip}(\mathbf{r})\xi_{ip}(t)]$$

where the functions a_{ip}, λ_{ip}, b_{ip}, ξ_{ip} are chosen so that h_p satisfies conditions (66), as illustrated later. Once the $u_p(\mathbf{r}, t)$, $p = 0$, 1, 2, are determined, the solution $u(\mathbf{r}, t)$ can be approximated by the sum of the first three terms in Eq. (60).

Note that the approach developed here avoids direct term-by-term differentiation of the eigenfunction series for each u_p which would be required for $L_1 u$ and $L_2 u$. This operation would be invalid here since each u_p satisfies different boundary conditions than the eigenfunctions, $\phi_n(\mathbf{r})$.

Examples

To illustrate the approach, two simple examples are given in this section.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

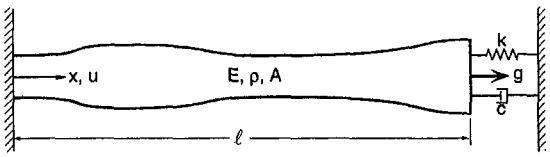


Fig. 1 Rod example

I: Rod Vibrations. Consider the longitudinal vibration of a viscoelastic rod shown in Fig. 1, where Eq. (1) applies with $u(x, t)$ being the longitudinal displacement at time t of a cross section at x and

$$S = (0, l), \quad \bar{S} = [0, l], \quad \partial S = \{0, l\}$$

$$m(x) = \rho(x)A(x), f(x, t)$$

= distributed longitudinal load per unit length

$$L_1 = \mu_E(x) - \frac{\partial}{\partial x}\left[\mu_I(x)A(x)\frac{\partial}{\partial x}\right],$$

$$L_2 = -\frac{\partial}{\partial x}\left[E(x)A(x)\frac{\partial}{\partial x}\right].$$

Here $m(x)$, $A(x)$, and $E(x)$ are the mass per unit length, the cross-sectional area and Young's modulus at location x along the rod, while $\mu_E(x)$ and $\mu_I(x)$ are non-negative viscous damping coefficients. From Fig. 1, the appropriate boundary conditions are geometric at $x = 0$ (i.e., ∂S_1) and a force balance at $x = l$ (i.e., $\partial S_1'$), which give

$$D_N^{(0)}u = u = 0 \text{ at } x = 0, \quad B_1 u + C_1 \dot{u} = g(t) \text{ at } x = l$$

where

$$B_1 = k + E(l)A(l)\frac{\partial}{\partial x}\bigg|_{x=l}, \quad C_1 = c + \mu_I(l)A(l)\frac{\partial}{\partial x}\bigg|_{x=l}$$

and $g(t)$ represents an applied axial force at $x = l$.

Using integration by parts and the appropriate inner product for this one-dimensional example, we get for all $\psi, \phi \in V \subset C^2(0, l) \cap C^1[0, l]$ satisfying the geometric boundary condition at $x = 0$:

$$\langle L_1\psi, \phi\rangle - \langle \psi, L_1\phi\rangle$$

$$= -\int_0^l \left[\phi\,\frac{d}{dx}\left(\mu_I A\frac{d\psi}{dx}\right) - \psi\,\frac{d}{dx}\left(\mu_I A\frac{d\phi}{dx}\right)\right]dx$$

$$= -\left[\mu_I A\frac{d\psi}{dx}\phi - \mu_I A\frac{d\phi}{dx}\psi\right]_0^l$$

$$= \left[\psi\left(c + \mu_I A\frac{d}{dx}\right)\phi - \phi\left(c + \mu_I A\frac{d}{dx}\right)\psi\right]_{x=l}$$

$$= [\psi C_1\phi - \phi C_1\psi]_{x=l} = (\psi, C_1\phi) - (C_1\psi, \phi).$$

According to (3) with $n_b = 1$, C_1 is therefore confirmed to be a compatible boundary operator for L_1. In a similar manner, B_1 can be confirmed to be a compatible boundary operator for L_2, as defined by Eq. (4).

For the undamped problem, the eigenvalues (frequencies) and eigenfunctions (modeshapes) are determined by

$$-[E(x)A(x)\phi_n'(x)]' = \omega_n^2 m(x)\phi_n(x) \quad \text{in } (0, l)$$

and

$$\phi_n(0) = 0, \quad B_1\phi_n|_{x=l} = E(l)A(l)\phi_n'(l) + k\phi_n(l) = 0$$

where the prime denotes d/dx. These equations define a Sturm-Liouville problem and so there is a countable set of eigenvalues $\{\omega_n^2 : n \in \mathbf{Z}^+\}$ and a complete set of orthonormal eigenfunctions $\{\phi_n : n \in \mathbf{Z}^+\}$ (Courant and Hilbert, 1989, Chapter 5).

For a classically damped system, we require from Eq. (11)

$$\mu_E(x)\phi_n(x) - [\mu_I(x)A(x)\phi_n'(x)]' = L_1\phi_n = \beta_n m(x)\phi_n(x).$$

This will be satisfied, for example, if we choose

$$\mu_E(x) = \alpha m(x), \quad \mu_I(x) = \beta E(x),$$

$$\beta_n = \alpha + \beta\omega_n^2, \quad \forall n \in \mathbf{Z}^+$$

where α and β are non-negative constants, which gives

$$L_{10} = \alpha m + \beta L_2$$

as a classical-damping operator. This is a continuous version of Rayleigh (or "proportional") damping. Also, from Eq. (12), we require

$$C_1\phi_n|_{x=l} = c\phi_n(l) + \beta E(l)A(l)\phi_n'(l) = (c - \beta k)\phi_n(l) = 0.$$

The presence of a spring at $x = l$ therefore necessitates a damper at $x = l$ with a particular damping coefficient $c = \beta k$ in order to give classical normal modes. For this value of c, the boundary damping operator becomes

$$C_{10} = \beta k + \beta E(l)A(l)\frac{\partial}{\partial x}\bigg|_{x=l} = \beta B_1.$$

Finally, the eigenvalues ρ_n, $n \in \mathbf{Z}^+$, for the classical-damping case are given by

$$\rho_n^2 + (\alpha + \beta\omega_n^2)\rho_n + \omega_n^2 = 0.$$

Now consider an almost classically-damped system where for a small parameter ϵ:

$$\mu_E(x) = \alpha m(x) + \epsilon\hat{\mu}_1(x), \quad \mu_I(x) = \beta E(x) + \epsilon\hat{\mu}_2(x),$$

$$c = \beta k + \epsilon c_1$$

so

$$L_1 = L_{10} + \epsilon L_{11}, \quad C_1 = C_{10} + \epsilon C_{11}$$

where

$$L_{11} = \hat{\mu}_1(x) - \frac{\partial}{\partial x}\left[\hat{\mu}_2(x)A(x)\frac{\partial}{\partial x}\right]$$

$$C_{11} = c_1 + \hat{\mu}_2(l)A(l)\frac{\partial}{\partial x}\bigg|_{x=l}.$$

Once the various field and boundary operators are defined, the next step involves the evaluation of the functions $\mu_p(x)$, $\upsilon_p(x)$, and $h_p(x, t)$ in Eqs. (28), (43), and (65), respectively, if both the eigenproblem and the forced vibration problem are of interest. The function $\mu_p(x)$ is evaluated by applying the conditions (29) on ∂S,

$$D_N^{(0)}\mu_p = \mu_p = 0 \quad \text{at} \quad x = 0$$

$$\mathcal{B}_{1p}\mu_p = \zeta_{1p} \quad \text{at} \quad x = l,$$

where

$$\mathcal{B}_{1p}\mu_p = B_1\mu_p + \rho_p C_{10}\mu_p = (1 + \beta\rho_p)B_1\mu_p$$

$$= (1 + \beta\rho_p)[k\mu_p(l) + E(l)A(l)\mu_p'(l)]$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\zeta_{1p} = -\rho_p C_{11}\phi_p = -\rho_p[c_1\phi_p(l) + \hat{\mu}_2(l)A(l)\phi_p'(l)].$$

Therefore, $\mu_p(x)$ can be chosen in the linear form,

$$\mu_p(x) = c_p x + d_p,$$

with the conditions at $x = 0$ and $x = l$ giving, respectively, $d_p = 0$ and

$$c_p = -\rho_p \frac{c_1\phi_p(l) + \hat{\mu}_2(l)A(l)\phi_p'(l)}{(1 + \beta\rho_p)[kl + E(l)A(l)]}.$$

This gives from Eqs. (36), (37), and (39)

$$\mu_{pq} = c_p \int_0^l xm(x)\phi_q(x)dx$$

$$\hat{\zeta}_{pq} = (C_{11}\phi_p, \phi_q)_{\partial S_1'} = [C_{11}\phi_p]_{x=l}\phi_q(l)$$

$$= c_1\phi_p(l)\phi_q(l) + \hat{\mu}_2(l)A(l)\phi_p'(l)\phi_q(l)$$

$$= [c_1 - k\hat{\mu}_2(l)/E(l)]\phi_p(l)\phi_q(l)$$

$$\phi_{pq} = \int_0^l \hat{\mu}_1(x)\phi_p(x)\phi_q(x)dx$$

$$- \int_0^l [\hat{\mu}_2(x)A(x)\phi_p'(x)]'\phi_q(x)dx$$

$$= \int_0^l \hat{\mu}_1(x)\phi_p(x)\phi_q(x)dx + \int_0^l \hat{\mu}_2(x)A(x)\phi_q'(x)dx$$

$$+ k\hat{\mu}_2(l)\phi_p(l)\phi_q(l)/E(l).$$

These parameters, together with function $\mu_p(x)$, permit the evaluation of all the parameters needed to determine the first-order eigenquantities σ_p and α_{pq} from Eqs. (41), (42), and (58).

In a similar fashion, application of condition (44) results in

$$D_N^{(0)}v_p|_{x=0} = v_p(0) = 0, \quad \mathcal{B}_{1p}v_p|_{x=l} = \eta_{1p}|_{x=l}$$

which in turn determine $v_p(x)$ in the form

$$v_p(x) = e_p x$$

where e_p is a known constant. This allows v_{pq} in Eq. (51) to be evaluated. The quantity $(\dot{\eta}_{pq} - \theta_{pq})$ can be evaluated as in Appendix A. The second-order eigenquantities τ_p and β_{pq} can then be determined from Eqs. (56), (57), and (59).

Finally, the functions $h_p(x, t)$, which are needed to determine the solution of the response problem up to some prescribed order in ϵ (e.g., taking $p = 0, 1, 2$) are chosen by applying conditions (66). In the present example, these conditions lead to

$$D_N^{(0)}h_p|_{x=0} = h_p(0, t) = 0$$

$$\mathcal{B}_1 h_p|_{x=l} = g_{1p}(t) = \begin{cases} g(t), & p = 0 \\ -C_{11}\dot{u}_{p-1}(l, t), & p \in \mathbf{Z}^+ \end{cases}$$

where $g(t)$ represents an applied axial force at $x = l$. Here

$$\mathcal{B}_1 h_p|_{x=l} = kh_p(l, t) + E(l)A(l)h_p'(l, t) + \beta k\dot{h}_p(l, t)$$

$$+ \beta E(l)A(l)\dot{h}_p'(l, t)$$

$$C_{11}\dot{u}_{p-1}(l, t) = c_1\dot{u}_{p-1}(l, t) + \hat{\mu}_2(l)A(l)\dot{u}_{p-1}'(l, t).$$

Next, choose h_p in the form

$$h_p(x, t) = \lambda_p(t)x + \xi_p(t).$$

Then, to satisfy the above conditions on h_p, we must have $\xi_p(t) = 0$ and

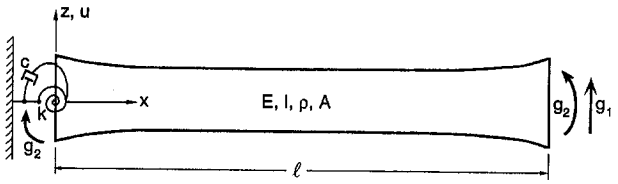


Fig. 2 Beam example

$$\beta\dot{\lambda}_p(t) + \lambda_p(t) = \frac{1}{[kl + E(l)A(l)]} g_{1p}(t)$$

where $g_{10}(t) = g(t)$ and $g_{1p}(t)$, $p \in \mathbf{Z}^+$, can be expressed as

$$g_{1p}(t) = -[c_1 + \hat{\mu}_2(l)A(l)]\dot{\lambda}_{p-1}(t)$$

$$+ \left[c_1 - \frac{k\hat{\mu}_2(l)}{E(l)}\right] \sum_{n=1}^{\infty} \dot{q}_{p-1,n}(t)\phi_n(l).$$

In practice, only a finite sum would be used to approximate $v_p(x, t)$ to sufficient accuracy and so the infinite sum in $g_{1p}(t)$ would be replaced by a finite one. The initial condition for the differential equation for $\lambda_p(t)$ depends on the choice of $h_p(x, 0)$, but this is arbitrary since whatever is chosen, it is corrected for by Eqs. (79) and (80). We take $h_p(x, 0) = 0$, so $\lambda_p(0) = 0$ and each $\lambda_p(t)$ can then be determined uniquely from a first-order linear initial value problem. This completes the determination of the functions $h_p(x, t)$. The functions $h_{pn}(t)$, and so $Q_{pn}(t)$, can be evaluated from Eqs. (74) and (78) and Appendix B. Then, the second-order linear initial value problem given by Eqs. (77), (79), and (80) can be solved to obtain each $q_{pn}(t)$. The function $v_p(x, t)$ can then be approximated by the first N terms in Eq. (69), where N is chosen to give the desired level of accuracy. Each $u_p(x, t)$ is then given by Eq. (65) to finally obtain an approximation of the solution $u(x, t)$ by Eq. (60).

II: Beam Vibrations. Consider the bending vibration of the beam shown in Fig. 2, where Eq. (1) applies with $u(x, t)$ being the transverse displacement at time t of a cross section at x and

$$S = (0, l), \quad \bar{S} = [0, l], \quad \partial S = \{0, l\}$$

$$m(x) = \rho(x)A(x),$$

$f(x, t)$ = distributed transverse load per unit length

$$L_1 = \mu_E(x) + \frac{\partial^2}{\partial x^2}\left[\mu_I(x)I(x)\frac{\partial^2}{\partial x^2}\right],$$

$$L_2 = \frac{\partial^2}{\partial x^2}\left[E(x)I(x)\frac{\partial^2}{\partial x^2}\right].$$

Here $m(x)$, $I(x)$, and $E(x)$ are the mass per unit length, the cross-sectional moment of inertia, and Young's modulus at location x along the beam, while $\mu_E(x)$ and $\mu_I(x)$ are non-negative viscous damping coefficients. Using Fig. 2, the boundary conditions (2) are

$$D_N^{(0)}u = u = 0 \text{ and } B_2(0)u + C_2(0)\dot{u} = g_2(0, t)$$

$$\text{at } x = 0$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$B_1(l)u + C_1(l)\dot{u} = g_1(l, t) \text{ and } B_2(l)u + C_2(l)\dot{u} = g_2(l, t) \quad \text{at } x = l$$

where

$$B_1(l) = -\frac{\partial}{\partial x}\left[E(x)I(x)\frac{\partial^2}{\partial x^2}\right]\Bigg|_{x=l},$$

$$C_1(l) = -\frac{\partial}{\partial x}\left[\mu_I(x)I(x)\frac{\partial^2}{\partial x^2}\right]\Bigg|_{x=l}$$

$$B_2(0) = -k\frac{\partial}{\partial x}\Bigg|_{x=0} + E(0)I(0)\frac{\partial^2}{\partial x^2}\Bigg|_{x=0},$$

$$C_2(0) = -c\frac{\partial}{\partial x}\Bigg|_{x=0} + \mu_I(0)I(0)\frac{\partial^2}{\partial x^2}\Bigg|_{x=0}$$

$$B_2(l) = E(l)I(l)\frac{\partial^2}{\partial x^2}\Bigg|_{x=l}, \quad C_2(l) = \mu_I(l)I(l)\frac{\partial^2}{\partial x^2}\Bigg|_{x=l}$$

and where $g_1(l, t)$ and $g_2(0, t)$, $g_2(l, t)$ are applied force and moments at the boundaries, while c and k are the damping and stiffness coefficients of the rotational dashpot and spring at $x = 0$.

Using integration by parts and the appropriate inner product, we get for all $\psi, \phi \in V \subset C^4(0, l) \cap C^3[0, l]$ satisfying the geometric boundary condition at $x = 0$,

$$\langle L_1\psi, \phi\rangle - \langle \psi, L_1\phi\rangle = [\psi C_1\phi - \phi C_1\psi]_0^l + [\psi' C_2\phi - \phi' C_2\psi]_0^l.$$

According to (3) with $n_b = 2$, C_1 and C_2 are therefore confirmed to be compatible boundary operators for L_1. In a similar manner, B_1 and B_2 can be confirmed to be compatible boundary operators for L_2.

For the undamped problem, the frequencies and modeshapes are determined by

$$[E(x)I(x)\phi_n''(x)]'' = \omega_n^2 m(x)\phi_n(x) \quad \text{in } (0, l)$$

$$\phi_n(0) = 0, \quad k\phi_n'(0) - E(0)I(0)\phi_n''(0) = 0$$

$$\frac{d}{dx}[E(x)I(x)\phi_n''(x)]|_{x=l} = 0, \quad E(l)I(l)\phi_n''(l) = 0.$$

We assume that this eigenvalue problem gives a countable set of eigenvalues $\{\omega_n^2: n \in \mathbf{Z}^+\}$ and a complete set of corresponding orthonormal eigenfunctions $\{\phi_n: n \in \mathbf{Z}^+\}$.

For a classically damped system, conditions (11) and (12) must be satisfied. This will be the case if we choose

$$\mu_E(x) = \alpha m(x), \quad \mu_I(x) = \beta E(x), \quad \beta_n = \alpha + \beta\omega_n^2$$

as in Example I, and then

$$C_1\phi_n|_{x=l} = -\beta\frac{d}{dx}[E(x)I(x)\phi_n''(x)]_{x=l} = 0$$

$$C_2\phi_n|_{x=l} = \beta E(l)I(l)\phi_n''(l) = 0$$

$$C_2\phi_n|_{x=0} = -c\phi_n'(0) + \beta E(0)I(0)\phi_n''(0) = (\beta k - c)\phi_n'(0) = 0$$

if the rotational dashpot and spring at $x = 0$ satisfy the condition $c = \beta k$ for classical normal modes. This means that the classical damping operators are

$$L_{10} = \alpha m + \beta L_2, \quad C_{10} = \beta B_1, \quad C_{20} = \beta B_2.$$

Now consider an almost classically damped system with:

$$\mu_E(x) = \alpha m(x) + \epsilon\hat{\mu}_1(x), \quad \mu_I(x) = \beta E(x) + \epsilon\hat{\mu}_2(x), \quad c = \beta k + \epsilon c_1,$$

so

$$L_1 = L_{10} + \epsilon L_{11} \quad \text{on } (0, l)$$

$$C_1 = C_{10} + \epsilon C_{11} \quad \text{at } x = l$$

$$C_2 = C_{20} + \epsilon C_{21} \quad \text{at } x = 0, l$$

where

$$L_{11} = \hat{\mu}_1(x) + \frac{\partial^2}{\partial x^2}\left[\hat{\mu}_2(x)I(x)\frac{\partial^2}{\partial x^2}\right]$$

$$C_{11}(l) = -\frac{\partial}{\partial x}\left[\hat{\mu}_2(x)I(x)\frac{\partial^2}{\partial x^2}\right]\Bigg|_{x=l}$$

$$C_{21}(0) = -c_1\frac{\partial}{\partial x}\Bigg|_{x=0} + \hat{\mu}_2(0)I(0)\frac{\partial^2}{\partial x^2}\Bigg|_{x=0}$$

$$C_{21}(l) = \hat{\mu}_2(l)I(l)\frac{\partial^2}{\partial x^2}\Bigg|_{x=l}.$$

At this point, the function $\mu_p(x)$ can be evaluated by applying the conditions (29) on the boundary

$$\mu_p(0) = 0, \quad \mathcal{B}_{1p}\mu_p = \zeta_{1p} \equiv -\rho_p C_{11}\phi_p \quad \text{at } x = l$$

$$\mathcal{B}_{2p}\mu_p = \zeta_{2p} \equiv -\rho_p C_{21}\phi_p \quad \text{at } x = 0, l$$

where from (24),

$$\mathcal{B}_{ip} = B_i + \rho_p C_{i0} = (1 + \beta\rho_p)B_i.$$

These conditions can be satisfied by a cubic of the form

$$\mu_p(x) = c_{p3}x^3 + c_{p2}x^2 + c_{p1}x$$

where the coefficients c_{p1}, c_{p2}, and c_{p3} are determined by the nonsingular linear system of equations

$$\frac{\partial}{\partial x}[E(x)I(x)(6c_{p3}x + 2c_{p2})]_{x=l} = \frac{\rho_p}{1 + \beta\rho_p}[C_{11}\phi_p]_{x=l}$$

$$kc_{p1} - 2E(0)I(0)c_{p2} = \frac{\rho_p}{1 + \beta\rho_p}[C_{21}\phi_p]_{x=0}$$

$$E(l)I(l)(6c_{p3}l + 2c_{p2}) = \frac{-\rho_p}{1 + \beta\rho_p}[C_{21}\phi_p]_{x=l}.$$

This allows the parameters μ_{pq} in (36) to be determined. Furthermore, ζ_{pq} and ϕ_{pq} in (37) and (39) are given by

$$\zeta_{pq} = [\phi_q C_{11}\phi_p]_{x=l} + [\phi_q' C_{21}\phi_p]_{x=0}^{x=l}$$

$$\phi_{pq} = \int_0^l \phi_q(x)L_{11}\phi_p(x)dx.$$

These parameters allow the first-order eigenvalue correction σ_p to be computed from (41) and the first-order eigenfunction correction $\chi_p(x)$ to be computed from (28), (32), (42), and (58).

In a similar fashion, $v_p(x)$ in (43) can be determined in the form

$$v_p(x) = e_{p3}x^3 + e_{p2}x^2 + e_{p1}x$$

by the conditions (44) that it must satisfy

$$v_p(0) = 0, \quad \mathcal{B}_{1p}v_p = \eta_{1p} \quad \text{at } x = l$$

$$\mathcal{B}_{2p}v_p = \eta_{2p} \quad \text{at } x = 0, l.$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

This allows the parameters v_{pq} in (51) to be determined. The quantity $(\hat{\eta}_{pq} - \theta_{pq})$ can be evaluated as in Appendix A. Therefore, the second-order eigenvalue correction τ_p can be computed from (56) and the second-order eigenfunction correction $\psi_p(x)$ can be computed from (43), (47), (57), and (59).

In order to determine the solution of the response problem up to, say, second order in ϵ, the functions $h_p(x, t)$ $(p = 0, 1, 2)$ in (65) must be specified so that they satisfy the conditions (66),

$$h_p(0, t) = 0, \quad B_1 h_p + \beta B_1 \dot{h}_p = g_{1p} \quad \text{at } x = l$$

$$B_2 h_p + \beta B_2 \dot{h}_p = g_{2p} \quad \text{at } x = 0, l,$$

where $g_{ip}(x, t)$ is given by (64). To satisfy these conditions, try a function of the form

$$h_p(x, t) = \sum_{i=1}^{2} [a_{ip}(x)\lambda_{ip}(t) + b_{ip}(x)\xi_{ip}(t)].$$

It follows that

$$0 = h_p(0, t) = \sum_{i=1}^{2} [a_{ip}(0)\lambda_{ip}(t) + b_{ip}(0)\xi_{ip}(t)]$$

$$g_{1p}(l, t) = \sum_{i=1}^{2} [B_1 a_{ip}(x)(\beta\dot{\lambda}_{ip}(t) + \lambda_{ip}(t)) + B_1 b_{ip}(x)(\beta\dot{\xi}_{ip}(t) + \xi_{ip}(t))]_{x=l}$$

$$g_{2p}(x, t) = \sum_{i=1}^{2} [B_2 a_{ip}(x)(\beta\dot{\lambda}_{ip}(t) + \lambda_{ip}(t)) + B_2 b_{ip}(x)(\beta\dot{\xi}_{ip}(t) + \xi_{ip}(t))] \quad \text{at } x = 0, l.$$

These equations are satisfied if

$$a_{1p}(0) = a_{2p}(0) = 0, \quad b_{1p}(0) = b_{2p}(0) = 0$$

$$B_1 a_{1p}|_{x=l} = B_1 a_{2p}|_{x=l} = 0, \quad B_1 b_{1p}|_{x=l} = 1, \quad B_1 b_{2p}|_{x=l} = 0$$

$$B_2 a_{1p}|_{x=l} = B_2 a_{2p}|_{x=l} = 0, \quad B_2 b_{1p}|_{x=l} = 0, \quad B_2 b_{2p}|_{x=l} = 1$$

$$B_2 a_{1p}|_{x=0} = 0, \quad B_2 a_{2p}|_{x=0} = 1, \quad B_2 b_{1p}|_{x=0} = B_2 b_{2p}|_{x=0} = 0$$

$$\beta\dot{\xi}_{1p}(t) + \xi_{1p}(t) = g_{1p}(l, t), \quad \beta\dot{\xi}_{2p}(t) + \xi_{2p}(t) = g_{2p}(l, t)$$

$$\beta\dot{\lambda}_{2p}(t) + \lambda_{2p}(t) = g_{2p}(0, t).$$

It is sufficient to set $a_{1p}(x) = 0$ on $[0, l]$ and so $\lambda_{1p}(t)$ is not required. Also, we can take $a_{2p}(x)$, $b_{1p}(x)$, and $b_{2p}(x)$ as cubics with the four coefficients in each cubic selected to satisfy the above four conditions on each of these functions. Finally, we can take $h_p(x, 0) = 0$ which gives the initial conditions $\xi_{1p}(0) = \xi_{2p}(0) = \lambda_{2p}(0) = 0$. This leads to a first-order linear initial value problem for each of $\xi_{1p}(t)$, $\xi_{2p}(t)$, and $\lambda_{2p}(t)$. Solution of these three initial value problems completes the determination of the functions $h_p(x, t)$.

The remainder of the procedure follows the description in Example I, where $h_{pn}(t)$, $Q_{pn}(t)$, and $q_{pn}(t)$ are determined successively and $v_p(x, t)$ in (69) is approximated by the first N terms to give the desired level of accuracy. Each $u_p(x, t)$ is then given by (65) to finally obtain an approximation of the beam displacement $u(x, t)$ by (60).

Summary

Conditions are presented for the existence of classical normal modes in continuous linear vibrating systems which have compatible (internal and boundary) stiffness and damping operators. A perturbation analysis is then developed for the response of almost classically damped continuous linear systems. The latter methodology provides approximate but sufficiently accurate analytical solutions for the eigenproblem as well as the response under general forcing functions.

The analysis is based on a decomposition of the internal and boundary damping operators into a part that results in classical normal mode response, with arbitrarily large damping ratios, plus a small perturbation. This permits application of standard modal analysis methodologies and provides the solution of the problem in terms of the frequencies and modes of the corresponding undamped problem only. A special treatment is presented for the series representations of the solutions in terms of the undamped eigenfunctions so that Gibbs phenomenon is avoided at the boundary. This ensures faster convergence of the eigenfunction expansions, that is, a smaller number of terms is required to give an approximation of the solutions to within a specified accuracy.

The methodology presented here is expected to result in computational benefits, especially when performing parametric studies to investigate damping effects for design purposes.

References

Bellos, J., and Inman, D. J., 1989, "A Survey of Nonproportional Damping," *Shock and Vibration Digest,* Vol. 21, No. 10, pp. 7–12.

Bergman, L. A., and Nicholson, J. W., 1985, "Forced Vibration of a Damped Combined Linear System," ASME *Journal of Vibration, Acoustics, Stress and Reliability in Design,* Vol. 107, pp. 275–281.

Caughey, T. K., 1960, "Classical Normal Modes in Damped Linear Dynamic Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 27, pp. 269–271.

Caughey, T. K., and O'Kelly, M. E. J., 1965, "Classical Normal Modes in Damped Linear Dynamic Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 32, pp. 583–588.

Chung, K. R., and Lee, C. W., 1986, "Dynamic Reanalysis of Weakly Nonproportional Damped Systems," *Journal of Sound and Vibration,* Vol. 111, pp. 37–50.

Courant, R., and Hilbert, D., 1989, *Methods of Mathematical Physics,* John Wiley and Sons, New York.

Cronin, D. L., 1976, "Approximation for Determining Harmonically Excited Response of Nonclassically Damped Systems," ASME *Journal of Engineering for Industry,* Vol. 98, pp. 43–47.

Erdelyi, A., 1956, *Asymptotic Expansions,* Dover, New York.

Foss, K. A., 1958, "Coordinates Which Uncouple the Equations of Motion of Damped Linear Dynamic Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 25, pp. 361–364.

Inman, D. J., and Andry, A. N., 1982, "The Nature of the Temporal Solutions of Damped Distributed Systems With Classical Normal Modes," ASME JOURNAL OF APPLIED MECHANICS, Vol. 49, pp. 867–870.

Inman, D. J., 1989, *Vibration with Control, Measurement and Stability,* Prentice-Hall, Englewood Cliffs, NJ.

Knowles, J. K., 1985, "On Damped Linear Dynamical Systems," *Ingenieur Archives,* Vol. 55, pp. 321–332.

Kevorkian, J., and Cole, J. D., 1981, *Perturbation Methods in Applied Mathematics,* Springer-Verlag, New York.

Meirovitch, L., 1967, *Analytical Methods in Vibrations,* Macmillan, New York.

Meirovitch, L., and Ryland, G., 1979, "Response of Slightly Damped Gyroscopic Systems," *Journal of Sound and Vibration,* Vol. 67, pp. 1–19.

Nashif, A. D., Jones, D. I. G., and Henderson, J. P., 1985, *Vibration Damping,* John Wiley and Sons, New York.

Natsiavas, S., 1993, "Dynamics of Multiple Degree of Freedom Oscillators with Colliding Components," *Journal of Sound and Vibration,* Vol. 165, pp. 439–453.

Natsiavas, S., and Beck, J. L., 1994, "Almost Classically Damped Linear Discrete Systems," *12th International Modal Analysis Conference,* Vol. 2, Honolulu, HI, pp. 1656–1661.

Nicholson, D. W., 1987, "Response Bounds for Nonclassically Damped Mechanical Systems Under Transient Loads," ASME JOURNAL OF APPLIED MECHANICS, Vol. 54, pp. 430–433.

Pan, H. H., 1966, "Vibration of a Viscoelastic Timoshenko Beam," *ASCE Journal of Engineering Mechanics,* Vol. 92, pp. 213–234.

Plaut, R. H., and Infante, E. F., 1972, "Bounds on Motions of Some Lumped and Continuous Dynamic Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 39, pp. 251–256.

Rayleigh, L., 1945, *Theory of Sound,* Vol. 1, Dover, New York.

Sharuz, S. M., and Ma, F., 1988, "Approximate Decoupling of the Equations of Motion of Linear Underdamped Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 55, pp. 716–720.

Snowdon, J. C., 1968, *Vibration and Shock in Damped Mechanical Systems,* John Wiley and Sons, New York.

Udwadia, F. E., and Esfandiari, R. S., 1990, "Nonclassically Damped Dynamic Systems: An Iterative Approach," ASME JOURNAL OF APPLIED MECHANICS, Vol. 57, pp. 423–433.

Udwadia, F. E., and Kumar, R., 1994, "Iterative Methods for Nonclassically Damped Dynamic Systems," *Earthquake Engineering and Structural Dynamics,* Vol. 23, pp. 137–152.

Yang, B., 1996, "Closed-Form Transient Response of Distributed Damped Systems, Part I: Modal Analysis and Green's Function Formula," ASME JOURNAL OF APPLIED MECHANICS, Vol. 63, pp. 997–1003.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

APPENDIX A

A simplified expression is derived for the difference of $\hat{\eta}_{pq}$ and θ_{pq} which are defined in Eqs. (52) and (54), respectively. This difference is needed to evaluate the second-order eigenvalue and eigenfunction corrections, τ_p and $\psi_p(\mathbf{r})$, given in Eqs. (56), (43), (44), (47), (57), and (59).

Start by defining

$$\gamma_{pq} = \sum_{i=1}^{n_b} (C_{i0}\chi_p, D_N^{(i-1)}\phi_q)_{\partial S_i'} \tag{81}$$

$$\hat{\gamma}_{pq} = \sum_{i=1}^{n_b} (C_{i1}\chi_p, D_N^{(i-1)}\phi_q)_{\partial S_i'}. \tag{82}$$

By substituting Eq. (22) into (52), η_{pq} can be expressed as

$$\hat{\eta}_{pq} = -\sigma_p\gamma_{pq} - \rho_p\hat{\gamma}_{pq} - \sigma_p\hat{\zeta}_{pq}. \tag{83}$$

Also, from Eq. (54),

$$\begin{aligned}\theta_{pq} &= \sigma_p\phi_{pq} + \langle(2m\rho_p\sigma_p + \sigma_pL_{10} + \rho_pL_{11})\chi_p, \phi_q\rangle \\ &= \sigma_p(2\rho_p + \beta_q)(\alpha_{pq} + \mu_{pq}) + \sigma_p\phi_{pq} \\ &\quad + \rho_p\langle L_{11}\chi_p, \phi_q\rangle - \sigma_p\gamma_{pq}.\end{aligned} \tag{84}$$

Here we use $\langle\chi_p, m\phi_q\rangle = \langle\hat{\chi}_p, m\phi_q\rangle + \langle\mu_p, m\phi_q\rangle = \alpha_{pq} + \mu_{pq}$, from Eqs. (28), (33), and (36), together with the compatibility Eq. (3) for L_{10} and the C_{i0}.

Subtracting (84) from (83),

$$\begin{aligned}\hat{\eta}_{pq} - \theta_{pq} &= -\rho_p\hat{\gamma}_{pq} - \sigma_p(\hat{\zeta}_{pq} + \phi_{pq}) \\ &\quad - \sigma_p(2\rho_p + \beta_q)(\alpha_{pq} + \mu_{pq}) - \rho_p\langle L_{11}\chi_p, \phi_q\rangle.\end{aligned} \tag{85}$$

This is the desired expression. Both $\hat{\gamma}_{pq}$ and $\langle L_{11}\chi_p, \phi_q\rangle$ can be evaluated using $\chi_p(\mathbf{r}) = \hat{\chi}_p(\mathbf{r}) + \mu_p(\mathbf{r})$ where $\hat{\chi}_p(\mathbf{r})$ is given by the eigenfunction expansion (32) on $\bar{S}$. In practice, only a finite number of terms, N, need be used in (32) to give the desired accuracy. Then $\hat{\gamma}_{pq}$ and $\langle L_{11}\chi_p, \phi_q\rangle$ can be approximated by

$$\begin{aligned}\hat{\gamma}_{pq} &= \sum_{i=1}^{n_b} (C_{i1}\hat{\chi}_p + C_{i1}\mu_p, D_N^{(i-1)}\phi_q)_{\partial S_i'} \\ &= \sum_{n=1}^{N} \alpha_{pn}\hat{\xi}_{nq} + \sum_{i=1}^{n_b} (C_{i1}\mu_p, D_N^{(i-1)}\phi_q)_{\partial S_i'}\end{aligned} \tag{86}$$

$$\begin{aligned}\langle L_{11}\chi_p, \phi_q\rangle &= \langle L_{11}\hat{\chi}_p + L_{11}\mu_p, \phi_q\rangle \\ &= \sum_{n=1}^{N} \alpha_{pn}\phi_{nq} + \langle L_{11}\mu_p, \phi_q\rangle.\end{aligned} \tag{87}$$

APPENDIX B

A simplified expression is derived for the sum $f_{pn}(t) + \hat{g}_{pn}(t)$ required in the differential Eq. (77) for $q_{pn}(t)$ where $p, n \in \mathbf{Z}^+$. (The $p = 0$ case is relatively straightforward.) Substituting Eqs. (62) and (64) into (76) and (75), we get

$$\begin{aligned}f_{pn} + \hat{g}_{pn} &= -\langle L_{11}\dot{u}_{p-1}, \phi_n\rangle - \sum_{i=1}^{n_b} (C_{i1}\dot{u}_{p-1}, D_N^{(i-1)}\phi_n)_{\partial S_i'} \\ &= -\langle\dot{u}_{p-1}, L_{11}\phi_n\rangle - \sum_{i=1}^{n_b} (D_N^{(i-1)}\dot{u}_{p-1}, C_{i1}\phi_n)_{\partial S_i'}\end{aligned}$$

where the compatibility condition (3) for L_{11} and the C_{i1} has been used. Equation (65) can now be used where $v_{p-1}(\mathbf{r}, t)$ is approximated by a finite number of terms, N, in the eigenfunction expansion (69), so that

$$\begin{aligned}f_{pn} + \hat{g}_{pn} &= -\sum_{m=1}^{N} \dot{q}_{p-1,m}(\phi_{nm} + \hat{\zeta}_{nm}) - \langle\dot{h}_{p-1}, L_{11}\phi_n\rangle \\ &\quad - \sum_{i=1}^{n_b} (D_N^{(i-1)}\dot{h}_{p-1,} C_{i1}\phi_n)_{\partial S_i'}\end{aligned} \tag{88}$$

where the definitions for $\hat{\zeta}_{nm}$ and ϕ_{nm} in (37) and (39) have been used. Equation (88) shows that the $q_{pn}(t)$ must be evaluated sequentially with respect to increasing p since q_{pn} depends on the $q_{p-1,m}$, $m = 1, 2, \ldots, N$.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

An Infrared Interferometer for Elastoplastic Crack-Tip Field Investigation of Homogeneous and Bimaterial Beams: A Study of Three-Dimensional Effects and *J*-Dominance

J. K. Sinha
Graduate Student

H. V. Tippur[1]
Associate Professor,
htippur@eng.auburn.edu,
Mem. ASME

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

An infrared interferometer capable of performing real-time full-field noncontacting deformation field measurements on optically rough surfaces is proposed as a tool for elastoplastic fracture mechanics investigations. The choice of the infrared wavelength allows interferometric measurements on fracture samples with little or no surface preparation and is more tolerant of the damage accumulation near the crack. The interferometer also bridges a sensitivity gap among existing techniques for out-of-plane deformation measurement. First, a rigorous Fourier optics analysis is provided for the interferometer and the range of surface roughness that can be studied using this interferometer is examined. The interferometer is then used for mapping deformations near elastoplastically deformed cracks in aluminum beams and solder-copper bimaterials. The regions of dominant three-dimensional effects and J-dominance are examined on the sample surface by evaluating measurements along with companion finite element analyses and the HRR fields.

1 Introduction

Currently there is a need for optical techniques with full-field real-time noncontact measurement capabilities preferably with little or no surface preparation. Most engineering material surfaces are "optically rough" (relative to the visible wavelength of light) and need elaborate surface preparation—polishing/thin film deposition to obtain a specular surface, printing gratings, etc.—prior to real-time metrology using visible wavelengths. Also, when damage accumulates, as in the case with the onset of plasticity, surface textural changes increase the roughness rendering visible light interferograms unclear or result in peeling of thin film deposition. For example, the deformation field near an elastoplastically deformed notch in an aluminum beam studied using the Coherent Gradient Sensor (CGS) (Tippur et al., 1991) is shown in Fig. 1. The sample surface was made specularly reflective through a series of processes consisting of lapping, polishing, and thin film deposition prior to measurements. With the onset of plasticity, clearly, the discernability of fringes near the notch tip is affected even when plasticity effects are relatively small (plastic zone size (r_p) 0.6 times the plate thickness). On the other hand, there are several optical methods—holography, laser speckle interferometry, shearography to name a few—that are capable of performing measurements on rough surfaces. However, they involve double-exposure (one of which is a reference exposure) schemes and hence do not offer real-time capability essential for many applications. Furthermore, surface textural changes and/or large deformations lead to speckle decorrelation effects resulting in a loss of fringe information and hence generally limited to elastic investigations.

A potential method of dealing with optically rough surfaces for real-time interferometric purposes is based on the theory of scattering of light by rough surfaces (Beckmann and Spizzichino, 1963). It is shown that the specular component of the scattered field increases with increasing wavelength for a randomly rough surface. Munnerlyn and Latta (1968) used mid-infrared radiation to relate fringe contrast and surface roughness based on average intensity measurement of an optical component in their studies. Optical component testing along with some quantitative insight into the influence of roughness of fringe contrast in the mid-infrared range of 10.6 μm is reported by Kwon et al. (1980). Recently, Lewandowski et al. (1989) have demonstrated infrared holographic interferometry facilitating fringe interpretations for large object displacements and minimizing the handicaps of visible holography. Persson (1993) has performed roughness measurement of machined surfaces using the speckle technique and has suggested that the measurement range could be extended if the experiments are performed in the infrared range.

In this paper an *infrared* Twyman-Green interferometer (wavelength 10.6 μm) has been developed as a new tool for elastoplastic crack-tip field mapping through rough surface measurements. The interferometer has real-time, full-field, and noncontact measurement capabilities. The textural changes and other surface roughness limitations commonly encountered when using visible radiation has been successfully overcome by a favorable roughness-to-wavelength ratio using infrared wavelengths. This offers greater flexibility in dealing with material surfaces in "as-is" conditions to perform real-time measurements. Furthermore, the measurement sensitivity offered 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.

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.

Manuscript received by the ASME Applied Mechanics Division, July 30, 1997; final revision, Jan. 16, 1998. Associate Technical Editor: I. M. Daniel.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

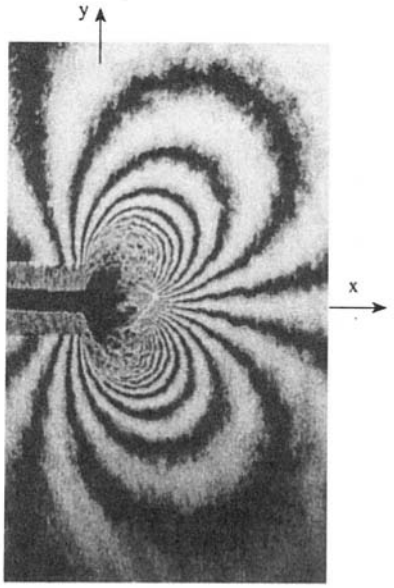


Fig. 1 Visible light interference patterns (contours of $\partial w/\partial y$) near an elastoplastically deformed notch in Al 6061 (Tippur et al., 1991); (plastic zone size ~0.6 times the plate thickness)

the proposed interferometer (~2.5 μm/half-fringe) bridges a large gap in the measurement sensitivities among existing methods of shadow and projection moiré methods (~15–25 μm/half-fringe) and visible light Twyman-Green interferometer (~0.3 μm/half-fringe). A rigorous Fourier optics analysis based on the Fresnel approximation and randomly rough surfaces is provided. Based on the result, the tolerable range of surface roughness by the interferometer is first investigated. Next, the proposed interferometer's capabilities for out-of-plane displacement field mapping near elastoplastically deformed cracks in homogenous and bimaterial fracture specimens is demonstrated. The measurements have been used for identifying the region of dominant three-dimensional deformations in homogeneous beam specimens and for evaluating the J-integral. Corresponding regions of three-dimensional effects and first experimental evidence of a HRR-like region near an elastoplastically deformed interfacial crack in a solder-copper bimaterial beam are provided.

2 Twyman-Green Interferometer: Optical Setup

The optical set-up for the infrared Twyman-Green interferometer is shown in Fig. 2. It consists of a CO_2 laser, lenses—L_1, L_2, and L_3, beamsplitters—BS1 and BS2, a reference mirror and a pyroviewer camera connected to a TV monitor, and a hard copy unit. The object to be investigated is a typical "rough" surface encountered in engineering. The optical interference is a result of the path difference between the reference and test paths as shown.

A water-cooled Invar stabilized CO_2 laser operating at 10.6 μm wavelength with continuous 7.5 W output power was used as a coherent light source. The unexpanded laser beam is 3.5 mm in diameter and is vertically polarized. A universal controller is used for general purpose interface between the user signal and CO_2 laser adding flexibility of power output control. The experiments reported in this paper have been conducted with low-power levels of approximately five percent of the nominal power of the laser. Two ZnSe lenses L_1 and L_2 were used to expand the laser beam to approximately 35-mm beam diameter. All lenses used were the meniscus type to minimize spherical aberrations. A ZnSe (50/50) beamsplitter was used to split the beam into two components as shown in Fig. 2. The optical coatings are the dielectric type and are optimized for 10.6 μm wavelength and 45 deg ± 2 percent incidence. The reference mirror is an aluminum-coated first-surface mirror. A Pyroviewer camera is used for imaging the interference fringes. This camera is designed to detect radiation in the 8–14 μm region. The recommended power threshold to prevent damage to the camera detector is 1 mW/cm^2. This being a rather small power, further reduction in the light intensity entering the optical set-up is accomplished by a 80/20 ZnSe beam splitter (BS2) that reflects away 80 percent of the incident energy as shown in Fig. 2.

3 Interference Analysis

Object surfaces encountered in engineering practice are generally "optically rough" and the differences are only in the degree and the type of roughness. Identifying the degree of roughness depends on both the geometry and the wavelength of the incident probe and is taken advantage of in this work. The types of roughness are generally classified into periodic roughness as in the case of machined surfaces and nonperiodic roughness, such as the random roughness of a sandblast surface. When a surface is illuminated with coherent light, the surface roughness and the orientation of the surface determine the general scattering pattern. As the surface roughness increases, the surface is perceived as changing from being smooth or specular to rough or diffuse. As the surface roughness increases the specular component of light decreases and the diffused light component increases. Thus specular reflection can be used as a measure of surface roughness. Rayleigh criterion characterizes the smoothness of a surface in terms of a parameter $\epsilon_R = (4\pi h/\lambda)\cos\theta_i (\epsilon_R \ll 1)$ where h is the surface peak amplitude, λ is wavelength of the incident light, and θ_i is the incident angle. Evidently, there are two ways of reducing ϵ_R: (i) by increasing

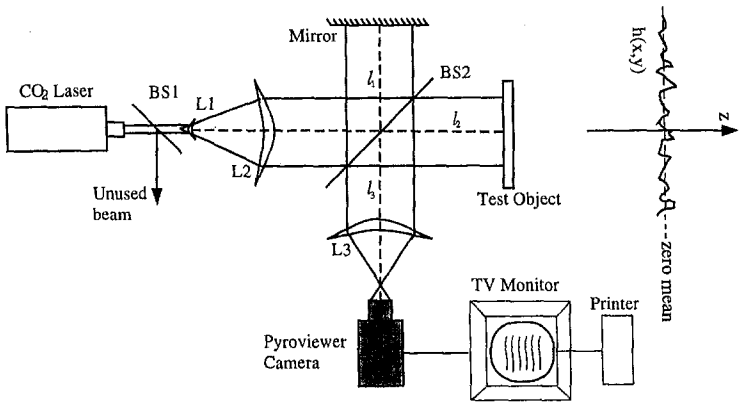


Fig. 2 Optical set-up for the Infrared Twyman-Green Interferometer

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

θ_i, and (*ii*) by increasing the wavelength (λ) of the incident light. Because the wavelength increase could be over an order of magnitude from the mid-visible to mid-infrared range and cosine function decreases when θ_i varies from 0–90 deg, it should be possible to extend the measurement techniques to surfaces that are traditionally considered far from being optically smooth.

In this work, the wavelength of light was increased in a Twyman-Green interferometer ($\theta_i = 0$) to perform out-of-plane deformation measurements on machined or natural surfaces. First, a study of the discernability of interference patterns in situations involving different surface roughness to test the feasibility of using infrared interferometry was performed. The fringe visibility and fringe contrast data were used to quantify operational limits of this interferometer in dealing with "rough" surfaces.

In the following a rigorous Fourier optics analysis of the interferometer that takes into account random surface roughness of the object is offered. The propagation of electromagnetic waves through the optical elements of the interferometer and the formation of interference patterns are systematically described by incorporating Fresnel diffraction principle based on the scalar theory of diffraction. The propagation of light waves from the jth plane to the $(j+1)$th plane located parallel to and at a distance z can be described as (Iizuka, 1985)

$$E_{j+1}(x_{j+1}, y_{j+1}) = \frac{1}{i\lambda z}\exp\left(ik\left\{z + \frac{x_{j+1}^2 + y_{j+1}^2}{2z}\right\}\right)\Im\{E_j(x_j, y_j)g(x_j, y_j)\},$$

$$\equiv \frac{1}{i\lambda z}\exp\left(ik\left\{z + \frac{x_{j+1}^2 + y_{j+1}^2}{2z}\right\}\right) \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_j(x_j, y_j)\exp\left(ik\frac{x_j^2 + y_j^2}{2z}\right) \times \exp\left(-\frac{ik}{z}\{x_j x_{j+1} + y_j y_{j+1}\}\right) dx_j dy_j, \quad (1)$$

where E represents the complex amplitudes on the respective planes, $i = \sqrt{-1}$, $g(x_j, y_j)$ is the point-source transfer function $g(x_j, y_j) = A\exp(ik((x_j^2 + y_j^2)/2z))$ and k $(=2\pi/\lambda)$ is the wave number with λ the wavelength of the light. In the above, $\Im(\cdot)$ denotes the two-dimensional Fourier transform operator. The positive quadratic phase term represents spherical waves emanating (diverging) from the current wave front to the next based on Huygens' principle which idealizes the wave front as a conglomeration of point sources.

Now, consider a plane wave incident on the object surface (x_o, y_o) and a planar reference mirror (x_r, y_r) as shown in Fig. 3. After reflecting from the object and the reference surfaces,

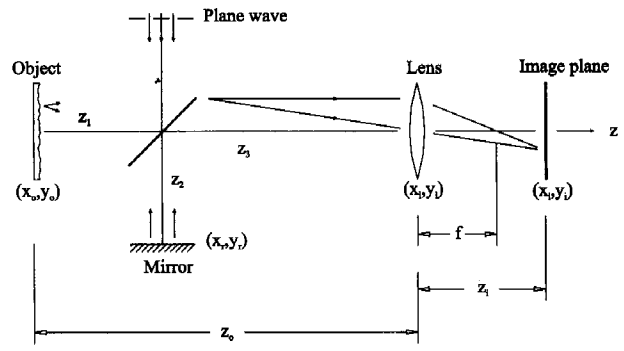


Fig. 3 Schematic of the interferometer used in the interference analysis

the waves propagate to the imaging lens located at the plane denoted by (x_l, y_l). Using Eq. (1) and letting $z_1 + z_3 = z_o$ and $z_1 + z_2 = z_r$, the complex amplitude distribution E_l^-, just before the lens plane, can be expressed as

$$E_l^-(x_l, y_l) = \frac{A_o}{i\lambda z_o}\exp\left(ik\left\{z_o + \frac{x_l^2 + y_l^2}{2z_o}\right\}\right) \times \Im\left\{E_o(x_o, y_o)\exp\left(ik\frac{x_o^2 + y_o^2}{2z_o}\right)\right\} + \frac{A_r}{i\lambda z_r}\exp\left(ik\left\{z_r + \frac{x_l^2 + y_l^2}{2z_r}\right\}\right) \times \Im\left\{E_r(x_r, y_r)\exp\left(ik\frac{x_r^2 + y_r^2}{2z_r}\right)\right\}, \quad (2)$$

where A_o and A_r denote the maximum amplitude of the object and reference waves upon reflection and $E_q(x_q, y_q)$ $(q = o, r, l, i)$ denote the complex amplitude distributions on different planes. A thin spherical lens is a phase object which modulates the incident waves as

$$E_l^+(x_l, y_l) = E_l^-(x_l, y_l)\exp\left(ik\left\{n_l\Delta_l - \frac{x_l^2 + y_l^2}{2f}\right\}\right), \quad (3)$$

where E_l^+ is the amplitude distribution exiting the lens and the negative sign in front of the quadratic phase term represents converging spherical waves to the back focal point when a positive lens is used. In the above, Δ_l is the maximum thickness, n_l the refractive index, and f the focal length of the lens. Using Eqs. (1)–(3), the amplitude distribution on the image plane (x_i, y_i) of the recording lens can be expressed as

$$E_i(x_i, y_i) = \frac{1}{i\lambda z_i}\exp\left(ik\left\{n_l\Delta_l + z_i + \frac{x_i^2 + y_i^2}{2z_i}\right\}\right) \times \Im\left\{E_l^-(x_l, y_l)\exp\left(ik\left\{-\frac{x_l^2 + y_l^2}{2f} + \frac{x_i^2 + y_i^2}{2z_i}\right\}\right)\right\}. \quad (4)$$

Substituting for $E_l^-(x_l, y_l)$ from Eq. (2) in the above we get

$$E_i(x_i, y_i) = \frac{1}{i\lambda z_i}\exp\left(ik\left\{n_l\Delta_l + z_i + \frac{x_i^2 + y_i^2}{2z_i}\right\}\right) \times \Im\left[\left(\frac{A_o}{i\lambda z_o}\exp\left(ik\left\{z_o + \frac{x_l^2 + y_l^2}{2z_o}\right\}\right)\right) \times \Im\left\{E_o(x_o, y_o)\exp\left(ik\frac{x_o^2 + y_o^2}{2z_o}\right)\right\} + \frac{A_r}{i\lambda z_r}\exp\left(ik\left\{z_r + \frac{x_l^2 + y_l^2}{2z_r}\right\}\right) \times \Im\left\{E_r(x_r, y_r)\exp\left(ik\frac{x_r^2 + y_r^2}{2z_r}\right)\right\}\right) \times \exp\left(ik\left\{-\frac{x_l^2 + y_l^2}{2f} + \frac{x_i^2 + y_i^2}{2z_i}\right\}\right)\right]. \quad (5)$$

Thus far, for clarity the coordinates of the object plane and the reference plane are distinguished through the use of subscripts o and r, respectively. In reality, the two are interchangeable, ($x_r \equiv x_o$ and $y_r \equiv y_o$) thereby implying that radiation from a generic

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

point on the object surface interferes with a plane wave generated by a corresponding point on the mirror. Now, for an equal arm interferometer with $z_o = z_r$ and when the imaging system is focussed on the object plane, we get

$$E_i(x_i, y_i) = c \exp\left(ik\left\{ n_l\Delta_l + z_i + \frac{x_i^2 + y_i^2}{2z_i} + z_o + \frac{x_o^2 + y_o^2}{2z_o} \right\} \right)$$

$$\times \{A_oE_o(-x_o, -y_o) + A_rE_r(-x_r, -y_r)\}, \quad (6)$$

where the identity $\Im(\Im\{E(x)\}) = E(-x)$ and the lens equation $(1/z_o) + (1/z_i) - (1/f) = 0$, have been utilized and c is a constant. The constant c will be dropped in the rest of the analysis since it has no bearing on the final outcome. The negative signs in front of the Cartesian coordinates in Eq. (6) physically suggests the inversion of the image of the object on the image plane.

To further describe the amplitude and hence the intensity distribution on the image plane, E_o and E_r need to be more specific. In the present interferometer, E_r denotes the amplitude distribution from a plane mirror perpendicular to the optical axis and hence $E_r(x_r, y_r) = 1$. For an object surface with a randomly varying roughness function $h(x_o, y_o)$ and deformation $w(x_o, y_o)$ relative to the initially flat plane

$$E_o(x_o, y_o) = \exp(2ik\{h(x_o, y_o) + w(x_o, y_o)\}), \quad (7)$$

where the factor 2 appears due to the reflective configuration of the interferometer. After substituting for E_o and E_r in Eq. (6), the intensity distribution on the image plane becomes

$$I = E_iE_i^*$$

$$= A_o^2 + A_r^2 + 2A_oA_r \cos[2k\{h(x_o, y_o) + w(x_o, y_o)\}], \quad (8)$$

where E_i^* is the complex conjugate of E_i. For a randomly rough object surface, $h(x_o, y_o)$ can be expressed statistically using probability density function $p(h)$. Among the available distribution forms, the most common one for random variations is the Gaussian distribution expressed in terms of rms roughness $\hat{\sigma}$ (Stover, 1995). A Gaussian surface height distribution with rms roughness $\hat{\sigma}$ has a probability distribution of the form

$$p(h) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} \exp\left(-\frac{h^2}{2\hat{\sigma}^2} \right). \quad (9)$$

Then the ensembled average value of the intensity distribution over the illuminated area of the object is

$$\langle I \rangle = \int_{-\infty}^{\infty} p(h)I(h)dh$$

$$= A_o^2 + A_r^2 + \frac{2A_oA_r}{\hat{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{h^2}{2\hat{\sigma}^2} \right)$$

$$\times \{\cos(2k\{h(x_o, y_o) + w(x_o, y_o)\})\}dh \quad (10)$$

where $\int_{-\infty}^{\infty} p(h)dh = 1$ has been used. Now, after expanding the cosine term in the above equation and simplifying, we get

$$\langle I \rangle = A_o^2 + A_r^2 + \frac{2A_oA_r \cos(2kw)}{\hat{\sigma}\sqrt{2\pi}}$$

$$\times \int_{-\infty}^{\infty} \exp\left(-\frac{h^2}{2\hat{\sigma}^2} \right) \{\cos(2kh)\}dh - \frac{2A_oA_r \sin(2kw)}{\hat{\sigma}\sqrt{2\pi}}$$

$$\times \int_{-\infty}^{\infty} \exp\left(-\frac{h^2}{2\hat{\sigma}^2} \right) \{\sin(2kh)\}dh$$

$$= A_o^2 + A_r^2 + 2A_oA_r \exp(-2k^2\hat{\sigma}^2) \cos(2kw(x_o, y_o)). \quad (11)$$

Equation (11) represents the intensity the interference fringes with maximas or minimas corresponding to the optical phase changes introduced by the out-of-plane deformation w. The result in Eq. (11) verifies the one obtained from an elementary analysis (Sinha and Tippur, 1997) wherein several implicit assumptions were made. The contrast of the interference fringes can now be expressed as

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2A_oA_r}{A_o^2 + A_r^2} \exp(-2k^2\hat{\sigma}^2). \quad (12)$$

Substituting for the wave number as $k = 2\pi/\lambda$, the fringe contrast can be expressed as a product of the contrast due to the surface roughness and the one due to the maximum amplitudes:

$$C = C_\alpha C_\sigma \quad \text{where}$$

$$C_\alpha = \frac{2A_oA_r}{A_o^2 + A_r^2} \quad \text{and} \quad C_\sigma = \exp\left(-\frac{8\pi^2\hat{\sigma}^2}{\lambda^2} \right). \quad (13)$$

In the presence of a specularly reflective object surface, ($C_\sigma = 1$), the equations simplify to the classical equations of a Twyman-Green interferometer. The deformation $w(x_o, y_o)$ can be determined from constructive interference fringes by equating the argument of the cosine in Eq. (11) to $2N\pi$ where N is an integer. This leads to

$$w(x_o, y_o) = \frac{N\lambda}{2}, \quad (14)$$

where $N = 0, \pm1, \pm2, \ldots$, represent the fringe orders. Thus, the light fringes occur for every $\lambda/2$ change in the optical path length and one could readily obtain a sensitivity of $\lambda/4$ (=2.65 μm) by utilizing both dark and light fringes. From the interference analysis it is clearly evident that the roughness of the surface essentially influences the contrast of the fringes and the fringe interpretation is unaltered. Thus, the above analysis can be viewed as a more general field description with specular surfaces as a special case.

4 Measurement Limits on Rough Surfaces

Experiments to determine the ability of the interferometer in dealing with various degrees of random roughness were first carried out. A circular aluminum blank was manually polished using different fixed abrasives to vary surface roughness. (This method also represents the most common method of preparing surfaces for optical metrology.) Fixed abrasives of various commercially available grit sizes (#2000 (mean grit size = 1 μm) to #60 (mean grit size = 250 μm) emery paper) were used under dry conditions to alter the surface roughness. The resulting roughness $\hat{\sigma}$ was measured using a surface profilometer. The samples were then placed in the interferometer to obtain "wedge" fringes by applying a known rigid rotation about the horizontal axis resulting in an out-of-plane displacement of the form $w(x, y) \approx c_1 + c_2y$, where c_i denote constants. Some representative fringes and the corresponding intensity distribution along the vertical diameter are shown in Fig. 4. These intensity profiles were subsequently analyzed to quantify average fringe contrast. The pixel intensity data along the vertical diameter over a narrow band were used in each case. A plot of normalized (with respect to an aluminum coated first surface mirror surface) fringe contrast as a function of the surface roughness parameter as well as grit size of the abrasive are shown in Figs. 5(a) and (b). Evidently, as the roughness of the surface increases, the fringe contrast deteriorates rapidly. The surface quality after polishing with #240 emery paper (mean abrasive grit size = 20 μm) was found to provide a sufficiently high fringe contrast. Experiments reported here used #600 fixed abrasives to obtain fringe contrast in excess of 0.8.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

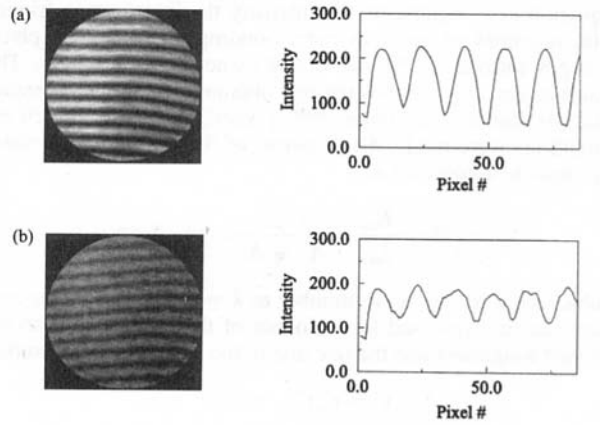


Fig. 4 Typical "wedge" fringes from the surfaces prepared using different grit size polishing papers (*a*) #600 (mean grit size 15 μm) (*b*) #60 (mean grit size 250 μm) and their corresponding intensity profile

5 Elastoplastic Crack-Tip Deformation Mapping

The ability of the interferometer to measure elastic deformations on rough surfaces has been demonstrated recently by the authors (Sinha and Tippur, 1997). In the following, crack-tip fields are mapped in homogeneous and bimaterial beams with little or no surface preparation to examine three-dimensional effects and J-dominance in the presence of substantial plastic deformations.

5.1 Homogeneous Cracked Beam. Aluminum (Al 6061-T6) beam samples were machined from a 6.25-mm rolled sheet of aluminum as obtained from the manufacturer. A through-the-thickness edge discontinuity of length (a) 6 mm and thickness (b) 150 μm was machined into the sample using a slitting saw. The specimen and its loading configuration (L = 100 mm, W = 25 mm) are shown in Fig. 6(a). *In this case, the sample was used without any surface preparation* (surface roughness 0.36 μm in the rolling direction and 0.34 μm in the transverse direction). It should be noted that this roughness ($\hat{\sigma}/\lambda \approx 0.03$) yields sufficiently high fringe contrast without any surface preparation. The specimen was subjected to symmetric three-point bending in a displacement-controlled loading device. A 0.3-mm diameter pencil lead (commonly used in mechanical pencils) was temporarily attached to the sample surface to obtain a sharp focus of the crack-tip region. The crack-tip deformations were mapped using the interferometer in real-time. Some representative interference patterns of contours of constant $w(x, y)$ fringes near the crack tip for different load levels are shown in Fig. 6(b) and (c). Knowing the nominal thickness B of the beam, the average out-of-plane strain ϵ_z ($\approx 2w(x, y)/B$) can be estimated from the fringes. In Fig. 6(c) crack-tip fringes represent a strain ϵ_z of approximately 0.02 while the outermost fringe corresponds to 0.0045. It should be noted that these values represent strain values beyond the yield strain ϵ_0 ($\approx$0.0038) for Al 6061-T6. Further, it will be shown later that the Dugdale plastic zone estimate (r_p) in this case is approximately 1.8B while the one shown in Fig. 1 is only 0.6B for the same material. As discussed earlier, the permanent deformations near the crack tip alter the surface texture and increase the local roughness and affects the fringe visibility when visible light is used. However, the discernability and contrast of the fringes near the crack-tip region are rather excellent in Fig. 6 with the use of infrared wavelength.

To obtain an estimate of the size of the zone of dominant three-dimensional deformations near the crack tip, a complementary elastoplastic plane-stress finite element analysis based on the incremental theory of plasticity and the J_2-flow rule was carried out using ANSYS software package. The finite element model consisted of 1728 eight-node plane-stress quadrilateral elements and 5377 nodes. No singular elements were embedded at the crack tip. The crack-tip region was discretized using a fine mesh with 576 crack-tip elements within a radius of r/B = 1 and 1440 elements within r/B = 2 from the crack tip. It was ensured that finite element discretization was adequate to satisfactorily capture the full-field singular behavior near the crack tip and the results have been reported elsewhere (Sinha et al., 1997). The stress-strain relations necessary for the elastoplastic simulations were based on a uniaxial material model

$$\left(\frac{\epsilon}{\epsilon_o}\right) = \left(\frac{\sigma}{\sigma_o}\right) \quad \text{for} \quad (\sigma \le \sigma_o)$$

$$\left(\frac{\epsilon}{\epsilon_o}\right) = \alpha\left(\frac{\sigma}{\sigma_o}\right)^m \quad \text{for} \quad (\sigma > \sigma_o). \tag{15}$$

In the above, ϵ_o, σ_o are yield strain and stress, m is the hardening index, and α is a constant. For Al 6061-T6, values of m = 18 and α = 1.22, were used (Chiang et al., 1988). The model was subjected to three-point bending similar to the experiments. Using material incompressibility, out-of-plane displacements can be computed around the crack tip using

$$\epsilon_z^p = -(\epsilon_r^p + \epsilon_\theta^p) \approx -\frac{2w}{B}, \tag{16}$$

where the superscript "p" denotes plastic component. The measured and computed values of out-of-plane displacements are plotted in Fig. 7 for P/P_o = 1.04 along several discrete directions θ ahead of the crack tip. Here the out-of-plane displacements and radial distances are normalized by the initial notch width b and specimen thickness B, respectively. In each of these plots, the region of disagreement between the two data sets by more than ten percent is identified as a region of dominant three-dimensional effects. Generally, three-dimensional deformations seem to dominate within $r/B \approx 0.35$, when 67.5 deg $< \theta <$ 67.5 deg. (In the regions beyond this value of θ, the optical information is rather limited and direct comparison was not carried out.) Beyond this region, plane-stress behavior

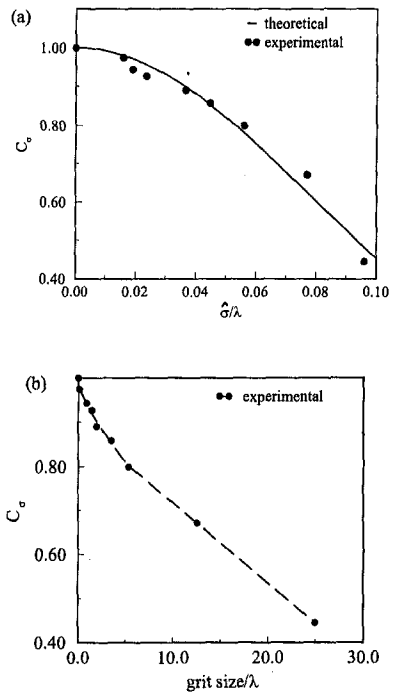


Fig. 5 Plot of fringe contrast as a function of (*a*) surface roughness (*b*) grit size

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

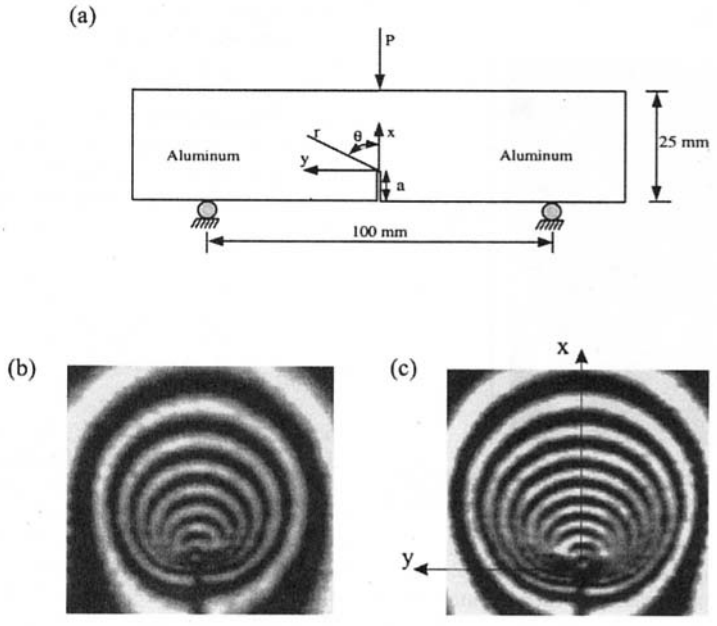


Fig. 6 Homogeneous three-point bending experiments: (*a*) loading configuration, (*b*) and (*c*) interference patterns representing contours of $w(x, y)$ fringes near the crack tip for $P/P_0 = 0.98$ and 1.04, respectively

is evident. The size of the three-dimensional zone seen here is generally comparable to the ones reported in the literature. A three-dimensional zone of size $r/B \cong 0.5$ is reported by Zehnder and Rosakis (1990) in three-point bend samples in steel samples with EDM cut notches using caustics and interferometry. A somewhat larger three-dimensional zone of size 0.8–1.0B is observed by Chiang and Hareesh (1988) in aluminum tensile sheets with extensive plasticity near fatigue cracks using geometric moiré methods and a more stringent error band of 2.5 percent. Crack tunneling on the midplane of the sheet and lower measurement sensitivity are other possible explanations for the difference.

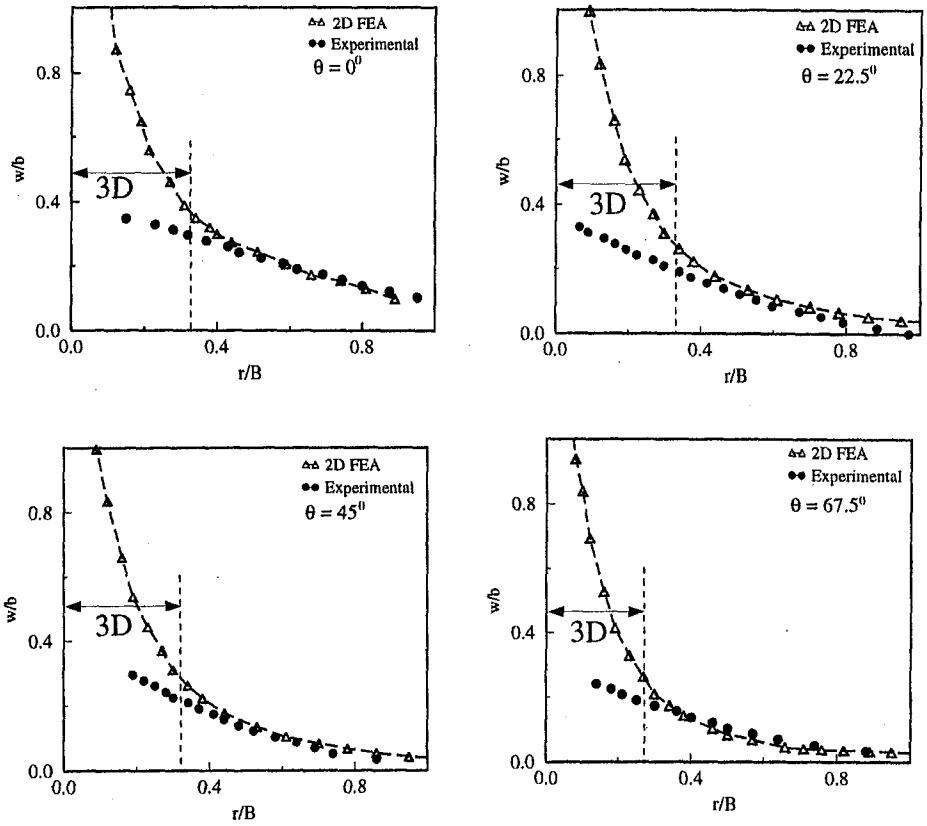


Fig. 7 Variation of out-of-plane displacements from the measurements and two-dimensional plane-stress simulations along different directions for the aluminum beam

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

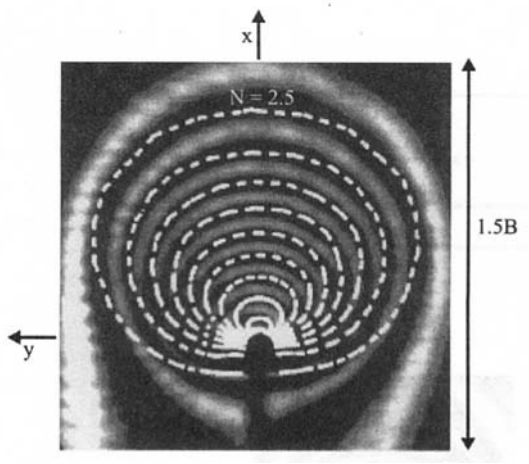


Fig. 8 Superposition of finite element contours (broken lines) from a three-dimensional simulation over the interferometric fringes for the aluminum beam

To ensure that the differences between the measurements and two-dimensional simulations are indeed due to three-dimensional effects, an elastoplastic three-dimensional finite element analysis was carried out. Four layers of eight-node (three-degree-of-freedom/node) brick elements were used in the model which consisted of 4000 elements and 5226 nodes. The contours of constant out-of-plane displacement were computed on the surface. A direct superposition of the experimental fringes and the finite element contours (broken lines) around the crack tip is shown in Fig. 8. A good agreement between the two results is readily evident even in the regions very close to the crack tip previously identified as regions of dominant three-dimensional effects. In Fig. 9(*a*), a representative plot of the measurements along with the ones from two-dimensional and three-dimensional simulations are shown for $\theta = 22.5$ deg and good agreement between measurements and three-dimensional simulation clearly becomes evident.

The J-integral is the fracture parameter which describes the strain amplification near a ductile crack tip (Rice, 1968). The optical measurements can be used to directly calculate the J-integral using the so-called HRR equation (Hutchinson, 1968; Rice and Rosengren, 1968) for out-of-plane strain

$$\epsilon_z^p \approx -\frac{2w}{B} = -\frac{\alpha\sigma_o}{E}\left(\frac{EJ}{\alpha\sigma_o^2 I_m r}\right)^{m/(m+1)} [\hat{\epsilon}_r(\theta; m) + \hat{\epsilon}_\theta(\theta; m)]. \quad (17)$$

In the above, functions $\hat{\epsilon}_r$, $\hat{\epsilon}_\theta$, and I_m are tabulated by Shih (1983) for known values of material hardening index m. Now, using Eqs. (17) and (14), the value of the J-integral for different fringe orders are evaluated and are plotted as a function of r/B in Fig. 9(*b*). Evidently, the value of the J-integral outside the region of three-dimensional deformations remains relatively constant over $0.35 < r/B < 0.65$ (shaded region) and can be identified as the region of J-dominance. As one moves farther away from the crack tip ($r/B > 0.65$), the far-field effects become significant and the evaluated values of the J-integral are no longer constant.

The value of the J-integral can also be estimated independently as $J = J_e + J_p$, where J_e is computed from the elastic stress intensity factor using an effective crack length. Consider the plane-stress condition

$$J_e = \frac{K_I^2(a_e)}{E}, \quad (18)$$

where

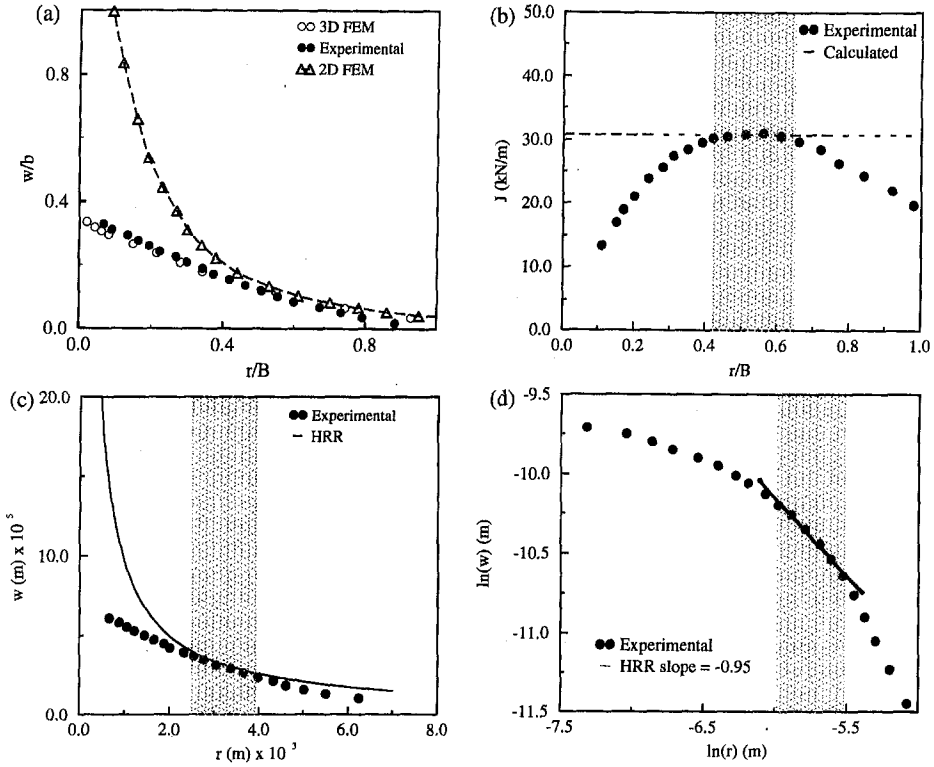


Fig. 9 Crack-tip three-dimensionality and *J*-dominance in aluminum beams ($\theta = 22.5$ deg); (*a*) plot of normalized displacements with normalized distance, (*b*) computed and measured values of the *J*-integral with normalized distance, (*c*), (*d*) comparison of optical measurements with HRR-prediction

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$a_e = a + \frac{1}{1 + \left(\frac{P}{P_o}\right)^2} \frac{1}{2\pi} \left(\frac{m-1}{m+1}\right) \left(\frac{K_I(a)}{\sigma_o}\right)^2 .$$

Now

$$K_I = \frac{P}{B\sqrt{W}} f\left(\frac{a}{W}\right), \quad \text{where} \quad f\left(\frac{a}{W}\right)$$

is a nondimensional function of the ratio a/W. The effective crack length (a_e) is evaluated iteratively until convergence is achieved. The value of J_p is estimated using

$$J_p = \alpha \epsilon_o \sigma_o (W - a) h_1 \left(\frac{P}{P_o}\right)^{m+1}, \qquad (19)$$

where $P_o (= (1.072B(W-a)^2\sigma_o)/L)$ is a reference load corresponding to a fully plastic ligament under plane-stress conditions. A detailed tabulation of the function $h_1(a/W)$ can be found in Kumar et al. (1981). From the above, the calculated value of the J-integral was found to be 30.8 kN/m and is in good agreement with the measured values of the J-integral in the shaded region of Fig. 9(b). Outside this region, ($0 < r/B < 0.35$ and $r/B > 0.65$) the J-integral is underestimated by the measurements.

The measured out-of-plane displacements are finite at the crack tip unlike the HRR predictions based on two-dimensional assumptions. The out-of-plane displacements based on HRR prediction and the measured value of the J-integral is plotted in Fig. 9(c) (solid line). The predictions and measurements deviate substantially very close to the crack tip while the agreement is rather good in the region $0.35 < r/B < 0.65$ (shaded region). Based on the measured value of the J-integral, one can get a Dugdale plastic zone estimate $r_p = (\pi/8)(JE/\sigma_o^2) \cong 1.8B$ for this load level. Previous studies (Zehnder and Rosakis, 1990) have shown that the HRR region exists within $r = 0.3r_p$. The shaded region indeed corresponds to such a prediction with three-dimensional region acting as the inner limit of the zone of J-dominance. From Eq. (17) it can be seen that the slope of log (w) versus log (r) plot should be $-(m/(m+1))$ (≈ -0.95 for Al 6061-T6) for HRR field to prevail. The data points corresponding to the shaded region in Fig. 9(c) indeed reflect the expected slope in Fig. 9(d) suggesting a J-dominant behavior over this distance.

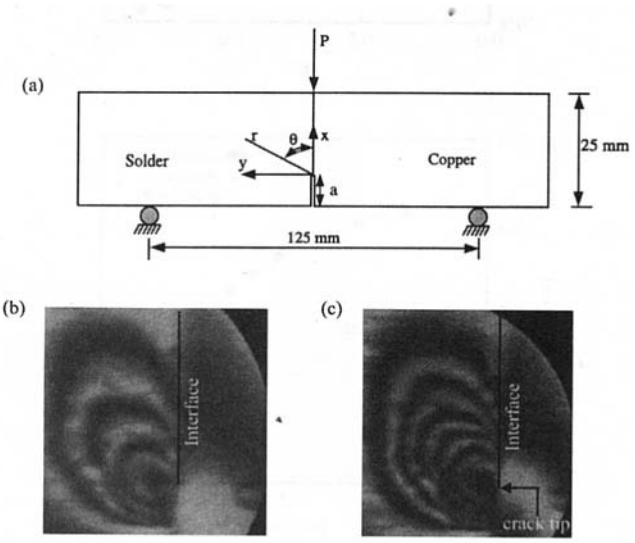


Fig. 10 Bimaterial three-point bending experiments; (*a*) loading configuration, (*b*) and (*c*) interference patterns representing contours of $w(x, y)$ fringes near the crack tip for $P/P_0 = 1.57$ and 1.65, respectively

Table 1 Material properties of the bimaterial samples

Material	E (GPa)	ν	σ_0 (MPa)	m	α
63/37 solder	20	0.4	20	11	1
Copper	124	0.33	70	—	—

5.2 Solder-Copper Bimaterial Beam With an Interfacial Crack. At the moment, optical investigation of deformation fields near elastoplastically deformed interfacial cracks is lacking. The method of infrared Twyman-Green interferometry is next extended to map out-of-plane displacements near elastoplastic interfacial cracks. Solder-copper bimaterials, which have a great deal of significance to the electronic industry, are examined using the new optical technique. The solder used in the investigation was obtained from 63/37 Sn-Pb solder ingots. Bimaterial samples (150 × 25 × 5 mm) with edge cracks ($a/W = 0.25$) were prepared by joining solder and copper bars (see Fig. 10(a)) in a specially designed aluminum fixture. An edge crack was introduced between the solder and copper bars during the bonding process. Soldering flux (Kester 185, type RMA) was applied to the edges of the bars where bonding was needed and was allowed to stay until the flux became tacky. The fixture was heated until the solder started melting. The bonded sample was then cooled to room temperature in air and subsequently machined to the final dimensions. The resulting bimaterial had interfacial cracks of thickness less than 50 μm. *The surface preparation consisted of simply polishing the surface using #600 grit fixed abrasive.* The beam samples were subjected to symmetric three-point bending with the loading point located <1 mm away from the interface on the copper side. A monotonically increasing load (strain rate in solder $\dot{\epsilon} \approx 10^{-4}$ sec^{-1}, Krishnamoorthy and Tippur, 1998) was applied to the sample in a displacement controlled loading device and the crack-tip deformations were measured in real-time. Typical interference patterns representing contours of constant $w(x, y)$ are shown in Figs. 10(b) and (c) for two load levels $P/P_o =$ 1.57 and 1.65. Here reference load P_o was calculated based on the yield stress of solder (see, Table 1). It is clearly evident that the deformations primarily occur on the solder side of the bimaterial and the ones on the copper side are negligible. Further, the deformations on the solder side of the bimaterial represent strain levels beyond the yield strain ϵ_o of solder ($\approx$0.001) in the entire region of observation. That is, the smallest measurable fringe order $N \approx 0.5$ denotes a strain component ϵ_z approximately equal to ϵ_o.

To obtain an estimate of the size of the zone of dominant three-dimensional deformations near the crack tip, a complementary elastoplastic plane-stress finite element analysis was carried out similar to the ones reported in the Section 5.1. The 63/37 solder is a relatively low-yield strength nonlinear material. Hence, permanent deformations mainly occur on the solder side of the bimaterial and the situation is similar to an elastoplastic material on the rigid substrate. Here, the solder is modeled as a power-law hardening material $(\epsilon/\epsilon_o) = \alpha(\sigma/\sigma_o)^m$ (for $\sigma > \sigma_o$) using the uniaxial test data reported by Skipor et al. (1996) for a strain rate of $\dot{\epsilon} \approx 10^{-4}$ sec^{-1}. The data were fitted with a power-law equation and the resulting parameters along with other relevant mechanical properties are shown in Table 1.

The model was subjected to three-point bending as in the experiments and the values of $w(x, y)$ were computed in the solder half using material incompressibility condition. To estimate the region of plane-stress behavior on the solder side, out-of-plane displacements are plotted against r/B along a few discrete directions θ in Fig. 11. In these plots the region with disagreement exceeding ten percent between the two data sets near the crack tip is identified as the region where the three-dimensional effects are dominant. The plots also suggest that a plane-stress assumption is reasonable beyond $r/B \approx 0.5$ and 30

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

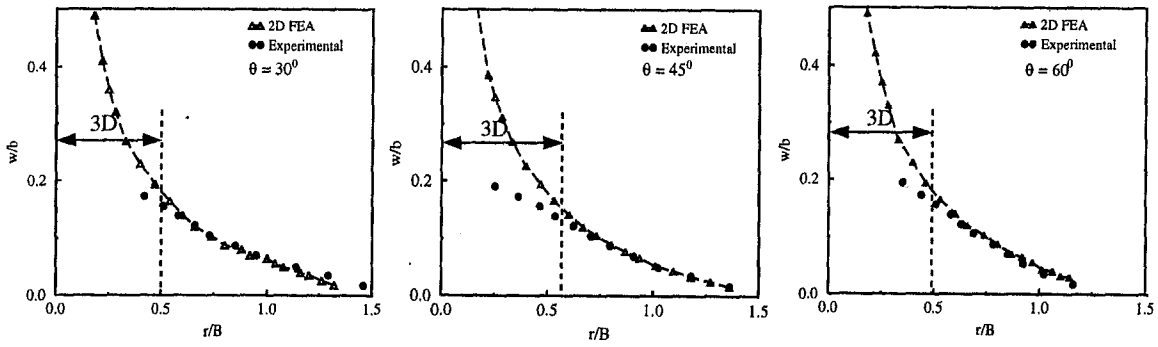


Fig. 11 Variation of out-of-plane displacements from the measurements and two-dimensional plane-stress simulations along different directions for the solder-copper bimaterial beam

deg $\leq \theta \leq 60$ deg. It should be noted that the optical data along other directions beyond $\theta = 60$ deg are rather limited for performing direct comparison. Here, we also need to recognize that three-dimensional effects dominate all along the interface ($0 < \theta < 30$ deg) and elastic counterpart investigations are reported by Lee and Rosakis (1993) and Sinha, et al. (1997).

To confirm that the disagreements in Fig. 11 are primarily due to the three-dimensional effects, an elastoplastic three-dimensional finite element analysis was carried out. The model is similar to the one described in the previous section. The contours of out-of-plane displacement $w(x, y)$ were generated from the computed results on the surface of the model. A direct superposition of experimental and finite element contours (broken lines) is shown in Fig. 12. A good agreement between the two is observed even very close to the crack tip. This is further demonstrated in Fig. 13(a) where normalized displacements from the experiments and two-dimensional and three-dimensional elastoplastic simulations are plotted against (r/B) for $\theta = 45$ deg for an applied load level of P/P_o equal to 1.65. Unlike the two-dimensional finite element results, the three-dimensional computations suggest a finite value of out-of-plane displacement at the crack tip. Moreover, the computed value of $w(x, y)$ at the crack tip is lower than the one at a small distance ahead of the crack tip and along the interface. It should be recognized that the copper half undergoes negligible out-of-plane deformation and hence copper can be viewed as a rigid substrate relative to solder. This in turn implies that the out-of-plane displacements along ($x \geq 0$, $y = 0$) are zero and correspondingly the simulations show a down turn in the out-of-plane displacement value as the crack tip is approached. The two-dimensional finite element results, however, deviate substantially from the experiments in the near vicinity of the crack predicting a singular behavior.

Although at the moment there are no explicit closed-form field solutions for a plastically deformed interfacial crack-tip region, there is evidence suggesting that for tensile dominated (or, "opening" dominated) loading, the stresses and strains are scaled by the J-integral (Shih, 1991) and bear striking resemblance to mixed-mode HRR-like fields in homogeneous media. The following form of the dominant term is suggested for strain components near an elastoplastic interfacial crack:

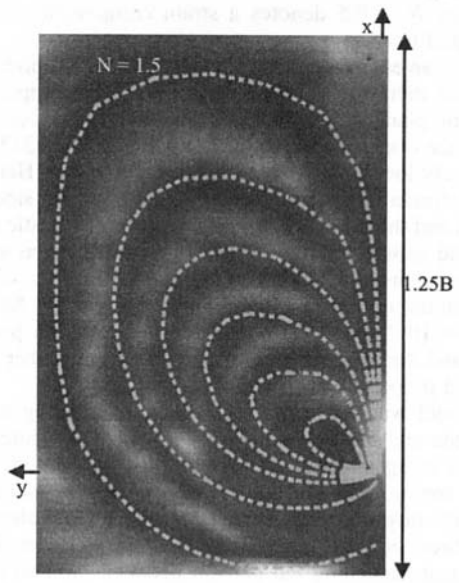


Fig. 12 Superposition of finite element contours (broken lines) from a three-dimensional simulation over the interferometric fringes for the solder-copper bimaterial beam

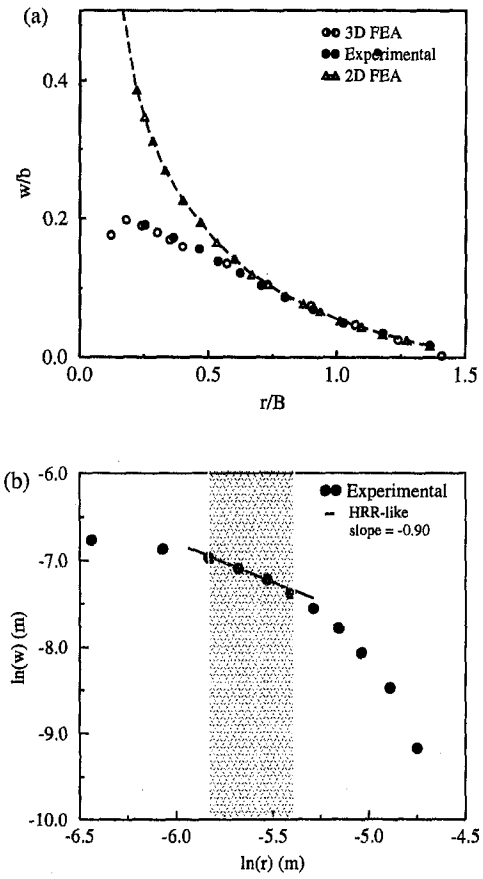


Fig. 13 Crack-tip three-dimensionality and J-dominance in solder-copper bimaterial beams ($\theta = 45$ deg); (a) plot of out-of-plane displacements, (b) comparison of optical measurements with HRR-prediction

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\epsilon_{ij} = \alpha\epsilon_o\left(\frac{J}{\alpha\epsilon_o\sigma_o I_m r}\right)^{m/(m+1)} \hat{\epsilon}_{ij}(\theta, \hat{r}, \xi, m), \qquad (20)$$

where $\hat{\epsilon}_{ij}$ is a slowly varying function of $\hat{r}$ ($=r\sigma_o/J$) and ξ is the so-called plastic mode mixity. In Eq. (20), σ_o, ϵ_o represent the yield stress and strain of the softer of the two materials. The above are deductions based on numerical simulation and direct experimental evidence of the same is currently lacking. In view of this, the experimental data is examined for evidence of a HRR-like behavior. In Fig. 13(b), log (w) is plotted against log (r) for the solder-copper interface. Knowing the hardening index m for 63/37 solder, when log (w) is plotted against log (r), a slope of $-(m/(m+1))(\approx -0.9)$ should be recoverable if a HRR-like field were to prevail in the crack-tip vicinity. In Fig. 13(b), such a behavior indeed is seen over $0.5 < r/B < 0.75$ (shaded region) and suggests J-dominance over this distance.

6 Conclusions

A novel infrared interferometer using 10.6-μm coherent infrared radiation has been developed for performing real-time full-field measurements on rough surfaces. A rigorous Fourier optics analysis that uses Fresnel diffraction principle and a Gaussian surface roughness profile is provided. The infrared interferometer is shown to be an effective tool for elastoplastic crack-tip field investigation in ductile solids. The range of surface roughness over which measurements can be effectively made is experimentally demonstrated. Fringe contrast in excess of 0.5 can be realized with surfaces with roughness parameter $\hat{\sigma} = 0.8$ μm (or, $\hat{\sigma}/\lambda \cong 0.08$) while fringe contrast in excess of 0.8 can be obtained with roughness of $\hat{\sigma} = 0.35$ μm (or, $\hat{\sigma}/\lambda \cong 0.03$). In fact, the latter is shown to be easily achievable simply by polishing the sample surface with a #600 (mean grit size 15 μm) fixed abrasive. The proposed interferometer bridges the sensitivity gap between the visible light Twyman-Green interferometer (submicron sensitivity) and shadow/projection moiré methods (25–50 μm/fringe) used for mapping out-of-plane deformations.

The region of three-dimensional deformations near elastoplastically deformed cracks in homogeneous and bimaterial beams are examined. The measurements are investigated along with companion two-dimensional and three-dimensional elastoplastic finite element simulations. In the beam configurations studied, three-dimensional effects dominate within $r/B = 0.35$ while plane-stress conditions exist beyond this region in homogeneous specimens. The optical measurements have been used to directly quantify the J-integral and the HRR predictions agree with the experiments over $0.35 < r/B < 0.65$. Experimental measurement of elastoplastic deformations near interfacial cracks are reported for solder-copper bimaterials. In the region $y > 0$, three-dimensional deformations seem to dominate over $r/B \leq 0.5$ and along the interface. In the absence of explicit elastoplastic interfacial crack-tip solutions, the existence of a region of a HRR-like deformation field is investigated and found to prevail over $0.5 \leq r/B \leq 0.75$ ahead of the crack.

Acknowledgments

Some of the equipment used in this research were acquired through NSF grant CMS9313153. Support for studying solder-copper interfaces were through AFoSR grant AF-F49620-97-10167.

References

Beckmann, P., and Spizzichino, A., 1963, *The Scattering of Electromagnetic Waves from Rough Surfaces,* Pergamon Press, New York.

Chiang, F. P., and Hareesh, T. V., 1988, "Three dimensional crack tip deformations: An experimental study and comparison to HRR field," *International Journal of Fracture,* Vol. 36, pp. 243–257.

Chiang, F. P., Hareesh, T. V., Liu, B. C., and Li, S., 1988, "Optical Analysis of HRR Field," *Optical Engineering,* Vol. 27, pp. 625–629.

Hutchinson, J. W., 1968, "Singular behavior at the end of a tensile crack tip in a hardening material," *Journal of the Mechanics and Physics of Solids,* Vol. 16, pp. 13–31.

Iizuka, K., 1985, *Engineering Optics,* 2nd Ed., Springer-Verlag, New York.

Krishnamoorthy, H., and Tippur, H. V., 1998, "Evaluation of Elasto-Plastic Interfacial Fracture Parameters in Solder-Copper Bimaterial Using Moiré Interferometry." ASME *Journal of Electronic Packaging,* Vol. 120, pp. 267–274.

Kumar, V., German, M. D., and Shih, C. F., 1981 "An engineering approach for elastic-plastic fracture analysis," EPRI Report NP-1931, Electric Power Research Institute, Palo Alto, CA.

Kwon, O., Wyant, J. C., and Hayslett, C. R., 1980, "Rough surface interferometry at 10.6 μm," *Applied Optics,* Vol. 19, pp. 1862–1869.

Lee, Y. J., and Rosakis, A. J., 1993, "Interfacial crack in plates: A three-dimensional numerical investigation," *International Journal of Solids & Structures,* Vol. 30, pp. 3139–3158.

Lewandowski, J., 1989 "Lateral shear interferometer for infrared and visible light," *Applied Optics,* Vol. 28, pp. 2373–2379.

Munnerlyn, C. R., and Latta, M., 1968, "Rough surface interferometry using CO_2 laser source," *Applied Optics,* Vol. 7, pp. 1858–1859.

Persson, U., 1993 "Roughness measurement of machined surfaces by means of the speckle technique in the visible and infrared regions," *Optical Engineering,* Vol. 32, pp. 3327–3332.

Rice, J. R., 1968 "A Path-Independent Integral and the Approximate Analysis of Strain Concentration by Notches and Cracks," ASME JOURNAL OF APPLIED MECHANICS, Vol. 35, pp. 379–386.

Rice, J. R., and Rosengren, G. F., 1968, "Plane strain deformation near a crack tip in a power law hardening material," *Journal of the Mechanics and Physics of Solids,* Vol. 16, pp. 1–12.

Shih, C. F., 1983, "Tables of Hutchinson-Rice-Rosengren Singular Fields Quantities," Materials Research Laboratory, Brown University, Providence, RI.

Shih, C. F., 1991, "Cracks in bimaterial interfaces: Elasticity and plasticity aspects," *Materials Science and Engineering,* Vol. A143, pp. 77–90.

Sinha, J. K., Tippur, H. V., and Xu, L., 1997, "An interferometric and finite element study of interfacial crack tip fields and the influence of mode-mixity on 3-D deformations," *International Journal of Solids and Structures,* Vol. 34, pp. 741–754.

Sinha, J. K., and Tippur, H. V., 1997 "Infrared Interferometry for rough surface measurements: Application to failure characterization and flaw detection," *Optical Engineering,* Vol. 36, No. 8, pp. 2233–2239.

Skipor, A. F., Harren, S. V., and Botsis, J., 1996, "On the Constitutive Response of 63/37 Sn/Pb Eutectic Solder," ASME *Journal of Engineering Materials and Technology,* Vol. 118, pp. 1–11.

Stover, J. C., 1995, *Optical Scattering,* SPIE Press, Bellingham, WA.

Timoshenko, S. P., and Goodier, J. N., 1970, *Theory of Elasticity,* 3rd Ed., McGraw-Hill, New York.

Tippur, H. V., S. Krishnaswamy, and Rosakis, A. J., 1991 "Optical Mapping of Crack Tip Deformations using the Methods of Transmission and Reflection Coherent Gradient Sensing: A Study of Crack Tip K-dominance," *International Journal of Fracture,* Vol. 52, pp. 91–117.

Zehnder, A. T., and Rosakis, A. J., 1990 "Three-Dimensional Effects Near a Crack Tip in a Ductile Three-Point Bend Specimen, Part II: An Experimental Investigation Using Interferometry and Caustics," ASME JOURNAL OF APPLIED MECHANICS, Vol. 57, pp. 618–626.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Asymptotics of Extensional Waves in Isotropic Elastic Plates

N. A. Losin
10834 N. 32nd Lane,
Phoenix, AZ 85029

The long and short-wave asymptotics of order $O(\eta^6)$, $\eta = kh$, for free extensional vibrations of an infinite isotropic elastic plate are studied. The asymptotic model for flexural and extensional wave motion applicable for both long and short-wave approximations and for any materials is developed. The velocity and frequency dispersion relations for extensional waves are derived in analytical form from the system of three-dimensional dynamic equations of linear elasticity. All dispersion equations and the group velocity formula are presented as explicit functions in material parameter $\gamma = c_S^2/c_L^2$, (the ratio of the velocities squared of the flexural and extensional waves) without any correction factors as in the Reissner-Mindlin theory. Variations of the velocity and frequency spectra depending on Poisson's ratio ν are illustrated graphically. The results are discussed and compared to those obtained and summarized by Mindlin (1960), Mindlin and Medick (1959), Tolstoy and Usdin (1953, 1957), Achenbach (1973), and Graff (1991).

1 Introduction

The problem of dispersive wave motion in infinite elastic plates has attracted the attention of many investigators for over a century. In spite of this, some interesting questions still need to be clarified. The Reference section of this paper lists just a few original works, reviews, and summaries that contain extensive bibliographies on this topic, but does not pretend to be comprehensive.

The advantages and disadvantages of the exact and approximate dynamic plate theories are well known and summarized in Achenbach (1973), Berdichevskii (1973), Graff (1991), Miklowitz (1966), Mindlin (1960), Redwood (1960), and Reissner (1985). In the exact theory, the detailed analysis of the transcendental Rayleigh-Lamb frequency equation, depending on the correlation between v, c_S, c_L in various regions, is not a simple problem. The asymptotic expansion of this equation for flexural waves, does not give an adequate approximation, even for velocities $v < c_s$.

Some popular approximate theories use well-substantiated, but still intuitive correction factors. Others are locally oriented to particular frequency ranges (short or long-wave asymptotics), but do not cover all aspects of a problem in the aggregate. These inconveniences motivate a search for simpler, adequate, and informative wave models. In a previous paper (Losin, 1997), the attempt was made to find dispersion relations for frequency and velocity spectra in analytical form as a result of a purely mathematical analysis of the elastodynamic problem for a plate. The asymptotic mathematical model of order $O(\eta^6)$ describing both flexural and extensional vibrations in plates was developed.

The study is based on the technique employed for solution of wave problems for thin-shelled structures in Protsenko (1980). Later, it was adopted for the computer-generated derivation of the propagation equations and numerical analysis of normal waves motion in cylindrical shells (Losin and Protsenko, 1984).

A few key points of this technique follows. The problem of free-wave motion in a plate is formulated as a boundary value problem for three-dimensional dynamic Navier's equations in matrix form. The presentation of the displacement vector in the form of a propagating harmonic wave turns the initial set of Navier's equations into a system of ordinary differential equations depending on several parameters including the direction vector of a traveling wave. The asymptotic nature of the problem is due to the approximate satisfaction of the boundary conditions. The stress vector at the free plate's surfaces is replaced by the partial sum of Taylor series expansion in the plate's thickness. The accuracy of results depends on the number of terms retained. The substitution of this approximation into modified boundary conditions converts them into a couple of the asymptotic matrix equations. An appropriate axes rotation simplifies the system by diagonalizing its matrix coefficients. This makes the first matrix boundary equation easily invertible, i.e., solvable for the derivative of the amplitude vector $\mathbf{u}'$ in analytical form. The exact inversion of this matrix equation, realized in Losin (1997), was extremely productive. This modification of the original technique substantially improved the accuracy of numerical results and extended the applicability of the theory. The substitution of this last expression for $\mathbf{u}'$ into the second matrix boundary condition reduces the problem to the final resolving operator equation.

The flexural vibrations of a plate were studied first. The results of this analysis were intriguing. The bicubic frequency and velocity dispersion equations, the group velocity formula, and the differential equation of the flexural waves were derived in a closed analytical form. The coefficients of all equations and formulas were obtained as explicit polynomials in material parameters γ or ν. The numerical results, i.e., the frequency and velocity spectra and some of their qualitative features, were encouraging enough in order to extend the analysis to the extensional motion of a plate.

2 Formulation

Consider free-wave motion in an isotropic plate of thickness $2h$, bounded by two planes $z = \pm h$, and infinite in the (x, y) directions. The corresponding dynamic boundary value problem of linear elasticity is formulated in matrix form

$$\{D\partial_z^2 + B_\partial \partial_z + C_\partial - \rho I \partial_t^2\}\mathbf{U} = \mathbf{0}, \quad -h < z < h, \quad (1)$$

$$\boldsymbol{\tau} = \{D\partial_z + A_\partial\}\mathbf{U} = \mathbf{0}, \quad z = \pm h, \quad (2)$$

$$B_\partial = \beta\begin{pmatrix} 0 & 0 & \partial_x \\ 0 & 0 & \partial_y \\ \partial_x & \partial_y & 0 \end{pmatrix}, \quad A_\partial = \begin{pmatrix} 0 & 0 & \mu\partial_x \\ 0 & 0 & \mu\partial_y \\ \lambda\partial_x & \lambda\partial_y & 0 \end{pmatrix},$$

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Sept. 27, 1997; final revision, Mar. 2, 1998. Associate Technical Editor: R. C. Benson.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$C_\partial = \begin{pmatrix} \alpha\partial_x^2 + \mu\partial_y^2 & \beta\partial_{xy}^2 & 0 \\ \beta\partial_{yx}^2 & \mu\partial_x^2 + \alpha\partial_y^2 & 0 \\ 0 & 0 & \mu(\partial_x^2 + \partial_y^2) \end{pmatrix},$$

$$D = \text{diag}\,\{\mu, \mu, \alpha\}, \quad B_\partial = A_\partial + A_\partial^t,$$

$$\alpha = \lambda + 2\mu, \quad \beta = \lambda + \mu.$$

Here, (1) is the system of Navier's equations written in terms of Lamé's constants λ, μ; the boundary conditions (2) express the absence of stresses at the free plate's surfaces $z = \pm h$; $\mathbf{U} = \mathbf{U}(x, y, z, t) = \{U_x, U_y, U_z\}^t$ is the displacement vector; $\boldsymbol{\tau} = \{\tau_{xz}, \tau_{yz}, \tau_{zz}\}^t$ is the stress vector at $z = \pm h$; ρ is the density of the plate; I is the unit matrix.

We seek solutions in the form of an harmonic wave

$$\mathbf{U}(x, y, z, t) = \mathbf{u}(z)\exp\{-i(k_x x + k_y y - \omega t)\}, \quad (3)$$

where $\mathbf{u}(z) = \{u_x(z), u_y(z), u_z(z)\}^t$ is the amplitude vector; $\omega = \omega(\mathbf{k})$ is the circular frequency (dispersion relation); $\mathbf{k} = \{k_x, k_y\}$ is the wave vector, $k = |\mathbf{k}|$. Differentiating (3) with respect to x, y, t, and replacing partial derivatives by their symbols

$$\partial_x \to -ik_x, \quad \partial_y \to -ik_y, \quad \partial_x^2 \to -k_x^2, \quad \partial_y^2 \to -k_y^2,$$

$$\partial_t \to i\omega, \quad \partial_t^2 \to -\omega^2, \quad \nabla^2 \to -k^2, \quad \nabla^4 \to k^4 \quad (4)$$

we may reduce (1) and (2) to the form

$$D\mathbf{u}'' - B_\mathbf{n}(k)\mathbf{u}' - C_\mathbf{n}(k)\mathbf{u} = \mathbf{0}, \quad -h < z < h, \quad (5)$$

$$\boldsymbol{\tau}(z, k) = D\mathbf{u}' - A_\mathbf{n}(k)\mathbf{u} = \mathbf{0}, \quad z = \pm h, \quad (6)$$

$$B_\mathbf{n}(k) = ik\beta\begin{pmatrix} 0 & 0 & n_1 \\ 0 & 0 & n_2 \\ n_1 & n_2 & 0 \end{pmatrix},$$

$$A_\mathbf{n}(k) = ik\begin{pmatrix} 0 & 0 & \mu n_1 \\ 0 & 0 & \mu n_2 \\ \lambda n_1 & \lambda n_2 & 0 \end{pmatrix},$$

$$C_\mathbf{n}(k) = k^2\begin{pmatrix} \alpha n_1^2 + \mu n_2^2 - \rho v^2 & \beta n_1 n_2 & 0 \\ \beta n_2 n_1 & \mu n_1^2 + \alpha n_2^2 - \rho v^2 & 0 \\ 0 & 0 & \mu(n_1^2 + n_2^2) - \rho v^2 \end{pmatrix},$$

$$\mathbf{u}' \equiv d\mathbf{u}(z)/dz, \quad \mathbf{n} = \{n_1, n_2\}, \quad n_1 = k_x/k,$$

$$n_2 = k_y/k, \quad \mathbf{k} = k\mathbf{n},$$

$$v(\mathbf{k}) = \omega(\mathbf{k})/k, \quad v(\mathbf{k}) = |\mathbf{v}(\mathbf{k})|, \quad (7)$$

where $\mathbf{v}$ is the phase velocity and v is the phase speed of a traveling wave; $\mathbf{k}$ is the wave vector; $\mathbf{n}$ is the unit direction vector; n_1, n_2 are direction cosines of $\mathbf{n}$.

Then, we multiply (5), (6) by the diagonal matrix $D^{-1} = \text{diag}\,\{\mu^{-1}, \mu^{-1}, \alpha^{-1}\}$ from the left, eliminate the Lamé's constants λ, μ, and express matrix coefficients through the unique parameter γ, the ratio of the velocities squared of the shear and extensional waves. Introducing the modified stress vector $\boldsymbol{\tau}_c$, we finally formulate the boundary value problem as the following system of ordinary differential equations:

$$\mathbf{u}'' = kB\mathbf{u}' + k^2 C\mathbf{u}, \quad -h < z < h, \quad (8)$$

$$\boldsymbol{\tau}_c(z, k) = \rho(\mathbf{u}' - kA\mathbf{u}) = \mathbf{0}, \quad z = \pm h, \quad (9)$$

$$B = i\begin{pmatrix} 0 & 0 & (\gamma^{-1} - 1)n_1 \\ 0 & 0 & (\gamma^{-1} - 1)n_2 \\ (1 - \gamma)n_1 & (1 - \gamma)n_2 & 0 \end{pmatrix},$$

$$A = i\begin{pmatrix} 0 & 0 & n_1 \\ 0 & 0 & n_2 \\ (1 - 2\gamma)n_1 & (1 - 2\gamma)n_2 & 0 \end{pmatrix},$$

$$C = \begin{pmatrix} \gamma^{-1}n_1^2 + n_2^2 - v_s^2 & (\gamma^{-1} - 1)n_1 n_2 & 0 \\ (\gamma^{-1} - 1)n_2 n_1 & n_1^2 + \gamma^{-1}n_2^2 - v_s^2 & 0 \\ 0 & 0 & \gamma(1 - v_s^2) \end{pmatrix},$$

where $\boldsymbol{\tau}_c(z, k) = \rho D^{-1}\boldsymbol{\tau}(z, k) = \{\tau_{xz}/c_s^2, \tau_{yz}/c_s^2, \tau_{zz}/c_L^2\}^t$ is the modified stress vector; c_s, c_L are velocities (speeds) of the shear and extensional waves, respectively,

$$\gamma = c_s^2/c_L^2 = \mu/(\lambda + 2\mu), \quad c_s^2 = \mu/\rho,$$

$$c_L^2 = (\lambda + 2\mu)/\rho, \quad v_s = v/c_s, \quad v_L = v/c_L.$$

3 Resolving Operator Equation

The corresponding asymptotic boundary value problem can be now formulated and reduced to the system of three propagation equations by applying the technique briefly mentioned in the Introduction. An interested reader is referred to paragraphs 3 and 4 of Losin (1997) for details.

The final resolving operator has the form

$$T\mathbf{u}(0) = O(\eta^6), \quad T = T_0 + \eta^2/6T_2 + \eta^4/120T_4, \quad (10)$$

which is analogous to Eq. (20) of the previous article where the expressions for T_0, T_2, T_4 are specified.

In the general case, the matrix $T = (t_{ij})$, $(i, j = 1, 2, 3)$, of the operator T has a block diagonal structure $T = \text{diag}\,\{T_L, T_S\}$, where T_S and T_L are blocks governing flexural and in-plane motion, respectively. In our particular case, the axes rotation reduces it to the diagonal form $T = \text{diag}\,\{t_{11}, t_{22}, t_{33}\}$, $T_L = \text{diag}\,\{t_{11}, t_{22}\}$, $T_S = t_{33}$, where

$$t_{jj} = k^2(t_{jj}^{(0)} + \eta^2/6t_{jj}^{(2)} + \eta^4/120t_{jj}^{(4)}), \quad (j = 1, 2, 3),$$

$$t_{11} = -k^2 m_{33}^{-1}(v_s^2 - c_{p_s}^2 - \eta^2/6P_4 + \eta^4/120Q_6), \quad (11)$$

$$t_{22} = k^2(1 - v_s^2)[1 + \eta^2/6(1 - v_s^2) + \eta^4/120(1 - v_s^2)^2].$$

$$P_4(v_s, \gamma) = p_0 v_s^4 - p_1 v_s^2 + p_2,$$

$$Q_6(v_s, \gamma) = q_0 v_s^6 - q_1 v_s^4 + q_2 v_s^2 - q_3,$$

$$p_0(\gamma) = 1 + 3\gamma, \quad p_1(\gamma) = 4(2 - \gamma^2), \quad p_2(\gamma) = 8(1 - \gamma),$$

$$q_0(\gamma) = 1 + 10\gamma + 5\gamma^2, \quad q_1(\gamma) = 4(4 + 10\gamma - 5\gamma^2 - \gamma^3),$$

$$q_2(\gamma) = 16(3 - 2\gamma^2), \quad q_3(\gamma) = 32(1 - \gamma),$$

$$c_{p_s}^2 = 4(1 - \gamma), \quad (12)$$

and $c_{p_s}^2 = c_p^2/c_s^2$. The element t_{33} and its components $t_{33}^{(0)}$, $t_{33}^{(2)}$, $t_{33}^{(4)}$ are specified in Losin (1997, Eq. (27)).

We can now approximate (10) by the linear homogeneous algebraic system $T\mathbf{u}(0) = \mathbf{0}$ with the asymptotic accuracy $O(\eta^6)$, assuming that $k^2 \neq 0$. The nontrivial solutions exist if and only if $\det T = \det T_L \cdot \det T_S = 0$. This equation is the three-dimensional analog of the Rayleigh-Lamb frequency equation for a plate that generates two dispersion relations. One of them, $\det T_S = 0$, or, $t_{33} = 0$, is the velocity equation for the flexural vibrations that were already analyzed in the previous paper.

4 Extensional Motion of a Plate

The second relation is the velocity equation for the extensional motion of a plate

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\det T_L = 0, \quad \text{or} \quad t_{11}t_{22} = 0. \tag{13}$$

The diagonal structure of the matrix $T = \text{diag}\{t_{11}, t_{22}, t_{33}\}$ implies that the corresponding diagonalized system of the propagation equations has the form $t_{jj}U_j = 0$, $(j = 1, 2, 3)$, where each equation affects only one of the displacement components $U_1 \equiv u$, $U_2 \equiv v$, $U_3 \equiv w$. All three equations of this system are independent, and can be analyzed separately. Therefore, we interpret the first equation $t_{11}u = 0$ as a propagation equation that governs longitudinal wave motion along the OX-axis, but the equation $t_{11} = 0$, that follows from (13), is a corresponding dispersion relation. They generate the frequency, velocity, and the differential equation of a plate's extensional motion. According to (11), the dispersion relation has the form

$$v_s^2 - c_{p_s}^2 - \eta^2/6(p_0v_s^4 - p_1v_s^2 + p_2) + \eta^4/120(q_0v_s^6 - q_1v_s^4 + q_2v_s^2 - q_3) = 0. \tag{14}$$

Equation (14) is the asymptotic form of the velocity equation for extensional vibrations, which may be transformed into the polynomial form

$$q_0v_s^6 - (q_1 + 20p_0/\eta^2)v_s^4 + (q_2 + 20p_1/\eta^2 + 120/\eta^4)v_s^2 - (q_3 + 20p_2/\eta^2 + 120c_{p_s}^2/\eta^4) = 0. \tag{15}$$

Equation (15) is the final velocity equation for extensional motion of a plate in terms of the dimensionless phase velocity $v_s = v/c_s$. The substitution $v_s = \omega_s/k$ turns Eq. (15) into the frequency equation

$$q_0\omega_s^6 - (20p_0 + q_1\eta^2)\omega_s^4 + (120 + 20p_1\eta^2 + q_2\eta^4)\omega_s^2 - (120c_{p_s}^2\eta^2 + 20p_2\eta^4 + q_3\eta^6) = 0. \tag{16}$$

Equation (16) is the frequency equation for extensional motion of a plate with respect to the dimensionless frequency $\omega_s = \omega h/c_s$. The backward substitution of symbols (4) by the corresponding partial derivatives generates the differential equation of extensional vibrations

$$\{\hat{\partial}_t^2 - c_{p_s}^2\nabla^2 + h^2/6(p_0\hat{\partial}_t^4 - p_1\hat{\partial}_t^2\nabla^2 + p_2\nabla^4) + h^4/120 \times (q_0\hat{\partial}_t^6 - q_1\hat{\partial}_t^4\nabla^2 + q_2\hat{\partial}_t^2\nabla^4 - q_3\nabla^6)\}u = 0, \tag{17}$$

where $\hat{\partial}_t = c_s^{-1}\partial_t$.

The second equation of the diagonalized system, $t_{22}v = 0$, governs the in-plane wave motion in the same direction and affects the v component of the displacement vector. Its velocity equation $t_{22} = 0$ is satisfied for $1 - v_s^2 = 0$ or $v^2 = c_s^2$, which is the unique real root of multiplicity one for long-wave, and of multiplicity three for short-wave asymptotics. This root corresponds to the nondispersive propagating flexural wave polarized in the XOY-plane. No more propagating modes are available in this case due to the remaining complex-conjugate roots of the velocity equation.

5 Velocity and Frequency Spectrums

The phase velocities (speeds) of a propagating wave are roots of the bicubic Eq. (15). Its coefficients depend on parameters $\eta = kh$ and γ, (or ν). The solutions of (15) give the phase velocities of three wave modes (in the general case) as functions of kh and γ. The graphs of velocity spectra for various Poisson's ratios are shown in Figs. 1–3. The wave motion under consideration is dispersive and the dispersion relation is given by Eq. (16). Its roots represent the frequency spectrum of vibrations depending on kh and γ. Differentiating Eq. (16) with respect to k, where $\omega_s = \omega_s(k)$, we obtain the following expression for the group velocity $V_g(k) = d\omega(k)/dk$ of the extensional motion in terms of $v_s(k) = \omega_s(k)/k$:

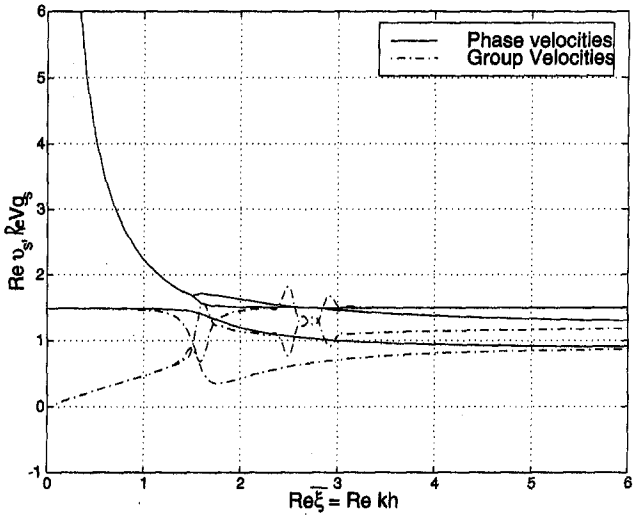


Fig. 1 Phase and group velocities of extensional modes for $\nu = 0.1$ ($\gamma = 0.4444$)

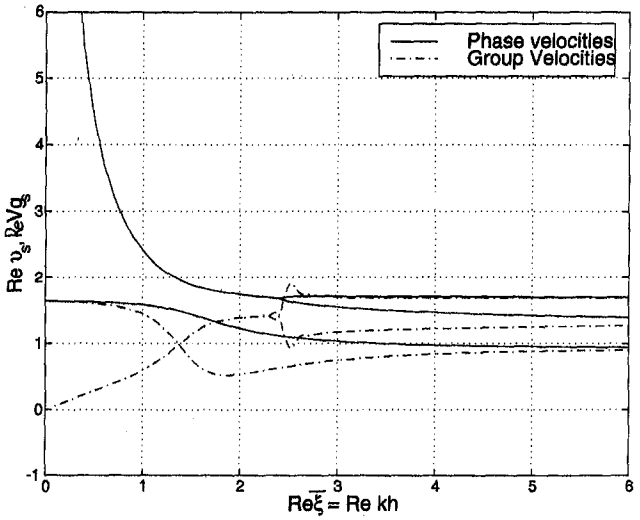


Fig. 2 Phase and group velocities of extensional modes for $\nu = \frac{1}{4}$ ($\gamma = \frac{1}{3}$)

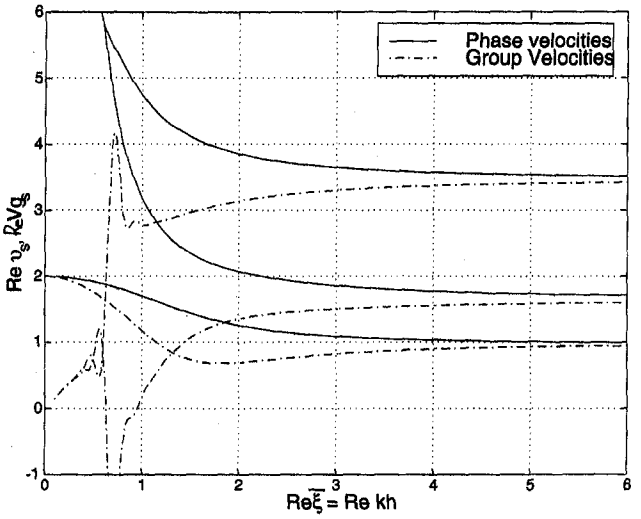


Fig. 3 Phase and group velocities of extensional modes for $\nu = 0.49$ ($\gamma = 0.0196$)

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

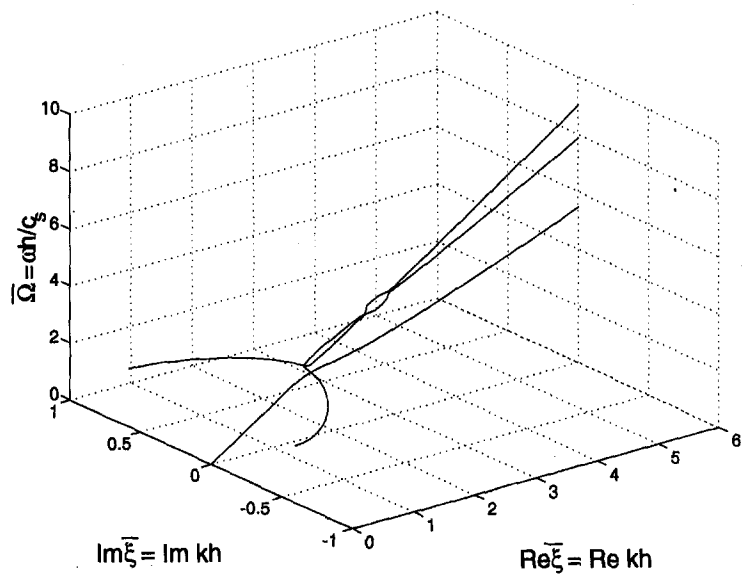


Fig. 4 Frequency spectrum of extensional modes for $\nu = 0.1$ ($\gamma = 0.4444$)

$$V_{g_s} = \frac{q_1 v_s^4 - 2v_s^2(q_2 + 10p_1/\eta^2) + 3q_3 + 40p_2/\eta^2 + 120c_{p_s}^2/\eta^4}{3q_0 v_s^5 - 2v_s^3(q_1 + 20p_0/\eta^2) + v_s(q_2 + 20p_1/\eta^2 + 120/\eta^4)}. \tag{18}$$

The group velocity and frequency spectra are shown in Figs. 1–3 and Figs. 4–6, respectively.

6 Long and Short-Wave Approximations

Both long and short-wave approximations are available from Eqs. (14)–(16). For long waves, the wavelength is very large compared to the thickness $2h$ of a plate, i.e., $\eta = kh \to 0$. The limiting form of Eq. (16) is

$$\omega_s^2(q_0\omega_s^4 - 20p_0\omega_s^2 + 120) = 0. \tag{19}$$

The trivial solution $\omega_s^2 = 0$ corresponds to the phase velocity of the fundamental mode $v_s^2 = c_{p_s}^2 = 4(1 - \gamma)$ that follows from Eq. (14) as $\eta = kh \to 0$. The biquadratic equation in (19), with coefficients depending on γ, gives two more roots in the general case. They are a pair of complex-conjugate roots in all the physical domain of Poisson's ratio $\nu \in [0, \frac{1}{2}]$ and for the thickness range $2h \in [\frac{1}{20}, \frac{1}{10}]$. The corresponding phase velocities are identical, since $v = \mathrm{Re}(\omega/k)$, and tend to infinity in Figs. 1–3.

For the short-wave approximation, the wavelength is very small compared to the thickness of a plate. The substitution of $\eta = kh \to \infty$ into (15) gives the limiting form of the velocity equation

$$q_0 v_s^6 - q_1 v_s^4 + q_2 v_s^2 - q_3 = 0. \tag{20}$$

The roots of this bicubic equation are phase velocities of short-wave modes depending only on the material parameter γ. It has three finite real roots that correspond to the velocities of the first three wave modes. The fundamental mode tends to the velocity of Rayleigh's surface wave c_R. The second mode approaches the velocity c_L of the longitudinal wave. The limit of the first mode is found below that of the longitudinal thin plate velocity c_p.

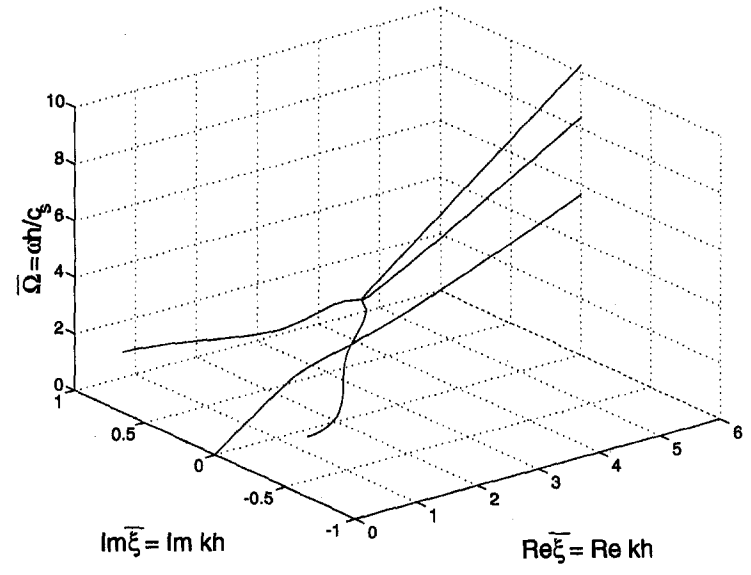


Fig. 5 Frequency spectrum of extensional modes for $\nu = \frac{1}{4}$ ($\gamma = \frac{1}{3}$)

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

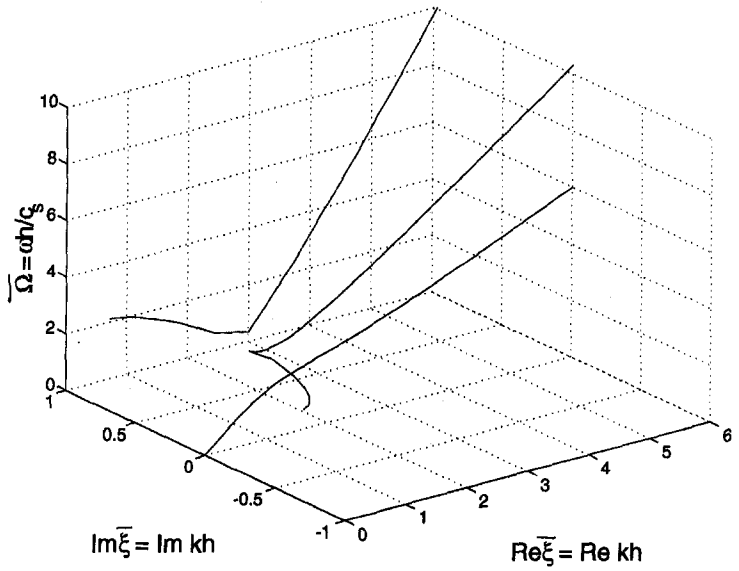


Fig. 6 Frequency spectrum of extensional modes for $\nu = \frac{1}{2}$ ($\gamma = 0$)

7 Discussion of Results

1 Equation (17) is the asymptotic approximation of order $O(\eta^6)$ to the differential equation for the extensional motion of a plate. All coefficients of the differential operator are presented as explicit functions (12) of the material parameter γ or Poisson's ratio ν, $\gamma = (1 - 2\nu)/(2(1 - \nu))$, without any correction factors. Equations (14)–(16) are velocity and frequency dispersion relations for longitudinal vibrations of a plate that are valid for long and short-wave asymptotics and for any materials. Comparing Eq. (17) and Eq. (31) in Losin (1997), we can see that both of them have a similar structure of the form

$$\{D_i + \eta^2/3!(p_{i2}\hat{\partial}_t^2 - q_{i2}\nabla^2)^2 + \eta^4/5!(p_{i3}\hat{\partial}_t^2 - q_{i3}\nabla^2)^3 + \ldots\}U_i = 0, \quad (i = 1, 3). \quad (21)$$

$U_1 = u$, $U_3 = w$, $D_1 = \hat{\partial}_t^2 - c_{p_s}^2\nabla^2$, $D_3 = \hat{\partial}_t^2$, $\hat{\partial}_t = c_s^{-1}\partial_t$, $c_{p_s} = c_p/c_s$, $p_{im}(\gamma)$, $q_{im}(\gamma)$ are polynomials in γ.

2 The velocity and frequency spectra, depending on Poisson's ratio, are shown in Figs. 1–3 and Figs. 4–6, respectively, in dimensionless coordinates

$$\bar{\xi} = \pi\xi/2, \quad \xi = 2kh/\pi, \quad \bar{\Omega} = \pi\Omega/2, \quad \Omega = 2\omega h/(\pi c_s).$$

The graphs of the phase and group velocities of the fundamental mode are approximated well by the real-valued curves (Figs. 1–3). They look similar to the corresponding graphs of Fig. 3 (Tolstoy and Usdin, 1953) and of Fig. 1 (Tolstoy and Usdin, 1957), have the common limit c_p at low frequencies, and tend to the velocity of Rayleigh surface waves c_R as $kh \to \infty$. The graph of the fundamental frequency mode is always a real-valued curve (propagating mode) that is entirely located in the real coordinate plane XOZ (Figs. 4–6). Its relative approximation error is $|e_r| < 1.5$ percent for $kh \leq 2$, and $|e_r| < 3.5$ percent for $kh \leq 6$, comparing with Fig. 8.27*a*, p. 510, Graff (1991), and Fig. 6.10, p. 231, Achenbach (1973), respectively, ($\nu = \frac{1}{4}$).

The spectra of the first and second modes are approximated reasonably well for short and intermediate wavelengths. One of these modes may be associated with some antisymmetric motion of a plate, as in Mindlin (1960), Achenbach (1973), and Graff (1991).

3 The comparison of results for long waves reveals some qualitative distinctions. In Achenbach (1973, Fig. 6.10, p. 231), the graphs of the first and second modes are completely real-valued curves with different cut-off frequencies due to the three real positive roots of the frequency equation. In Mindlin (1960, Fig. 23, p. 219), only two curves of the fundamental and second modes are entirely real-valued. In our case (Figs. 4–6), only the graph of the fundamental mode is completely real-valued (propagating mode), while the first and second modes are partially complex-valued branches (standing waves). This is the case of one real and two complex-conjugate roots, which is also valid according to Decartes' sign rule (Graff, 1991, p. 508). In Figs. 1–3, the coupling effect of these two modes is due to the equal real parts of two complex-conjugate roots of the velocity equation. This phenomenon is better illustrated by three dimensional pictures of the corresponding frequency spectra in Figs. 4–6, where two complex-conjugate branches (standing waves) are symmetrical with respect to the real XOZ plane.

4 The portion of the L(2) branch between cut-off frequency and the minimum point was referred in Graff (1991, p. 455), as a puzzling feature of the spectrum, since the negative group velocity was detected. In our studies of both extensional and flexural waves, the negative group velocity was found only in relatively small neighborhoods of some branch points at low frequencies, and seems to be local in nature due to the bifurcations. One more interesting moment is the branch interchange for the first and second modes described in Mindlin and Medick (1959) and mentioned in Graff (1991). This phenomenon is observed in both spectra over Poisson's ratio interval $\nu \in [0, 0.2]$. It is shown in Fig. 4 for frequency, and in Fig. 1 for velocity spectra, where it generates an elegant local interlacing of the corresponding group velocity branches.

5 The basic distinctions in spectrum behavior of the first and second modes, described in Mindlin (1960), Achenbach (1973), and Graff (1991), were surprisingly found for long-wave approximation, where the theory is substantiated the best ($kh \to 0$). This fact seems all the more strange, since the model is obviously sensitive enough, exactly at long waves, to simulate such nontrivial qualitative phenomena as the mode coupling, bifurcation, branch interchange and interlacing, and a negative group velocity. Conversely, the numerical results at short waves ($kh \to \infty$), that have to be affected by a greater approximation error, look acceptable.

6 There is a significant difference in spectra behavior of longitudinal and flexural waves (Losin, 1997). All figures for extensional motion have only one bifurcation point which di-

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

vides the domain into two different parts. In the left portion, the graphs of the first and the second modes coincide, while they bifurcate into two separate branches in the right part. The position of the branch point depends on Poisson's ratio. This qualitative picture is typical for longitudinal waves and holds for any materials in contrast to the flexural waves where the situation is more complicated.

For flexural motion, zero, one, or two bifurcation points were found in spectra essentially depending on the material parameter. The three flexural modes can coexist without any intersections and branching for materials with $\nu \geq 0.37$. In case of one branch point, the situation is similar to the extensional vibrations. Two bifurcation points divide the domain into three intervals, where the first and second modes are separated at low and high frequencies, but coincide at the intermediate wavelengths.

Conclusion

An asymptotic model applicable for both extensional and flexural motion, long and short-wave approximations, and for any materials is developed. The bicubic dispersion relations for frequency and velocity spectra and the group velocity formula are derived. The coefficients of all equations and formulas are presented as explicit polynomials in γ or ν without any correction factors as in the Mindlin-Reissner theory.

In the general case, the wave motion in a plate is three-dimensional since three waves can propagate in a certain direction (for example, along the OX-axis). Two of them, the longitudinal wave and the flexural wave polarized in the XOZ-plane, are dispersive. The third flexural wave, polarized in the plate's undeformed midplane XOY, is not dispersive.

The phase, group velocities, and the frequency of the fundamental extensional mode are approximated well for all wavelengths. The spectra of the first and second modes are approximated reasonably well at the intermediate and short waves. The numerical results of the long-wave asymptotics are valid theoretically, but do not match those predicted by the exact theory. In particular, the real values of cut-off frequencies for the first and second extensional modes were not observed.

The mathematical model developed is informative. It does simulate such high-order effects as mode coupling, bifurcation, branch interchange and interlacing, and a negative group velocity. It also prompts a general structure of the propagation equations for extensional and flexural motion of a plate with any order of the asymptotic accuracy.

The major problems of the model are the results verification for long-wave asymptotics, and the evaluation of the approximation error for short-wave asymptotics.

Acknowledgment

The author is grateful to Prof. A. M. Protsenko, Moscow, Russia, for the statement of the problem and initial guidance; and to MathWorks, Inc., for such an excellent research environment as the MATLAB is.

References

Achenbach, J. D., 1973, *Wave Propagation in Elastic Solids,* North-Holland, Amsterdam.

Bache, T. C., and Hegemier, G. A., 1974, "On Higher-Order Elastodynamic Plate Theories," ASME JOURNAL OF APPLIED MECHANICS, Vol. 41, pp. 423–428.

Bedford, A., and Drumheller, D. S., 1994, *Introduction to Elastic Wave Propagation,* John Wiley and Sons, New York.

Bland, D. R., 1988, *Wave Theory and Applications,* Oxford University Press, New York.

Berdichevskii, V. L., 1973, "Dynamic Theory of Thin Elastic Plates," *Izvestia AN SSSR, Mechanics of Solids,* Vol. 8, No. 6, pp. 99–109.

Brekhovskikh, L. M., and Goncharov, V., 1994, *Mechanics of Continua and Wave Dynamics,* Springer-Verlag, New York.

Davis, J. L., 1988, *Wave Propagation in Solids and Fluids,* Springer-Verlag, New York.

Doyle, J. F., 1989, *Wave Propagation in Structures,* Springer-Verlag, New York.

Eringen, A. C., and Şuhubi, E. S., 1975, *Elastodynamics,* Vol. 2, Academic Press, New York.

Ewing, W. M., Jardetzky, W. S., and Press, F., 1957, *Elastic Waves in Layered Media,* McGraw-Hill, New York.

Goldenveizer, A. L., Kaplunov, J. D., and Nolde, E. V., 1993, "On Timoshenko-Reissner Type Theories of Plates and Shells," *International Journal of Solids and Structures,* Vol. 30, No. 5, pp. 675–694.

Graff, K. F., 1991, *Wave Motion in Elastic Solids,* Dover, New York.

Losin, N. A., 1997, "Asymptotics of Flexural Waves in Isotropic Elastic Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 336–342.

Losin, N. A., and Protsenko, A. M., 1984, "Spectrum of Normal Waves in a Cylindrical Shell," *Izvestia AN SSSR, Mechanics of Solids,* Vol. 19, No. 3, pp. 186–189.

Miklowitz, J., 1966, "Elastic Wave Propagation," *Applied Mechanics Surveys,* H. N. Abramson et al., eds., Spartan Books, Washington, pp. 809–839.

Miklowitz, J., 1978, *The Theory of Elastic Waves and Waveguides,* North-Holland, Amsterdam.

Mindlin, R. D., 1960, "Waves and Vibrations in Isotropic, Elastic Plates," *Structural Mechanics,* J. N. Goodier, and N. J. Hoff, eds., Pergamon Press, New York, pp. 199–232.

Mindlin, R. D., and Medick, M. A., 1959, "Extensional Vibrations of Elastic Plates," ASME JOURNAL OF APPLIED MECHANICS, Vol. 26, pp. 561–569.

Protsenko, A. M., 1980, "Asymptotics of Wave Problems for a Cylindrical Shell," *Izvestia AN SSSR, Applied Mathematics and Mechanics,* Vol. 44, No. 3, pp. 507–515.

Redwood, M., 1960, *Mechanical Waveguides,* Pergamon Press, New York.

Reissner, E., 1985, "Reflections on the Theory of Elastic Plates," ASME *Applied Mechanics Reviews,* Vol. 38, pp. 1453–1464.

Tolstoy, I., and Usdin, E., 1953, "Dispersive Properties of Stratified Elastic and Liquid Media: A Ray Theory," *Geophysics,* Vol. 18, pp. 844–870.

Tolstoy, I., and Usdin, E., 1957, "Wave Propagation in Elastic Plates: Low and High Mode Dispersion," *Journal of the Acoustical Society of America (J.A.S.A.),* Vol. 29, No. 1, pp. 37–42.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Three-Dimensional Interface Cracks in Anisotropic Bimaterials: The Non-Oscillatory Case

Jianmin Qu
Mem. ASME

Yibin Xue

G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405

Two-dimensional interface cracks in anisotropic bimaterials have been studied extensively in the literature. However, solutions to three-dimensional interface cracks in anisotropic bimaterials are not available. In this paper, a penny-shaped crack on the interface between two anisotropic elastic half-spaces is considered. A formal solution is obtained by using the Stroh method in two-dimensional elasticity in conjunction with the Fourier transform method. Fracture mechanics parameters such as the stress intensity factor, crack-opening displacement, and energy release rate are obtained in terms of the interfacial matrix $\mathbf{M}$. *To illustrate the solution procedure, a circular delaminations in a unidirectional and a cross-ply composite are considered. Numerical results for the stress intensity factors and energy release rate along the crack front are presented.*

1 Introduction

Interface crack problems have been studied extensively. For example, just in the past decade, numerous papers on two-dimensional interface cracks in isotropic bimaterials have appeared in the literature (e.g., Hutchinson et al., 1987; Rice, 1988, etc, just to name a few). Two-dimensional interface cracks in anisotropic solids also have attracted much attention. The generalized plane-strain problem of a crack on the interface between two generally anisotropic elastic solids has been considered by Clements (1971), Willis (1971a), Ting (1986), Qu and Bassani (1989), Bassani and Qu (1989), Suo (1990), Qu and Li (1991), and Qu and Bassani (1993).

In comparison with the two-dimensional problem, very few solutions are available for three-dimensional interface crack problems. A solution for a penny-shaped interface crack in an isotropic bimaterial under uniform normal loading was given by Mossakovskii and Rykba (1964). The axiysymmetric problem of a penny-shaped crack on the interface between an isotropic solid and a rigid substrate was given by Keer (1967). When general polynomial loads are applied to the crack surface, Willis (1972) presented the solution to a penny-shaped crack on the interface between two dissimilar isotropic solids. To the author's knowledge, solutions to three-dimensional interface cracks in anisotropic bimaterials are not available in the literature.

The focus of this paper is to formulate a general approach to solve the problems of three-dimensional interface cracks in anisotropic solids. In addition to their scientific value, solutions to such problems have important applications in various engineering fields. Grain boundary fracture in polycrystalline materials, delamination in fiber-reinforced composite laminates, and thin film decohesion from silicon wafers in semiconductor fabrications are just a few examples of industrial applications for the solutions to this problem.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS.

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.

Manuscript received by the ASME Applied Mechanics Division, Oct. 14, 1997; final revision, June 18, 1998. Associate Technical Editor: J. W. Ju.

The general approach developed here combines the Fourier transform method used by Willis (1967, 1968, 1971b, 1972) and the Stroh method used by Ting (1996). The derivation follows their formulations closely. As pointed out by Ting (1996) the Stroh method, originally developed for two-dimensional anisotropic elasticity, has the potential to be extended to three-dimensional anisotropic elasticity analysis. The present study is another example of how the Stroh method can be extended to three-dimensional problems.

In using the two-dimensional Stroh method, care must be taken when the elastic stiffness tensor degenerates into an isotropic tensor (Ting and Hwu, 1988). However, it seems from the present study that degenerating to isotropic solids is not a problem at all when the Stroh method is applied to three-dimensional problems. The formal solution derived using the general anisotropic stiffness matrix can be reduced to solutions for isotropic materials by simply substituting the isotropic elastic constants into the formal solution. Taking advantage of this, some well-known solutions for penny-shaped cracks in isotropic solids can be easily reproduced by the present method. A few examples of such solutions are presented here to illustrate how simple and direct the present method is in comparison to the method used originally to derive these solutions.

It is well known that the linear elastic solutions of two-dimensional interface cracks behave pathologically near the crack tip; namely, the crack faces come into contact, and the singular crack-tip fields are oscillatory (Rice, 1988). The extent to which such pathological behaviors exist depends on the mismatch parameter $\beta = [-0.5 \text{ tr } [\mathbf{D}^{-1}\mathbf{W}]^2]^{1/2}$ which reduces to the second Dundurs constant in the isotropic case (Qu and Li, 1991). Similar pathological behaviors also exist in the solutions to three-dimensional interface cracks (Willis, 1972). It can be shown that the extent to which the pathological behavior exists depends on a mismatch parameter similar to that in the two-dimensional case (Xue, 1998). For many cases of practical interest, the mismatch parameter is rather small, so the pathological behavior exists only within a very small region near the crack front. For practical applications, one may simply ignore the mismatch parameter and treat it as zero. In this paper, the formulation is derived based on the general case. The integral equation, however, is solved only for the case when the mismatch parameter $\beta(\mathbf{n}) = [-0.5 \text{ tr } [\mathbf{D}^{-1}(\mathbf{n})\mathbf{W}(\mathbf{n})]^2]^{1/2}$ is zero.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Solutions to the nonzero $\beta(\mathbf{n})$ are presented elsewhere (Xue, 1998).

In Section 2 of this paper, the general three-dimensional planar interface crack problem is stated as a boundary value problem. General equations in the three-dimensional Stroh method are derived in Section 3. By using these equations, the boundary value problem stated in Section 2 is formulated into an integral equation for the crack-opening displacement. Up to this point, the formulation is general in that they are valid for no-zero $\beta(\mathbf{n})$ as well. Starting from Section 5, it is assumed that $\beta(\mathbf{n}) = 0$. The integral equation derived in Section 4 is solved in Section 5 based on this assumption and the crack-opening displacement is obtained. Fracture parameters such as the stress intensity factors and energy release rate are obtained in Section 6 in terms of the crack-opening displacement. In Section 7, several numerical examples are presented to illustrate the solution procedure.

2 Problem Statement

Consider an anisotropic bimaterial of infinite extent. Let a Cartesian coordinate (x_1, x_2, x_3) be such that the bimaterial interface is at $x_3 = 0$. The two material half-spaces are labeled material 1 for $x_3 > 0$, and material 2 for $x_3 < 0$. Let a planar crack be located on the bimaterial interface occupying the region R.

It is assumed that both elastic half-spaces are linear elastic. They are perfectly bonded together on the bimaterial interface except the cracked region R. Furthermore, it is assumed that traction vectors of equal amplitude and opposite sign are prescribed on the upper and lower surfaces of the interfacial crack, respectively. For traction-free interfacial cracks under remote loading, the solution can be obtained from the one discussed below through a superposition with proper T-stresses added (Bassani and Qu, 1990; Qu and Bassani, 1993).

Under the above conditions, the governing equations for the displacement and stress fields are

$$\sigma_{ij,j} = 0, \tag{2.1}$$

where

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl} = C_{ijkl}u_{k,l}, \tag{2.2}$$

and C_{ijkl} is the fourth-order stiffness tensor of the bimaterial represented by

$$C_{ijkl} = \begin{cases} C^{(1)}_{ijkl}, & x_3 > 0, \\ C^{(2)}_{ijkl}, & x_3 < 0. \end{cases} \tag{2.3}$$

For future reference, all indicial notations in this paper obey the summation convention unless otherwise noted. Also, lower case letters are used to represent vectors and scalars, while capital letters are used for higher-order tensors.

The proper boundary conditions on the interface $x_3 = 0$ are

$$\sigma_{i3}(x_1, x_2, 0^+) = \sigma_{i3}(x_1, x_2, 0^-), \quad \text{for} \quad (x_1, x_2) \notin R \tag{2.4a}$$

$$u_i(x_1, x_2, 0^+) = u_i(x_1, x_2, 0^-), \quad \text{for} \quad (x_1, x_2) \notin R \tag{2.4b}$$

and

$$\sigma_{i3}(x_1, x_2, 0^+) = \sigma_{i3}(x_1, x_2, 0^-) = -t_i(x_1, x_2), \quad \text{for} \quad (x_1, x_2) \in R \tag{2.5}$$

where $t_i(x_1, x_2)$ is the traction on the crack faces. Equations (2.4a, b) describe the continuity of traction and displacement across the interface outside the cracked region, while (2.5) expresses the self-balanced traction loading on the two crack faces. To ensure the uniqueness of the solution, asymptotic conditions at infinity are needed, i.e.,

$$\sigma_{ij} \to 0 \quad \text{as} \quad |\mathbf{x}| \to \infty. \tag{2.6}$$

In addition, the strain energy density must be integrable near the crack edge.

Equations (2.1)–(2.6) form a boundary value problem for the displacement and stress fields. In the next section, this boundary value problem will be solved by using the three-dimensional Stroh method.

3 The Three-Dimensional Stroh Method

The original Stroh method (Stroh, 1958) can be extended to three-dimensional anisotropic elasticity (Ting, 1996). Following Ting's approach, a brief derivation of the Stroh method applied to three-dimensional analysis is presented in this section.

First, consider the two-dimensional Fourier transform pairs defined by

$$\tilde{f}(\mathbf{y}) = \iint_{\infty} f(\mathbf{x}) e^{i\mathbf{y}\cdot\mathbf{x}} d\mathbf{x},$$

$$f(\mathbf{x}) = \frac{1}{4\pi^2} \iint_{\infty} \tilde{f}(\mathbf{y}) e^{-i\mathbf{y}\cdot\mathbf{x}} d\mathbf{y}, \tag{3.1a, b}$$

where $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$. The application of (3.1a) to (2.1) in conjunction with (2.2) yields (Willis, 1966)

$$C_{i\alpha k\beta} y_\alpha y_\beta \tilde{u}_k + i(C_{i\alpha k3} + C_{i3k\alpha}) y_\alpha \tilde{u}_{k,3} - C_{i3k3}\tilde{u}_{k,33} = 0, \tag{3.2}$$

where $\alpha, \beta = 1, 2$.

Next, let the general solution to (3.2) be written as

$$\tilde{\mathbf{u}}(y_1, y_2, x_3) = \mathbf{a} \exp(-ip\rho x_3), \tag{3.3}$$

where $\mathbf{a}$ and p are to be determined and

$$\mathbf{y} = \rho\mathbf{n}, \quad \mathbf{n} = (n_1, n_2, 0)^T, \quad \mathbf{m} = (0, 0, 1)^T. \tag{3.4a, b, c}$$

Substitution of (3.3) into (3.2) yields

$$[\mathbf{Q} + p(\mathbf{R} + \mathbf{R}^T) + p^2\mathbf{T}]\mathbf{a} = \mathbf{0}, \tag{3.5}$$

where

$$Q_{ik} = C_{ijks} n_j n_s, \quad R_{ik} = C_{ijks} n_j m_s, \quad T_{ik} = C_{ijks} m_j m_s. \tag{3.6}$$

Given that the stiffness tensor C_{ijkl} is positive definite with the usual symmetries, and m_i, n_i are nonzero unit direction vectors, one can show that $\mathbf{Q}$ and $\mathbf{T}$ are symmetric and positive definite (Ting, 1996).

To ensure that (3.3) is a solution to (3.2), $\mathbf{a}$ and p must satisfy the eigenvalue problem (3.5), which is the Stroh eigenrelation for the oblique plane spanned by the vectors $\mathbf{n}$ and $\mathbf{m}$ (Ting, 1996). Therefore, it can be shown that there are three pairs of complex eigenvalues, $p_1, p_2 \ldots p_6$. Let the eigenvalues be arranged in such a way that the first three have a positive imaginary part, i.e., $\mathrm{Im}\{p_n\} > 0$, for $n = 1, 2, 3$. One can then define

$$\mathbf{P} = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{bmatrix}, \quad \mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3],$$

$$\mathbf{B} = \mathbf{R}^T\mathbf{A} + \mathbf{T}\mathbf{A}\mathbf{P}, \tag{3.7}$$

where $\mathbf{a}_n$ is the eigenvector associated with p_n.

Now, the displacement and stress fields in the Fourier space can be written as (Ting, 1996)

$$\tilde{\mathbf{u}}(\mathbf{y}, x_3) = \mathbf{A}\langle \exp(-i\rho p_m x_3)\rangle \mathbf{q} + \bar{\mathbf{A}}\langle \exp(-i\rho \bar{p}_m x_3)\rangle \mathbf{h}, \tag{3.8}$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\tilde{\boldsymbol{\tau}}(\mathbf{y}, x_3) = -i\rho[\mathbf{B}\langle\exp(-i\rho p_m x_3)\rangle\mathbf{q} + \bar{\mathbf{B}}\langle\exp(-i\rho\bar{p}_m x_3)\rangle\mathbf{h}], \quad (3.9)$$

where $\mathbf{q}$ and $\mathbf{h}$ are arbitrary constant vectors and

$$\tilde{\tau}_i(\mathbf{y}, x_3) = \tilde{\sigma}_{i3}(\mathbf{y}, x_3), \quad (3.10)$$

$$\langle\exp(-i\rho p_m x_3)\rangle \equiv \text{diag}\,[\exp(-i\rho p_1 x_3), \exp(-i\rho p_2 x_3), \exp(-i\rho p_3 x_3)]. \quad (3.11)$$

In (3.8)–(3.9) and the sequel, an overbar stands for complex conjugate.

Finally, taking the inverse Fourier transform on (3.8)–(3.9) yields

$$\mathbf{u}(\mathbf{x}, x_3) = \frac{1}{4\pi^2}\iint_\infty [\mathbf{A}\langle\exp(-i\rho p_m x_3)\rangle\mathbf{q} + \bar{\mathbf{A}}\langle\exp(-i\rho\bar{p}_m x_3)\rangle\mathbf{h}]\exp(-i\mathbf{y}\cdot\mathbf{x})d\mathbf{y}, \quad (3.12)$$

$$\boldsymbol{\tau}(\mathbf{x}, x_3) = \frac{-i}{4\pi^2}\iint_\infty \rho[\mathbf{B}\langle\exp(-i\rho p_m x_3)\rangle\mathbf{q} + \bar{\mathbf{B}}\langle\exp(-i\rho\bar{p}_m x_3)\rangle\mathbf{h}]\exp(-i\mathbf{y}\cdot\mathbf{x})d\mathbf{y}. \quad (3.13)$$

These are the formal integral representations of the displacement and traction fields. The unknowns in these representations are the vectors $\mathbf{q}$ and $\mathbf{h}$.

To close this section, it is noted that in addition to being functions of the elastic constants, matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{P}$ are also functions of the Fourier transform variable $\mathbf{y}$. To be more precise, they are functions of the unit vector $\mathbf{n}$ given by (3.4*b*). In general, the unknown vectors $\mathbf{q}$ and $\mathbf{h}$ are also functions of $\mathbf{y}$.

4 Derivation of the Integral Equation

In this section, the three-dimensional interface crack problem stated in Section 2 is formulated into a vector integral equation by using the Stroh method discussed in the previous section. To this end, by making use of (3.12)–(3.13), one can write the displacement and traction on the interface ($x_3 = 0$) for each of the two half-spaces as

$$\mathbf{u}(x_1, x_2, 0^+) = \frac{1}{4\pi^2}\iint_\infty \bar{\mathbf{A}}_1(\mathbf{n})\mathbf{h}_1(\mathbf{y})\exp(-i\mathbf{y}\cdot\mathbf{x})d\mathbf{y}, \quad (4.1a)$$

$$\boldsymbol{\tau}(x_1, x_2, 0^+) = \frac{-i}{4\pi^2}\iint_\infty \rho\bar{\mathbf{B}}_1(\mathbf{n})\mathbf{h}_1(\mathbf{y})\exp(-i\mathbf{y}\cdot\mathbf{x})d\mathbf{y}, \quad (4.1b)$$

$$\mathbf{u}(x_1, x_2, 0^-) = \frac{1}{4\pi^2}\iint_\infty \mathbf{A}_2(\mathbf{n})\mathbf{q}_2(\mathbf{y})\exp(-i\mathbf{y}\cdot\mathbf{x})d\mathbf{y}, \quad (4.1c)$$

$$\boldsymbol{\tau}(x_1, x_2, 0^-) = \frac{-i}{4\pi^2}\iint_\infty \rho\mathbf{B}_2(\mathbf{n})\mathbf{q}_2(\mathbf{y})\exp(-i\mathbf{y}\cdot\mathbf{x})d\mathbf{y}, \quad (4.1d)$$

where the subscripts on $\mathbf{A}(\mathbf{n})$ and $\mathbf{B}(\mathbf{n})$ are used to indicate the materials with which they are associated. For clarity, their dependence on the Fourier transformation variable $\mathbf{y} = \rho\mathbf{n}$ is also explicitly indicated. In deriving (4.1), the radiation condition (2.6) was used to ensure the finiteness of the fields at infinity.

Substitution of (4.1*b*) and (4.1*d*) into the traction continuity conditions (2.4*a*) and (2.5) yields

$$\bar{\mathbf{B}}_1(\mathbf{n})\mathbf{h}_1(\mathbf{y}) = \mathbf{B}_2(\mathbf{n})\mathbf{q}_2(\mathbf{y}) \equiv \mathbf{g}(\mathbf{y}). \quad (4.2a, b)$$

In terms of the newly introduced vector $\mathbf{g}(\mathbf{y})$, one has

$$\mathbf{h}_1(\mathbf{y}) = \bar{\mathbf{B}}_1^{-1}(\mathbf{n})\mathbf{g}(\mathbf{y}), \quad \mathbf{q}_2(\mathbf{y}) = \mathbf{B}_2^{-1}(\mathbf{n})\mathbf{g}(\mathbf{y}). \quad (4.3a, b)$$

Next, define the crack-opening displacement

$$\mathbf{b}(\mathbf{x}) = \begin{cases} \mathbf{u}(x_1, x_2, 0^+) - \mathbf{u}(x_1, x_2, 0^-) & (x_1, x_2) \in R \\ 0 & (x_1, x_2) \notin R. \end{cases} \quad (4.4)$$

It then follows from substituting (4.1*a*) and (4.1*c*) into the displacement continuity condition (2.4*b*) that

$$\frac{1}{4\pi^2}\iint_\infty \mathbf{M}^{-1}(\mathbf{n})\mathbf{g}(\mathbf{y})\exp(-i\mathbf{y}\cdot\mathbf{x})d\mathbf{y} = -i\mathbf{b}(\mathbf{x}), \quad (4.5)$$

where

$$\mathbf{M}^{-1}(\mathbf{n}) = \bar{\mathbf{M}}_1^{-1}(\mathbf{n}) + \mathbf{M}_2^{-1}(\mathbf{n}), \quad (4.6a)$$

$$\mathbf{M}_n^{-1}(\mathbf{n}) = i\mathbf{A}_n(\mathbf{n})\mathbf{B}_n^{-1}(\mathbf{n}), \quad n = 1, 2 \text{ (no summation)}. \quad (4.6b)$$

Taking the inverse Fourier transform on (4.5) yields

$$\mathbf{g}(\mathbf{y}) = -i\mathbf{M}(\mathbf{n})\iint_R \mathbf{b}(\mathbf{x})\exp(i\mathbf{y}\cdot\mathbf{x})d\mathbf{x}. \quad (4.7)$$

Finally, by substituting (4.7) into (4.2), then into (4.1*b*), then into (2.5), one obtains

$$\iint_\infty \rho\mathbf{M}(\mathbf{n})\left[\iint_R \mathbf{b}(\mathbf{z})\exp(i\mathbf{y}\cdot\mathbf{z})d\mathbf{z}\right]\exp(-i\mathbf{y}\cdot\mathbf{x})d\mathbf{y} = 4\pi^2\mathbf{t}(\mathbf{x}), \quad \text{for} \quad \mathbf{x} \in R. \quad (4.8)$$

Solution to (4.8) yields the crack-opening displacement $\mathbf{b}(\mathbf{x})$ on the crack surface. Once $\mathbf{b}(\mathbf{x})$ is known, other fracture parameters such as the stress intensity factor and energy release rate can be found from (4.1) by using (4.7) and (4.3). However, sophisticated discretization schemes must be used to obtain numerical solutions to (4.8) for the general case. In this paper, analytical solutions are of particular interest for some special cases of (4.8).

To obtain an analytical solution, it is convenient to derive an alternative form of (4.8). To this end, some special properties of the matrix $\mathbf{M}(\mathbf{n})$ will be noted. In the literature, $\mathbf{M}(\mathbf{n})$ is referred to as the interfacial impedance matrix (Lothe and Barnett, 1976). It can be shown that $\mathbf{M}(\mathbf{n})$ is a Hermitian matrix and, therefore, can be written as (Ting, 1996),

$$\mathbf{M}(\mathbf{n}) = \tilde{\mathbf{L}}(\mathbf{n}) - i\tilde{\mathbf{W}}(\mathbf{n}), \quad (4.9)$$

where

$$\tilde{\mathbf{W}}(\mathbf{n}) = \mathbf{D}^{-1}(\mathbf{n})\mathbf{W}(\mathbf{n})\tilde{\mathbf{L}}(\mathbf{n}),$$

$$\tilde{\mathbf{L}}(\mathbf{n}) = [\mathbf{D}(\mathbf{n}) + \mathbf{W}(\mathbf{n})\mathbf{D}^{-1}(\mathbf{n})\mathbf{W}(\mathbf{n})]^{-1}, \quad (4.10a, b)$$

and $\mathbf{D}(\mathbf{n})$ is symmetric and positive definite, $\mathbf{W}(\mathbf{n})$ is antisymmetric. They are known as the interfacial matrices (Qu and Li, 1991). The elastic mismatch of the bimaterial is characterized by $\mathbf{D}(\mathbf{n})$ and the pathological crack-tip behavior (oscillatory stresses and crack face interpenetration) is directly related to $\mathbf{W}(\mathbf{n})$ (Qu and Bassani, 1993). In terms of the Barnett-Lothe tensors, $\mathbf{L}(\mathbf{n})$ and $\mathbf{S}(\mathbf{n})$ (Barnett and Lothe, 1976),

$$\mathbf{D}(\mathbf{n}) = \mathbf{L}_1^{-1}(\mathbf{n}) + \mathbf{L}_2^{-1}(\mathbf{n}),$$

$$\mathbf{W}(\mathbf{n}) = \mathbf{S}_1(\mathbf{n})\mathbf{L}_1^{-1}(\mathbf{n}) - \mathbf{S}_2(\mathbf{n})\mathbf{L}_2^{-1}(\mathbf{n}), \quad (4.11a, b)$$

where, again, the subscripts 1 and 2 indicate with which material half-space the Barnett-Lothe tensors are associated.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

It can be shown (Ting, 1996) that the $\mathbf{L}(\mathbf{n})$ and $\mathbf{S}(\mathbf{n})$ are even and odd functions of $\mathbf{n}$, respectively, i.e.,

$$\mathbf{L}(\mathbf{n}) = \mathbf{L}(-\mathbf{n}), \quad \mathbf{S}(\mathbf{n}) = -\mathbf{S}(-\mathbf{n}). \quad (4.12a, b)$$

Consequently,

$$\mathbf{D}(\mathbf{n}) = \mathbf{D}(-\mathbf{n}), \quad \mathbf{W}(\mathbf{n}) = -\mathbf{W}(-\mathbf{n}). \quad (4.13a, b)$$

Although $\mathbf{M}(\mathbf{n})$ can be obtained by solving the eigenvalue problem (3.5) then using (3.7) and (4.6a, b), a more convenient alternative method of obtaining it is presented in the Appendix.

Now, introduce the Radon transform

$$\hat{\mathbf{b}}(t, \mathbf{n}) = \iint_R \mathbf{b}(\mathbf{z})\delta(t - \mathbf{n}\cdot\mathbf{z})d\mathbf{z}, \quad (4.14)$$

where $\delta(x)$ is the Dirac delta function. Since the crack occupies a finite region R, the properties of the Dirac delta function imply that there is a finite region $[-a, a]$ so that $\hat{\mathbf{b}}(t, \mathbf{n}) = 0$ for $t \notin [-a, a]$. Thus, it is easy to show that (4.8) can be recast into

$$\iint_\infty \rho\mathbf{M}(\mathbf{n}) \int_{-a}^{a} \hat{\mathbf{b}}(t, \mathbf{n}) \exp[i\rho(t - \mathbf{n}\cdot\mathbf{x})]dtd\mathbf{y} = 4\pi^2\mathbf{t}(\mathbf{x}), \quad (4.15)$$

or

$$\oint_{|\mathbf{n}|=1} \mathbf{M}(\mathbf{n}) \frac{-\partial^2}{\partial(\mathbf{n}\cdot\mathbf{x})^2} \int_{-a}^{a} \hat{\mathbf{b}}(t, \mathbf{n}) \times \int_0^\infty \exp[i\rho(t - \mathbf{n}\cdot\mathbf{x})]d\rho dtd\mathbf{n} = 4\pi^2\mathbf{t}(\mathbf{x}). \quad (4.16)$$

By using the integral (Yang and Qu, 1993),

$$\int_0^\infty \exp[i\rho(t - \mathbf{n}\cdot\mathbf{x})]d\rho = \frac{i}{t - \mathbf{n}\cdot\mathbf{x}} + \pi\delta(t - \mathbf{n}\cdot\mathbf{x}), \quad (4.17)$$

Eq. (4.16) can be further simplified to

$$\oint_{|\mathbf{n}|=1} \mathbf{M}(\mathbf{n}) \frac{-\partial^2}{\partial(\mathbf{n}\cdot\mathbf{x})^2} \int_{-a}^{a} \hat{\mathbf{b}}(t, \mathbf{n}) \left[\frac{i}{t - \mathbf{n}\cdot\mathbf{x}} + \pi\delta(t - \mathbf{n}\cdot\mathbf{x})\right] dtd\mathbf{n} = 4\pi^2\mathbf{t}(\mathbf{x}). \quad (4.18)$$

Next, define a sectionally analytical function of z

$$\mathbf{f}(z, \mathbf{n}) = \frac{1}{2\pi i} \int_{-a}^{a} \frac{\hat{\mathbf{b}}(t, \mathbf{n})}{t - z} dt,$$

$$\mathbf{f}''(z, \mathbf{n}) = \frac{\partial^2 \mathbf{f}(z, \mathbf{n})}{\partial z^2}. \quad (4.19a,b)$$

Then, by using the Plemelj formulas (Muskhelishvili, 1953),

$$\mathbf{f}_+(\mathbf{n}\cdot\mathbf{x}, \mathbf{n}) + \mathbf{f}_-(\mathbf{n}\cdot\mathbf{x}, \mathbf{n}) = \frac{1}{\pi i} \int_{-a}^{a} \frac{\hat{\mathbf{b}}(t, \mathbf{n})}{t - \mathbf{n}\cdot\mathbf{x}} dt, \quad (4.20a)$$

$$\mathbf{f}_+(\mathbf{n}\cdot\mathbf{x}, \mathbf{n}) - \mathbf{f}_-(\mathbf{n}\cdot\mathbf{x}, \mathbf{n}) = \hat{\mathbf{b}}(\mathbf{n}\cdot\mathbf{x}, \mathbf{n}), \quad (4.20b)$$

one can finally write (4.18) as

$$\oint_{|\mathbf{n}|=1} [\mathbf{M}(\mathbf{n})\mathbf{f}''_-(\mathbf{n}\cdot\mathbf{x}, \mathbf{n}) - \bar{\mathbf{M}}(\mathbf{n})\mathbf{f}''_+(\mathbf{n}\cdot\mathbf{x}, \mathbf{n})]d\mathbf{n} = 4\pi\mathbf{t}(\mathbf{x}). \quad (4.21)$$

In deriving (4.21), only the real part of the left-hand side of (4.18) is used since the right-hand side is real. Equation (4.21) was derived by Willis (1972) using a somewhat different approach. Besides a numerical factor, the interfacial matrix $\mathbf{M}(\mathbf{n})$ is the same as the matrix $\mathbf{A}$ in Willis' (1972) paper.

5 The Non-Oscillatory Solution

It is well known for two-dimensional interface cracks that the crack-tip fields are pathological in that the near-tip stresses are oscillatory and the crack faces interpenetrate each other near the tip (e.g., Erdogan, 1968). Such pathological behavior is also known as the β-effect (Rice, 1988) for it disappears when the second Dundurs (1969) constant β of an isotropic bimaterial vanishes. Fortunately, for most isotropic engineering bimaterial systems, β is rather small. Thus for many practical applications, the linear elasticity solution predicts the pathological behavior only over a very small region near the crack tip (Hutchinson, et al., 1987).

For anisotropic bimaterials, the condition for non-oscillatory solutions to exist becomes more complex. Qu and Bassani showed (1989) that the necessary and sufficient condition for a two-dimensional interface crack-tip field to be non-oscillatory is that the bimaterial constant $\beta = [-0.5 \text{ tr } [\mathbf{D}^{-1}\mathbf{W}]^2]^{1/2} = 0$. Analogous to the two-dimensional case, one can show that the three-dimensional interface crack-tip field is non-oscillatory if the bimaterial is such that $\beta(\mathbf{n}) = [-0.5 \text{ tr } [\mathbf{D}^{-1}(\mathbf{n})\mathbf{W}(\mathbf{n})]^2]^{1/2} = 0$, which is equivalent to $\mathbf{W}(\mathbf{n}) = \mathbf{0}$, where $\mathbf{D}(\mathbf{n})$ and $\mathbf{W}(\mathbf{n})$ are given in (4.11). Proof of the above statement can be found in (Xue, 1998). In the present paper, it is assumed in the sequel that $\beta = 0$, or $\mathbf{W}(\mathbf{n}) = \mathbf{0}$, so only the non-oscillatory solutions are investigated.

To this end, consider a circular crack (penny-shape) of radius a on the interface, i.e.,

$$R = \{\mathbf{x} \in [x_1^2 + x_2^2 = a^2, x_3 = 0]\}. \quad (5.1)$$

When $\mathbf{W}(\mathbf{n}) = \mathbf{0}$, it follows from (4.10) that $\mathbf{M}(\mathbf{n}) = \tilde{\mathbf{L}}(\mathbf{n})$ is a real, symmetric, and even matrix of $\mathbf{n}$. Thus, (4.21) can be simplified to

$$\oint_{|\mathbf{n}|=1} \tilde{\mathbf{L}}(\mathbf{n}) \frac{\partial^2 \hat{\mathbf{b}}(\mathbf{n}\cdot\mathbf{x}, \mathbf{n})}{\partial(\mathbf{n}\cdot\mathbf{x})^2} d\mathbf{n} = -4\pi\mathbf{t}(\mathbf{x}), \quad (5.2)$$

where the Plemelj formula (4.20b) was used again. This equation can be solved if the applied traction $\mathbf{t}(\mathbf{x})$ is a polynomial of degree k, e.g., $\mathbf{t}(\mathbf{x}) = \mathbf{t}_k(\mathbf{x})$. In this case (Willis, 1972)

$$\hat{\mathbf{b}}(t, \mathbf{n}) = (a^2 - t^2)\mathbf{g}_k(t, \mathbf{n}), \quad \mathbf{b}(\mathbf{x}) = (a^2 - |\mathbf{x}|^2)^{1/2}\mathbf{v}_k(\mathbf{x}), \quad |t| \text{ and } |\mathbf{x}| < a, \quad (5.3a, b)$$

where $\mathbf{g}_k(t, \mathbf{n})$ and $\mathbf{v}_k(\mathbf{x})$ are, respectively, polynomials of degree k in t and $\mathbf{x}$. In other words, the crack-opening displacement is a polynomial of degree k if the applied traction on the crack surface is a polynomial of degree k.

For example, consider uniform loading ($k = 0$), i.e., $\mathbf{t}(\mathbf{x}) = \mathbf{t}_0 = \text{const}$. According to (5.3) and (4.14), one has

$$\mathbf{b}(\mathbf{x}) = (a^2 - r^2)^{1/2}\mathbf{v}_0, \quad \hat{\mathbf{b}}(t, \mathbf{n}) = \frac{\pi}{2}(a^2 - t^2)\mathbf{v}_0, \quad |t| \text{ and } r < a, \quad (5.4a, b)$$

where $\mathbf{v}_0$ is a constant vector to be determined and r is the polar coordinate defined by $x_1 = r\cos\theta$, $x_2 = r\sin\theta$ (see Fig. 1). Substitution of (5.4b) into (5.2) yields an equation for the unknown vector $\mathbf{v}_0$

$$\left[\oint_{|\mathbf{n}|=1} \tilde{\mathbf{L}}(\mathbf{n})d\mathbf{n}\right]\mathbf{v}_0 = 4\mathbf{t}_0. \quad (5.5)$$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

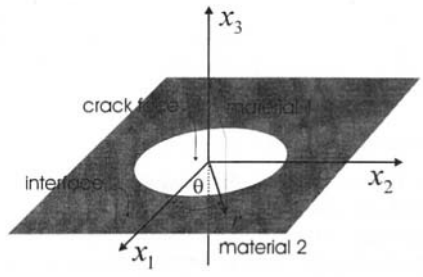


Fig. 1 A three-dimensional interface crack in an anisotropic bimaterial

Thus,

$$\mathbf{v}_0 = \frac{2}{\pi}\tilde{\mathbf{L}}_0^{-1}\mathbf{t}_0, \tag{5.6}$$

where

$$\tilde{\mathbf{L}}_0 = \frac{1}{2\pi}\oint_{|\mathbf{n}|=1}\tilde{\mathbf{L}}(\mathbf{n})d\mathbf{n}. \tag{5.7}$$

Then, it follows from (5.4*a*) that the crack-opening displacement is obtained as

$$\mathbf{b}(r,\theta) = \frac{2}{\pi}(a^2 - r^2)^{1/2}\tilde{\mathbf{L}}_0^{-1}\mathbf{t}_0. \tag{5.8}$$

6 Fracture Mechanics Parameters

First, consider the interfacial traction vector outside the crack. Analogous to the derivation from (4.14) to (4.21), it can be shown that the interfacial traction outside the crack surface is given by

$$\boldsymbol{\tau}(x_1, x_2, 0) = \frac{-1}{2\pi}\oint_{|\mathbf{n}|=1}\mathbf{M}(\mathbf{n})\mathbf{f}''(\mathbf{n}\cdot\mathbf{x}, \mathbf{n})d\mathbf{n}, \qquad |\mathbf{x}| > a \tag{6.1}$$

where, according to (4.19) and Muskhelishvili (1953),

$$\mathbf{f}(z, \mathbf{n}) = \frac{1}{2\pi i}\int_{-a}^{a}\frac{\hat{\mathbf{b}}(t, \mathbf{n})}{t - z}dt, \tag{6.2}$$

$$\mathbf{f}''(\mathbf{n}\cdot\mathbf{x}, \mathbf{n}) = -\frac{1}{2}\frac{\partial^2\hat{\mathbf{b}}(\mathbf{n}\cdot\mathbf{x}, \mathbf{n})}{\partial(\mathbf{n}\cdot\mathbf{x})^2} + \frac{1}{2\pi i}\frac{\partial^2}{\partial(\mathbf{n}\cdot\mathbf{x})^2}\int_{-a}^{a}\frac{\hat{\mathbf{b}}(t, \mathbf{n})}{t - \mathbf{n}\cdot\mathbf{x}}dt. \tag{6.3}$$

Again, it is assumed that the bimaterial is such that $\mathbf{W} = \mathbf{0}$. It then follows from the previous section that for a circular crack of radius a subjected to uniform surface traction $\mathbf{t}_0$, one has from (5.4*b*) and (6.3) that

$$\mathbf{f}''(\mathbf{n}\cdot\mathbf{x}, \mathbf{n}) = \frac{\pi}{2}\mathbf{v}_0 H(a - \mathbf{n}\cdot\mathbf{x}) - \frac{1}{2i}\left[\ln\left(\frac{a - \mathbf{n}\cdot\mathbf{x}}{a + \mathbf{n}\cdot\mathbf{x}}\right) + \frac{2a\mathbf{n}\cdot\mathbf{x}}{(\mathbf{n}\cdot\mathbf{x})^2 - a^2}\right]\mathbf{v}_0. \tag{6.4}$$

Substitution of (6.4) into (6.1) yields the interfacial traction outside the cracked region. Of primary interest in fracture mechanics is the asymptotic behavior near the crack edge, $r = a$. It is easy to confirm that the contributions from all terms in (6.4) are bounded as $r \to a$, except the last term in the brackets. Therefore, asymptotically,

$$\begin{aligned}\boldsymbol{\tau}(x_1, x_2, 0) &\sim \frac{a}{2\pi i}\oint_{|\mathbf{n}|=1}\frac{(\mathbf{n}\cdot\mathbf{x})\tilde{\mathbf{L}}(\mathbf{n})\mathbf{v}_0}{(\mathbf{n}\cdot\mathbf{x})^2 - a^2}d\mathbf{n} \quad \text{as} \quad r \to a + 0 \\ &= \frac{a^2}{2\pi i}\oint_{|\mathbf{n}|=1}\frac{[(\mathbf{n}\cdot\hat{\mathbf{x}})\tilde{\mathbf{L}}(\mathbf{n}) - \tilde{\mathbf{L}}(\hat{\mathbf{x}})]\mathbf{v}_0}{(\mathbf{n}\cdot\mathbf{x})^2 - a^2}d\mathbf{n} \\ &\quad + \frac{a^2\tilde{\mathbf{L}}(\hat{\mathbf{x}})\mathbf{v}_0}{2\pi i}\oint_{|\mathbf{n}|=1}\frac{d\mathbf{n}}{(\mathbf{n}\cdot\mathbf{x})^2 - a^2},\end{aligned} \tag{6.5}$$

where $\hat{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|$ and $\mathbf{v}_0$ is defined by (5.5). Notice that the first integral is finite and the second term gives (Willis, 1968),

$$\oint_{|\mathbf{n}|=1}\frac{d\mathbf{n}}{(\mathbf{n}\cdot\mathbf{x})^2 - a^2} = \frac{2\pi i}{a\sqrt{|\mathbf{x}|^2 - a^2}}. \tag{6.6}$$

Thus, the crack-edge singular stresses are given by

$$\boldsymbol{\tau}(r, \theta) \sim \frac{2a}{\pi\sqrt{r^2 - a^2}}\tilde{\mathbf{L}}(\hat{\mathbf{x}})\tilde{\mathbf{L}}_0^{-1}\mathbf{t}_0 \quad \text{as} \quad r \to a \tag{6.7}$$

Or, one can write

$$\boldsymbol{\tau}(r, \theta) \sim \frac{\mathbf{k}(\hat{\mathbf{x}})}{\sqrt{2\pi(r - a)}}, \quad \hat{\mathbf{x}} = (\cos\theta, \sin\theta)^T \tag{6.8}$$

where the stress intensity vector is defined as

$$\mathbf{k}(\hat{\mathbf{x}}) = \frac{2}{\pi}\sqrt{\pi a}\,\tilde{\mathbf{L}}(\hat{\mathbf{x}})\tilde{\mathbf{L}}_0^{-1}\mathbf{t}_0. \tag{6.9}$$

The crack-opening displacement, the interfacial traction, and the stress intensity vector in the cylindrical coordinate system $(r\theta z)$ can be obtained through a simple coordinate transformation

$$\mathbf{b}_R(r, \theta) = \begin{bmatrix} u_r(r, 0^+) - u_r(r, 0^-) \\ u_\theta(r, 0^+) - u_\theta(r, 0^-) \\ u_z(r, 0^+) - u_z(r, 0^-) \end{bmatrix} = \boldsymbol{\alpha}(\theta)\mathbf{b}(\mathbf{x}), \tag{6.10}$$

$$\boldsymbol{\tau}_R(r, \theta) = \begin{bmatrix} \sigma_{rz}(r, \theta, 0) \\ \sigma_{\theta z}(r, \theta, 0) \\ \sigma_{zz}(r, \theta, 0) \end{bmatrix} = \boldsymbol{\alpha}(\theta)\boldsymbol{\tau}(r, \theta), \tag{6.11}$$

$$\mathbf{k}_R(\theta) = \begin{bmatrix} K_{II}(\theta) \\ K_{III}(\theta) \\ K_I(\theta) \end{bmatrix} = \boldsymbol{\alpha}(\theta)\mathbf{k}(\hat{\mathbf{x}}), \tag{6.12}$$

where $K_I(\theta)$, $K_{II}(\theta)$, $K_{III}(\theta)$ are the conventional stress intensity factors of mode I, II, and III, respectively. The coordinate transformation matrix in (6.10) – (6.12) is given by

$$\boldsymbol{\alpha}(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6.13}$$

Now assuming that under load the crack edge at $r = a$ advances along the interface to $r = a + \delta a$. The energy release rate associated with this crack growth is defined by

$$G(\theta) = \lim_{\delta a \to 0}\left\{\frac{1}{\delta a}\int_a^{a+\delta a}\frac{1}{2}\boldsymbol{\tau}(r, \theta)\cdot\mathbf{b}(r - \delta a, \theta)dr\right\}. \tag{6.14}$$

In the case of uniform load, $\mathbf{t}_0$, the above integral can be evaluated by using (5.4*a*), and (6.7) to give

$$G(\theta) = \frac{a}{\pi}\mathbf{t}_0^T\tilde{\mathbf{L}}_0^{-1}\tilde{\mathbf{L}}(\hat{\mathbf{x}})\tilde{\mathbf{L}}_0^{-1}\mathbf{t}_0, \tag{6.15a}$$

Or, in terms of the stress intensity vector,

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 1 Elastic constants of a unidirectional graphite-epoxy fiber-reinforced composite

C_{11}	C_{22}	C_{12}	C_{23}	C_{55}
100.3×10^9 Pa	13.2×10^9 Pa	5.0×10^9 Pa	6.42×10^9 Pa	6.16×10^9 Pa

$$G(\theta) = \tfrac{1}{4}\mathbf{k}(\hat{\mathbf{x}})^T\tilde{\mathbf{L}}^{-1}(\hat{\mathbf{x}})\mathbf{k}(\hat{\mathbf{x}}). \qquad (6.15b)$$

It is worth mentioning that the above method also provides an easier way to handle three-dimensional cracks in homogeneous media. For example, by using the explicit expressions of $\tilde{\mathbf{L}}^{-1}(\hat{\mathbf{x}})$ derived in the Appendix, Eqs. (5.8) and (6.9) can be easily reduced to the well-known results

$$b_z(r) = \frac{4(1-\nu)\sigma_0}{\pi\mu}\sqrt{a^2 - r^2}, \quad K_I = \frac{2\sigma_0}{\pi}\sqrt{\pi a}, \qquad (6.16a, b)$$

for a penny-shaped crack of radius a subjected to uniform mode I loading σ_0. When the penny-shaped crack is subjected to uniform shear τ_0 in the x_1-direction, i.e., when $\mathbf{t}_0 = (\tau_0, 0, 0)^T$, Eqs. (6.9)–(6.12) yield

$$b_r(r, \theta) = \frac{8(1-\nu)\tau_0}{\pi\mu(2-\nu)}\sqrt{a^2 - r^2}\cos\theta, \qquad (6.17a)$$

$$b_\theta(r, \theta) = \frac{-8(1-\nu)\tau_0}{\pi\mu(2-\nu)}\sqrt{a^2 - r^2}\sin\theta, \qquad (6.17b)$$

$$K_{II}(\theta) = \frac{4\cos\theta}{\pi(2-\nu)}\tau_0\sqrt{\pi a}, \qquad (6.18a)$$

$$K_{III}(\theta) = \frac{-4(1-\nu)\sin\theta}{\pi(2-\nu)}\tau_0\sqrt{\pi a}. \qquad (6.18b)$$

These results were obtained by Sneddon (1969). Comparing with Sneddon's derivations, the simplicity of the current method is obvious.

7 Numerical Examples for Transversely Isotropic Solids

Consider a transversely isotropic solid of infinite extent. Let the Cartesian coordinate system be such that the x_1-axis coincides with the axis of symmetry of the transversely isotropic solid. The elastic constants of the transversely isotropic solids are listed in Table 1. These values are taken from the effective properties of a graphite-epoxy unidirectional fiber-reinforced composite.

Assume a circular crack of radius a is located in the x_1x_2-plane. First, consider a uniform opening load, $\mathbf{t}(\mathbf{x}) = (0, 0, \sigma_0)^T$. In this case, the crack-edge singular field is purely mode I (opening mode). Shown in Fig. 2 is the mode I stress intensity factor when only a uniform mode I load is applied. Obviously, if the material were isotropic, the stress intensity factor should be a constant along the crack edge, i.e., $K_I(\theta)$ is independent of the polar angle θ. Due to material anisotropy, the mode I stress intensity factor varies as shown in Fig. 2. The maximum $K_I(\theta)$ is at $\theta = 0$ deg, 180 deg, i.e., in the x_1-direction (or the fiber direction), while the minimum occurs at $\theta = 90$ deg, 270 deg, i.e., in the x_2-direction (perpendicular to the fiber direction). The difference between the maximum and minimum is more than 35 percent. This may not be surprising since the material is fairly anisotropic in these two directions, e.g., $C_{11}^{(1)}/C_{22}^{(1)} > 7.5$. The pattern of the stress intensity factor distribution along the crack edge indicates that unlike a crack in an isotropic solid, a circular crack is not a stable shape for a crack in anisotropic solids. For example, a circular crack in a transversely isotropic laminate solid will tend to grow into an elliptical shape with its major axis in the fiber direction.

Next, consider a uniform shearing load in the x_1-direction, $\mathbf{t}(\mathbf{x}) = (\tau_0, 0, 0)^T$. In this case, mixed-mode singular stress fields near the crack edge develop, i.e., both mode II and mode III may co-exist. Shown in Fig. 3 are the normalized stress intensity factors of mode II and mode III, respectively, as functions of the polar angle. The corresponding energy release rate computed from (6.15) is also shown in Fig. 3. As expected, the maximum energy release rate is at $\theta = 0$ due to the nature of the load. It is interesting to notice that unlike in the isotropic case, the maximum value of $K_{III}(\theta)$ is not at $\theta = 90$ deg due to material anisotropy.

Now, consider a uniform shearing load in the x_2-direction, $\mathbf{t}(\mathbf{x}) = (0, \tau_0, 0)^T$. Again, mixed-mode singular stress fields near the crack edge develop. Shown in Fig. 4 are the normalized stress intensity factors of mode II and mode III, respectively, as functions of the polar angle. The corresponding energy release rate computed from (6.15) is also shown in Fig. 4. Surprisingly, the energy release rate is almost a constant along the crack edge, particularly between $\theta = 40$ deg and $\theta = 140$ deg, even though the individual stress intensity factors vary significantly.

The next example considered here is an interface crack in an anisotropic bimaterial consisting of two transversely isotropic half-spaces of identical properties with perpendicular orientations. Assume that the axis of isotropy for material 1 is in the x_1-direction and the axis of material 2 is in the x_2-direction. A circular crack of radius a is located on the interface between the two transversely isotropic half-spaces. This configuration mimics a circular delamination in a cross-ply composite. The

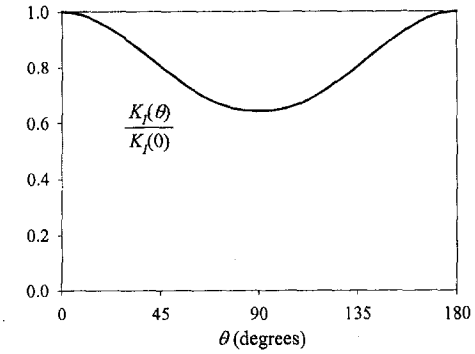


Fig. 2 Fracture parameters for a circular crack in a transversely isotropic solid under uniform normal loading, $\mathbf{t}(\mathbf{x}) = (0, 0, \sigma_0)^T$

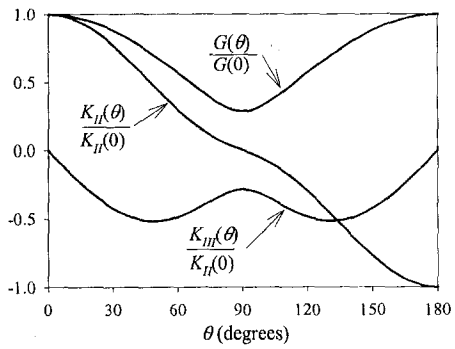


Fig. 3 Fracture parameters for a circular crack in a transversely isotropic solid under uniform shear loading in the x_1-direction, $\mathbf{t}(\mathbf{x}) = (\tau_0, 0, 0)^T$

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

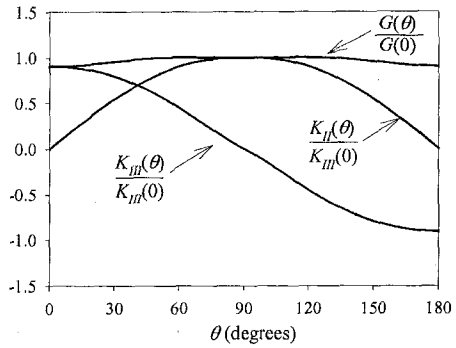


Fig. 4 Fracture parameters for a circular crack in a transversely isotropic solid under uniform shear loading in the x_2-direction, $\mathbf{t}(\mathbf{x}) = (0, \tau_0, 0)^T$

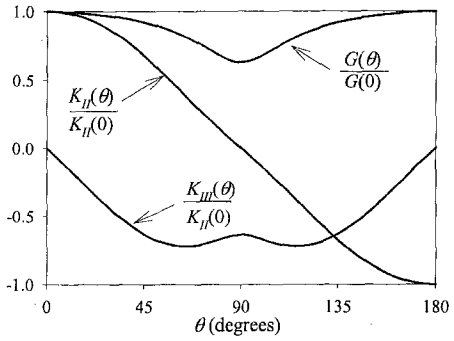


Fig. 6 Fracture parameters for a circular delamination in a cross-ply composite under uniform shear loading in the x_1-direction, $\mathbf{t}(\mathbf{x}) = (\tau_0, 0, 0)^T$

elastic constants for material 1 are listed in Table 1. For material 2, the elastic constants are obtained from those in Table 1 through a 90-deg rotation about the x_3-axis.

To proceed with the non-oscillatory solution discussed in previous sections, one needs to justify the assumption $\mathbf{W}(\mathbf{n}) = \mathbf{0}$. A measure of the "magnitude" of $\mathbf{W}(\mathbf{n})$ is the so-called oscillatory index $\beta(\mathbf{n}) = [-0.5 \text{ tr } [\mathbf{D}^{-1}(\mathbf{n})\mathbf{W}(\mathbf{n})]^2]^{1/2}$, which reduces to the second Dundurs constant β for isotropic bimaterials. For the cross-ply bimaterial considered here, it is found that $|\beta(\mathbf{n})| \leq 0.03065$. This justifies the use of the non-oscillatory solutions.

In all the numerical examples given below, the interfacial impedance matrix $\mathbf{M}(\mathbf{n})$ was computed first based on the material constants given in Table 1. Then, $\tilde{\mathbf{L}}(\mathbf{n})$ was obtained by taking the real part of $\mathbf{M}(\mathbf{n})$. Once $\tilde{\mathbf{L}}(\mathbf{n})$ is known as a function of the polar angle, the stress intensity factors and energy release rate can be evaluated.

First, consider a uniform opening load, $\mathbf{t}(\mathbf{x}) = (0, 0, \sigma_0)^T$. In this case, the crack-edge singular field is purely mode I (opening mode). Shown in Fig. 5 is the mode I stress intensity factor when only a uniform mode I load is applied. Obviously, if the bimaterial were isotropic, the stress intensity factor should be a constant along the crack edge, i.e., $K_I(\theta)$ is independent of the polar angle θ. Due to material anisotropy, one might expect some variation along the crack edge. However, it is seen from Fig. 5 that $K_I(\theta)$ varies less than 2.5 percent along the crack edge, even though the bimaterial is fairly anisotropic, $C_{11}^{(1)}/C_{22}^{(1)} > 7.5$. This is distinctively different from the example shown in Fig. 2. Comparison of these two cases indicates that the interface between two cross-plies is much more delamination tolerant than the interface between two unidirectional plies. In other words, a circular delamination between cross-plies tends to grow self-similarly while a circular delamination between unidirectional plies tends to grow into elliptical shape.

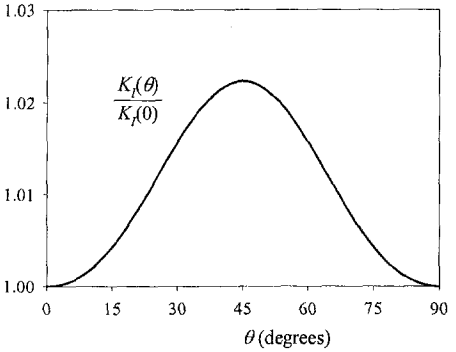


Fig. 5 Fracture parameters for a circular delamination in a cross-ply composite under uniform normal loading, $\mathbf{t}(\mathbf{x}) = (0, 0, \sigma_0)^T$

Energetically, the latter is more favorable for crack growth given that all other conditions are the same. This qualitatively explains why the well-known low-velocity impact-induced "peanut-shaped" delamination always has its major axis in the fiber direction (Liu, 1988).

Next, consider a uniform shearing load in the x_1-direction, $\mathbf{t}(\mathbf{x}) = (\tau_0, 0, 0)^T$. Again, both mode II and mode III exist. Their normalized values are shown in Fig. 6 as functions of the polar angle. The corresponding energy release rate computed from (6.15) is also shown in Fig. 6. As expected, the maximum energy release rate is at $\theta = 0$ due to the nature of the load. It is interesting to notice that at the maximum value of $K_{III}(\theta)$ is not at $\theta = 90$ deg, due to the anisotropy of the cross-ply composite.

8 Summary and Concluding Remarks

In this paper, the original Stroh method is extended to three-dimensional anisotropic elasticity analysis. By using this method, a planar crack on the interface between two dissimilar anisotropic elastic materials is considered. The solution to a circular interface crack in an anisotropic bimaterial is obtained, which is believed to be new. It is found that when the oscillatory index is small, the singular stress fields near the crack edge can be expressed in terms of the interfacial impedance matrix. In addition, the method also provides a very simple and convenient way to solve three-dimensional crack problems in homogeneous anisotropic solids.

Several numerical examples are given to problems of circular delamination in cross-ply composites. Stress intensity factors and energy release rates are obtained numerically. Among other things, it is found that the interface between cross-plies is more delamination tolerant than the interface between unidirectional plies.

References

Barnett, D. M., and Lothe, J., 1976, "Synthesis of the Sextic and the Integral Formalism for Dislocations, Green's Functions and Surface Waves in Anisotropic Elastic Solids," *Phys., Norv.*, Vol. 7, pp. 13–19.

Bassani, J. L., and Qu, J., 1989, "Finite Crack on the Bimaterial and Bicrystal Interfaces," *J. Mech. Phys. Solids*, Vol. 37, pp. 435–454.

Bassani, J. L., and Qu, J., 1990, "Interfacial discontinuities and Average Bimaterial Properties," *Acta/Scripta Metallurgic Proceedings Series*, Vol. 4, pp. 401–406.

Clements, D. L., 1971, "A Crack between Dissimilar Anisotropic Media," *Int. J. Engng. Sci.*, Vol. 9, pp. 353–368.

Dundurs, J., 1969, discussion, ASME JOURNAL OF APPLIED MECHANICS, Vol. 36, pp. 650–652.

Erdogan, F., 1968, "Simultaneous Dual Integral Equations with Trigonometric and Bessel Kernels," *ZAMM*, Vol., 48, pp. 217–225.

Hutchinson, J. W., Mear, M. E., and Rice, J. R., 1987, "Crack Paralleling and Interface Between Dissimilar Materials," ASME JOURNAL OF APPLIED MECHANICS, Vol. 54, pp. 828–832.

Keer, L. K., 1967, "Mixed Boundary Value Problems for an Elastic Half-Space," *Proc. Camb. Phil. Soc.*, Vol. 63, pp. 1379–1386.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Liu, D., 1988, "Impact-Induced Delamination—A view of Bending Stiffness Mismatching," *J. Composite Materials,* Vol. 22, pp. 674–692.

Lothe, J., and Barnett, D. M., 1976, "On the Existence of Surface-Wave Solutions for Anisotropic Elastic Half-Spaces With Free Surface," *J. Appl. Phys.,* Vol. 47, pp. 428–433.

Mossakovskii, V. I., and Rybka, 1964, "Generalization of the Griffith-Sneddon Criterion for the Case of a Nonhomogeneous Body," *Prikl. Mat. Meh.,* Vol. 28, pp. 1061–1069.

Muskhelishvili, N. I., 1953, *Singular Integral Equations,* P. Noordhoff, Groningen, Holland.

Qu, J., and Bassani, J. L., 1989, "Cracks on Bimaterial and Bicrystal Interfaces," *J. Mech., Phys. Solids,* Vol. 37, pp. 417–433.

Qu, J., and Li, Q., 1991, "Interfacial dislocation and Its Applications to Interface Crack in Anisotropic Bimaterials," *J. Elasticity,* Vol. 26, pp. 169–195.

Qu, J., and Bassani, J. L., 1993, "Interfacial Fracture Mechanics for Anisotropic Bimaterials," ASME JOURNAL OF APPLIED MECHANICS, Vol. 60, pp. 422–431.

Rice, J. R., 1988, "Elastic Fracture Mechanics Concepts for Interfacial Cracks," ASME JOURNAL OF APPLIED MECHANICS, Vol. 55, pp. 98–103.

Sneddon, I. N., 1969, *Crack Problems in the Classical Theory of Elasticity,* John Wiley and Sons, New York.

Sou, Z., 1990, "Singularities, Interfaces and Cracks in Dissimilar Anisotropic Media," *Proc. Royal Soc. London,* Vol. A447, p. 331–358.

Stroh, A. N., 1958, "Dislocations and Cracks in Anisotropic Elasticity," *Phil. Mag.,* Vol. 3, pp. 625–646.

Ting, T. C. T., 1986, "Explicit Solutions and Invariance of Singularities at an Interface Crack in Anisotropic Composites," *Int. J. Solids Structures,* Vol. 18, pp. 139–152.

Ting, T. C. T., and Hwu, C., 1988, "Sextic Formalism in Anisotropic Elasticity for Almost Non-Semisimple Matric N," *Int. J. Solids Structures,* Vol. 24, pp. 65–76.

Ting, T. C. T., 1996, *Anisotropic Elasticity: Theory and Application,* Oxford University Press, New York.

Willis, J. R., 1966, "Hertzian Contact of Anisotropic Bodies," *J. Mech. Phys. Solids,* Vol. 14, pp. 163–176.

Willis, J. R., 1967, "Boussinesq Problems for an Anisotropic Half-Space," *J. Mech. Phys. Solids,* Vol. 15, pp. 331–339.

Willis, J. R., 1968, "The Stress Field around an Elliptical Crack in an Anisotropic Elastic Medium," *Int. J. Engen. Sci.,* Vol. 6, pp. 253–263.

Willis, J. R., 1971a, "Fracture Mechanics of Interfacial Cracks," *J. Mech. Phys. Solids,* Vol. 19, pp. 353–368.

Willis, J. R., 1971b, "Interfacial Stresses Induced by Arbitrary Loading of dissimilar Elastic Half-Spaces Joined Over a Circular Region," *J. Inst. Maths Applics,* Vol. 7, pp. 179–197.

Willis, J. R., 1972, "The Penny-shaped Crack on an Interface," *Quart. J. Mech. and Appl. Math.,* Vol. 25, pp. 368–385.

Xue, Y., 1998, "Three-Dimensional Interface Cracks in Anisotropic Bimaterials," Ph.D. thesis, School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA.

Yang, L., and Qu, J., 1993, "Fracture Mechanics Parameters for Cracks on Slightly Undulating Interfaces," *Int. J. Fracture,* Vol. 64, pp. 79–91.

APPENDIX

The equations needed to compute $\mathbf{S}(\mathbf{n})$ and $\mathbf{L}(\mathbf{n})$ are listed here. Derivations of these equations can be found in (Barnett and Lothe, 1973; Ting, 1996).

$$\mathbf{S}(\mathbf{n}) = \frac{1}{\pi}\int_0^{\pi} \mathbf{N}_1(\varphi)d\varphi, \quad \mathbf{L}(\mathbf{n}) = \frac{-1}{\pi}\int_0^{\pi} \mathbf{N}_3(\varphi)d\varphi, \quad \text{(A1)}$$

where

$$\mathbf{N}_1(\varphi) = -\mathbf{T}^{-1}(\varphi)\mathbf{R}^T(\varphi),$$

$$\mathbf{N}_3(\varphi) = \mathbf{R}(\varphi)\mathbf{T}^{-1}(\varphi)\mathbf{R}^T(\varphi) - \mathbf{Q}(\varphi), \quad \text{(A2}a, b\text{)}$$

$$Q_{ik}(\varphi) = C_{ijks}n_j^* n_s^*, \quad R_{ik}(\varphi) = C_{ijks}n_j^* m_s^*,$$

$$T_{ik}(\varphi) = C_{ijks}m_j^* m_s^*. \quad \text{(A3}a, b, c\text{)}$$

$$\mathbf{n}^* = \mathbf{n}\cos\varphi + \mathbf{m}\sin\varphi,$$

$$\mathbf{m}^* = -\mathbf{n}\sin\varphi + \mathbf{m}\cos\varphi. \quad \text{(A4}a, b\text{)}$$

In (A4), the unit vectors $\mathbf{n}$ and $\mathbf{m}$ are defined in (3.4), i.e.,

$$\mathbf{y} = \rho\mathbf{n}, \quad \mathbf{n} = (n_1, n_2, 0)^T, \quad \mathbf{m} = (0, 0, 1)^T. \quad \text{(A5}a, b, c\text{)}$$

Then, it follows from Ting (1996) that for isotropic materials with Poission's ratio ν and shear modulus μ,

$$\mathbf{S}(\mathbf{n}) = \frac{1-2\nu}{2(1-\nu)}\begin{bmatrix} 0 & 0 & -n_1 \\ 0 & 0 & -n_2 \\ n_1 & n_2 & 0 \end{bmatrix}, \quad \text{(A6)}$$

$$\mathbf{L}(\mathbf{n}) = \frac{\mu}{(1-\nu)}\begin{bmatrix} 1-\nu n_2^2 & \nu n_1 n_2 & 0 \\ \nu n_1 n_2 & 1-\nu n_1^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{(A7)}$$

$$\mathbf{L}^{-1}(\mathbf{n}) = \frac{1}{\mu}\begin{bmatrix} 1-\nu n_1^2 & -\nu n_1 n_2 & 0 \\ -\nu n_1 n_2 & 1-\nu n_2^2 & 0 \\ 0 & 0 & 1-\nu \end{bmatrix}, \quad \text{(A8)}$$

$$\mathbf{S}(\mathbf{n})\mathbf{L}^{-1}(\mathbf{n}) = \frac{1-2\nu}{2\mu}\begin{bmatrix} 0 & 0 & -n_1 \\ 0 & 0 & -n_2 \\ n_1 & n_2 & 0 \end{bmatrix}. \quad \text{(A9)}$$

It then follows from (4.11) that for an isotropic bimaterial,

$$\mathbf{D}^{-1}(\mathbf{n})\mathbf{W}(\mathbf{n}) = \beta\begin{bmatrix} 0 & 0 & -n_1 \\ 0 & 0 & -n_2 \\ n_1 & n_2 & 0 \end{bmatrix}, \quad \text{(A10)}$$

where

$$\beta = \frac{\mu_2(1-2\nu_1) - \mu_1(1-2\nu_2)}{2\mu_2(1-\nu_1) + 2\mu_1(1-\nu_2)} \quad \text{(A11)}$$

is the second Dundurs constant for two-dimensional bimaterials (Dundurs, 1969). Making use of (A10) in (4.10) gives the explicit expressions of $\tilde{\mathbf{L}}(\mathbf{n})$ and $\tilde{\mathbf{W}}(\mathbf{n})$ for isotropic bimaterials.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

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, 345 East 47th Street, New York, N. Y. 10017, 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.

On the Effect of Loading Rate on the Compressive Strength of an AS4/PEEK Composite

T. J. Vogler[1,2] and S. Kyriakides[1,3]

Introduction

It is now widely accepted that the compressive strength of aligned fiber composites is governed by buckling of the fibers in the inelastic matrix. The simplest, yet still realistic, model of the governing events is the zero kink band angle model of Budiansky (1983) shown schematically in Fig. 1 (see also Argon (1972)). The fibers are assumed to be inextensional and uniformly inclined to the direction of the applied stress (σ) by a small angle θ_o. Equilibrium of the element relates the applied stress to the induced shear stress $\tau(\gamma)$ as follows:

$$\sigma = \frac{\tau(\gamma)}{(\theta_o + \gamma)} \tag{1}$$

where γ is the shearing strain also assumed small. The shear response of typical fiber/polymeric matrix composites is highly nonlinear with a relatively low "yield" stress. The interaction between this material nonlinearity and the geometric nonlinearity of the fiber misalignment (denominator of Eq. (1)) results in a maximum value for σ which represents the strength of the composite. Recent studies of the failure mechanism (e.g., Budiansky and Fleck, 1993; Kyriakides et al., 1995; Hsu et al., 1998) have confirmed the validity of this simple model.

It is well known that the shear response of this class of polymeric matrix composites is rate dependent. For example, the rate dependence of the AS4/PEEK composite herein can be seen in the set of pure shear responses from constant strain rate shear tests spanning rates of five decades ($10^{-5} \leq \dot{\gamma} \leq 10^{-1}\ s^{-1}$). In this brief note, we revisit the subject of compressive strength with the purpose of establishing the effect of rate dependence of the composite on the strength. For brevity, we will limit attention to two straight fiber models of the type used by Budiansky and Fleck (1993) (see model classification in Hsu et al. (1998)) in which we incorporate the material rate dependence. Experimentally we examine a composite loaded in compression at prescribed constant end-displacement rate $\dot{\delta}$.

Straight-Inextensional Fibers Model

The incremental form of Eq. (1) is

$$\dot{\sigma} = \frac{\dot{\tau}}{(\theta_o + \gamma)} - \frac{\tau}{(\theta_o + \gamma)^2}\dot{\gamma}. \tag{2}$$

As in Hsu et al. (1998), we assume the composite to shorten elastically ($\epsilon \equiv$ axial strain) but, for simplicity, this deformation will not be coupled to the shear response. For small deformations, the end-shortening rate is then

$$\frac{\dot{\delta}}{L} \approx (\theta_o + \gamma)\dot{\gamma} + \dot{\epsilon}. \tag{3}$$

Due to the very specific nature of this effort, the nonlinearity and rate dependence of the shear response is represented through an elastic power-law viscoplastic model because of the simplicity it affords (see also Schapery (1993) who modeled the material as a viscoelastic solid, and Slaughter et al. (1993) who used a somewhat different viscoplastic model in life prediction calculations). The shearing rate is decomposed into an elastic and an inelastic part

$$\dot{\gamma} = \dot{\gamma}^e + \dot{\gamma}^p. \tag{4}$$

The elastic part is linearly related to the stress rate through

$$\dot{\gamma}^e = \frac{\dot{\tau}}{G_{12}} \tag{5}$$

while the inelastic strain rate obeys the following power law:

$$\left(\frac{\dot{\gamma}^p}{\dot{\gamma}_o}\right)^m = \frac{\tau}{T(\gamma^p)|_{\dot{\gamma}_o}}. \tag{6}$$

Here m is a rate exponent determined experimentally, $\dot{\gamma}_o$ is a base rate, and $T(\gamma^p)$ is the inelastic part of the shear response at the base rate. The model is completed by including the independent, linearly elastic axial stress-strain relationship

[1] Research Center for Mechanics of Solids, Structures & Materials, The University of Texas at Austin, WRW 110, Austin, TX 78712.
[2] Assoc. Mem. ASME.
[3] Fellow ASME.

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, Dec. 1, 1997; final revision, Mar. 2, 1998. Associate Technical Editor: L. T. Wheeler.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

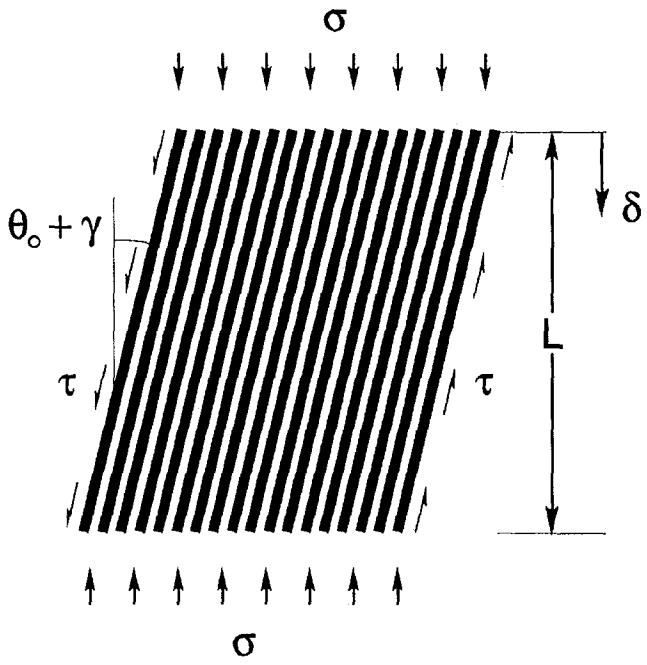


Fig. 1 Idealized composite element used in the straight fiber models

$$\dot{\epsilon} = \frac{\dot{\sigma}}{E_{11}} . \tag{7}$$

The viscoplastic model was calibrated to the shear response at a rate of $\dot{\gamma}_o = 10^{-3}\ \mathrm{s}^{-1}$ which yielded the parameters listed in Table 1. The calibrated model then was used to predict the responses at the other four strain rates. The predictions are seen in Fig. 2 to be in good agreement with the corresponding experimental results.

This formulation was used to calculate the strength of our AS4/PEEK composite compressed at various constant end-displacement rates ($\dot{\delta}/L$). A sample of calculated axial stress-end shortening responses at three different displacement rates is shown in Fig. 3. The specimen analyzed had an initial misalignment angle of $\theta_o = 1.5$ deg. Initially, the inelastic deformation is very limited and the response is almost linear. The response exhibits a limit load, and it is in this neighborhood that inelastic behavior and rate play a role. In this case, we see that increasing the displacement rate by two decades increases the strength approximately 6.7 percent.

Experimental Results and Model Performance

A set of compression experiments were performed on AS4/PEEK composite solid circular rods to test the validity of the model. The rods were machined from the same 0.75-in. thick unidirectional plate as the shear specimens. They were 3.0-in. (76.2-mm) long with a test section diameter of 0.325 in. (8.26 mm). The specimens were mounted in an axisymmetric aligning fixture consisting of two steel end supports which slide inside a stiff aligning collar as shown in the inset in Fig. 4 (see Fig. 1 in Kyriakides et al. (1995)). The tests were performed at room temperature at prescribed end-displacement rates ($\dot{\delta}$). The actual strain rate ($\dot{\epsilon}$) was measured with strain gages in the test section of each specimen ($\dot{\epsilon}$ is somewhat smaller than $\dot{\delta}/L$). Tests were performed at four strain rates spanning five decades ($10^{-6} \le \dot{\epsilon} \le 4 \times 10^{-1}$).

Table 1 Mechanical properties of the composite

E_{11} msi (GPa)	G_{12} ksi (GPa)	v_f	$\dot{\gamma}_o\ \mathrm{s}^{-1}$	m
18.6 (128)	800 (5.52)	0.60	10^{-3}	0.019

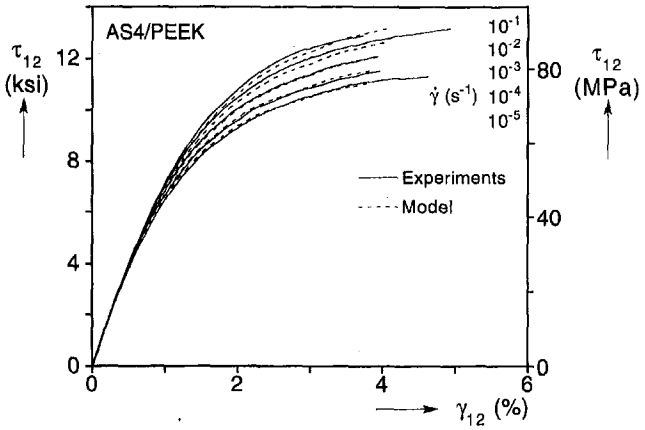


Fig. 2 Measured and predicted composite shear responses at different strain rates

Fiber imperfections are a decisive factor in the strength of such composites. As reported in Kyriakides and Ruff (1997), our composite has two types of imperfections. The first is in the plane of the plies and is introduced during the manufacture of the prepreg. The second is introduced during the lamination process and is in the stacking plane. The two types of imperfections have distinct but different spatial distributions. As a result, the predominant imperfections which end up in the rods vary from specimen to specimen. Consequently, the measured strengths exhibit some scatter. For this reason, we performed at least three compression tests at each of the four strain rates examined.

The fiber imperfections described in Kyriakides and Ruff include small pockets of interlaminar resin-rich regions irregularly distributed in the plate. In addition to their detrimental effect on strength, resin-rich regions also affect the local elastic modulus. In the compression tests performed, a variation in axial modulus (E'_{11}) of approximately ±3 percent from the average value listed in Table 1 was recorded by the local strain gage measurements. To a first approximation, the strength is inversely proportional to $(1 - v_f)$, and for the higher fiber volume fractions of interest here $v_f \approx E'_{11}/E_f$. Thus, in order to reduce the scatter in the measured strengths (σ_C) the data were normalized as follows:

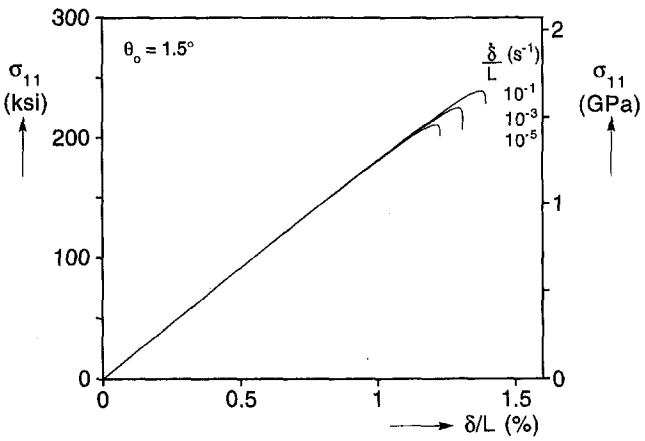


Fig. 3 Calculated axial stress-end shortening responses for the rate sensitive composite with an initial misalignment angle of 1.5 deg

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

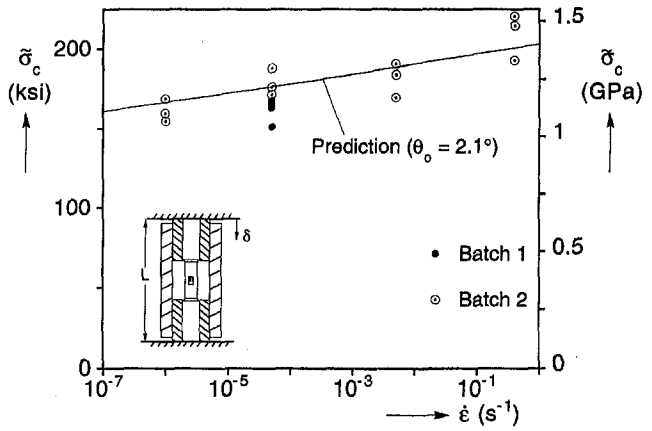


Fig. 4 Measured and predicted compressive strength at various strain rates

$$\tilde{\sigma}_C = \sigma_C \left(\frac{E_f - E'_{11}}{E_f - E_{11}} \right) \tag{8}$$

where E_f is the fiber modulus. The normalized strengths are plotted against strain rate in Fig. 4 (Batch 2). Included in the plot are five tests reported in Kyriakides et al. (1995) using similar specimens from a different plate (Batch 1). Despite the scatter in the data, the strengths exhibit a clear upward trend with rate.

Although the two and three-dimensional bending fiber models we employed before to study the strength of composites can be modified to include the effect of rate, here we limit attention to the simpler straight fiber models. To use such models, an "equivalent" uniform initial fiber misalignment angle must first be established for the composite empirically. The value of θ_o = 2.1 deg was chosen by making the model match the value of $\tilde{\sigma}_C$ = 176 ksi (8.37 GPa) measured in one of the tests conducted at the strain rate of $\dot{\epsilon} = 5 \times 10^{-5}$. With this value of θ_o and the material parameters in Table 1, the composite strength was calculated for the strain rates of interest. The predictions are plotted as a solid line in Fig. 4. The predicted strengths exhibit an upward trend with strain rate which reasonably represents the trend of the experimental results.

The model was then used to conduct similar calculations for 0.3 deg $\leq \theta_o \leq$ 3 deg and for strain rates of 10^{-1}, 10^{-3} and 10^{-5} s^{-1}. The calculated strengths are plotted against θ_o in Fig. 5. The well-known strong imperfection sensitivity of the problem is clearly demonstrated along with the effect of rate. A minor weakness of the present model is that the axial deformation of the fibers is not coupled to that of the matrix. Budiansky and Fleck (1993) addressed this by idealizing the matrix as a J_2 deformation theory solid and the fibers as linearly elastic and using the rule of mixtures to establish the axial and shear response of the composite. Hsu et al. (1998) followed a similar procedure but represented the matrix as a J_2 flow theory elastoplastic solid.

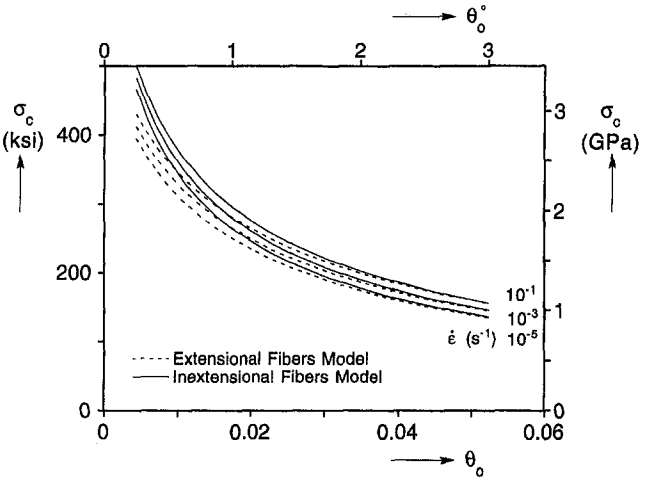


Fig. 5 Composite strength predictions from two models as a function of initial misalignment angle for three strain rates

Table 2 Fiber and matrix properties for two-dimensional extensional fibers composite model

E_f msi (GPa)	ν_f	E_m ksi (GPa)	υ_m	σ_{om} ksi (MPa)	$\dot{\epsilon}_o$ s^{-1}	m
31 (214)	0.263	893 (6.14)	0.356	10.57 (72.9)	10^{-3}	0.020

For completeness, we have modified the formulations in Hsu et al. by making the matrix a power-law viscoplastic solid and repeated the strength calculations in Fig. 5. The properties of the matrix were extracted through the rule of mixtures applied incrementally. The major variables of the two constituents are listed in Table 2. The nonlinear response of the matrix has the same shape as the one in Fig. 2 of Hsu et al. but at the base rate of $\dot{\epsilon}_o = 10^{-3}$ s^{-1} it has a lower yield stress listed in Table 2 ($\sigma_{om} \equiv$ stress at a strain offset of 0.2 percent). The shear responses predicted by this model are very similar to those shown in Fig. 2.

The strengths predicted using this extensional model are drawn in dashed lines in Fig. 5. The trend is similar to that of the results of the inextensional model as is the sensitivity to rate. As was the case for the rate independent material, at lower values of θ_o (<1 deg) the strengths predicted by the extensional model are somewhat lower than those from the inextensional one.

In summary, the nonlinear shear response of polymeric matrix composites exhibits a sensitivity to rate even at room temperature. For the particular time scales considered this rate sensitivity has been shown to be well represented by power-law viscoplastic models. This constitutive model was incorporated in the straight fiber models used for first cut estimates of the compressive strength. These models were shown to yield good estimates of the rate dependence of the compressive strength of the material.

Acknowledgments

The financial support of the Office of Naval Research under contract N00014-91-J-4091 acknowledged with thanks. The work of TJV was also supported by the Fannie and John Hertz foundation and the W. M. Keck Foundation.

References

Argon, A. S., 1972, "Fracture of composites," *Treatise on Material Science and Technology,* Vol. 1, H. Herman, ed., Academic Press, New York, pp. 79–114.

Budiansky, B., 1983, "Micromechanics," *Computers & Structures,* Vol. 16, pp. 3–12.

Budiansky, B., and Fleck, N. A., 1993, "Compressive failure of fiber composites," *J. Mechanics & Physics of Solids,* Vol. 41, pp. 183–211.

Hsu, S.-Y., Vogler, T. J., and Kyriakides, S., 1998, "Compressive Strength Predictions for Fiber Composites," ASME JOURNAL OF APPLIED MECHANICS, Vol. 65 pp. 7–16.

Kyriakides, S., Arseculeratne, R., Perry, E. J., and Liechti, K. M., 1995, "On the compressive failure of fiber reinforced composites," *Int'l J. Solids & Structures* Vol. 32, pp. 689–738.

Kyriakides, S., and Ruff, A. E., 1997, "Aspects of failure and postfailure of fiber composites in compression," *J. Composite Materials,* Vol. 30, No. 20, pp. 2000–2037.

Schapery, R. A., 1993, "Compressive strength and failure time based on local buckling in viscoelastic composites," ASME *Appl. Mech. Reviews,* Vol. 46, No. 11, Pt. 2, pp. S221–S228.

Slaughter, W. S., Fleck, N. A., and Budiansky, B., 1993, "Compressive Failure of Fiber Composites: The Roles of Multiaxial Loading and Creep," ASME *Journal of Engineering Materials and Technology,* Vol. 115, pp. 308–313.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

On the Decomposition of Deformation Into Elastic-Plastic Parts

Li-jun Shen[1]

This paper analyzes the decomposition of finite deformation into elastic and plastic parts. The analysis argues that Lee's multiplicative decomposition of the deformation gradient is not the real separation of the total deformation into elastic and plastic parts and that the Green strain cannot be decomposed into elastic and plastic parts. The paper also analyzes the decomposition of the velocity gradient and presents an exact elastic-plastic decomposition of deformation rate.

1 Introduction

It is generally held that the plastic strain can be defined as some measure of permanent deformation. However, the measure of the plastic strain is still at issue in the case of finite deformations except rigid plastic deformation (see, for example, Naghdi, 1990). The decomposition of deformation into elastic and plastic parts has evoked heated arguments (see Naghdi, 1990). Many have attempted to decompose the finite deformation and to define the plastic strain in terms of an intermediate stress-free configuration. They have held that the deformation gradient $\mathbf{F}$ can be decomposed into elastic (e) and plastic (p) parts, $\mathbf{F}_p$ and $\mathbf{F}_e$ (see Lee, 1969), in the form

$$\mathbf{F} = \mathbf{F}_e \cdot \mathbf{F}_p. \tag{1}$$

If Green strain $\mathbf{E}$, elastic strain $\mathbf{E}_e$, and plastic strain $\mathbf{E}_p$ are, respectively, defined by

$$\mathbf{E} = \tfrac{1}{2}(\mathbf{F}^T \cdot \mathbf{F} - \mathbf{I}), \quad \mathbf{E}_e = \tfrac{1}{2}(\mathbf{F}_e^T \cdot \mathbf{F}_e - \mathbf{I}),$$

$$\mathbf{E}_p = \tfrac{1}{2}(\mathbf{F}_p^T \cdot \mathbf{F}_p - \mathbf{I}), \tag{2a, b, c}$$

we have

$$\mathbf{E} = \mathbf{E}_p + \mathbf{F}_p^T \cdot \mathbf{E}_e \cdot \mathbf{F}_p \tag{3}$$

where the super index T denotes the transpose and $\mathbf{I}$ is the identify tensor. Obviously the total strain is not equal to the sum of the elastic and plastic strains. If velocity gradient $\mathbf{L}$, elastic and plastic velocity gradients $\mathbf{L}_e$ and $\mathbf{L}_p$ are, respectively, defined by

$$\mathbf{L} = \dot{\mathbf{F}} \cdot \mathbf{F}^{-1}, \quad \mathbf{L}_e = \dot{\mathbf{F}}_e \cdot \mathbf{F}_e^{-1}, \quad \mathbf{L}_p = \dot{\mathbf{F}}_p \cdot \mathbf{F}_p^{-1}, \tag{4a, b, c}$$

where a superposed dot denotes the material time derivative and "-1" denotes the inverse of tensor, we obtain

$$\mathbf{L} = \mathbf{L}_e + \mathbf{F}_e \cdot \mathbf{L}_p \cdot \mathbf{F}_e^{-1}. \tag{5}$$

In the analysis of rate-type elastic-plastic constitutive equations, many usually assume that the velocity gradient $\mathbf{L}$ or the deformation rate $\mathbf{D}$ (the symmetric part of velocity gradient) is decomposed into elastic and plastic parts $\mathbf{L}_e$ and $\mathbf{L}_p$ or $\mathbf{D}_e$ and $\mathbf{D}_p$, respectively, in the following forms:

$$\mathbf{L} = \mathbf{L}_e + \mathbf{L}_p, \quad \mathbf{D} = \mathbf{D}_e + \mathbf{D}_p, \tag{6a, b}$$

and they postulate the constitutive equations for $\mathbf{D}_p$ and $\mathbf{D}_e$ (see, for example, Naghdi, 1990). But the relations (6a, b) are valid only if the elastic deformation is negligible in comparison with the plastic deformation (see Metzger and Dubey, 1987). Many (see, for example, Metzger and Dubey, 1987) have believed that an error is perhaps introduced into the constitutive equations by the assumption (6b). The objectivity under superposed rigid-body motions (s.r.b.m) of the $\mathbf{F}$, $\mathbf{F}_e$, $\mathbf{F}_p$, $\mathbf{E}_e$, and $\mathbf{E}_p$ is also at issue (see Naghdi, 1990). Casey and Naghdi (1980) have supposed that superposing rigid-body motions into elastic-plastic deformations leads to transformations $\mathbf{F} \to \mathbf{F}^+$, $\mathbf{F}_e \to \mathbf{F}_e^+$ and $\mathbf{F}_p \to \mathbf{F}_p^+$ with

$$\mathbf{F}^+ = \mathbf{Q} \cdot \mathbf{F}, \quad \mathbf{F} = \mathbf{F}_e^+ \cdot \mathbf{F}_p^+ \tag{7a, b}$$

and

$$\mathbf{F}_e^+ = \mathbf{Q} \cdot \mathbf{F}_e \cdot \bar{\mathbf{Q}}^T, \quad \mathbf{F}_p^+ = \bar{\mathbf{Q}} \cdot \mathbf{F}_p \tag{8a, b}$$

where $\mathbf{Q} = \mathbf{Q}(t)$ and $\bar{\mathbf{Q}} = \bar{\mathbf{Q}}(t)$ are proper orthogonal tensor-valued functions of time and are different. That is to say, rigid-body rotations independent of the total rigid-body motion can be superposed into the intermediate stress-free configuration. The above puzzling problems will be clarified in this note.

2 The Decomposition of Finite Deformation Into Elastic and Plastic Parts

First we analyze the multiplicative decomposition of the deformation gradient through the simple tension of elastic-plastic materials. As space is limited, the lateral contraction is neglected, and the motion in this deformation is given by

$$x_1 = (1 + \delta/l)X_1, \quad x_2 = X_2, \quad x_3 = X_3 \tag{9a, b, c}$$

where x_i ($i = 1, 2, 3$) and X_A ($A = 1, 2, 3$) are, respectively, the rectangular Cartesian coordinates of the current and initial positions of a material point and the total elongation δ is equal to the sum of the elastic and plastic parts e and p, i.e., $\delta = e + p$. We obtain the component matrix of the deformation gradient (only the diagonal elements are written out when other elements are zero)

$$\mathbf{F} = [\partial x_i / X_A] = ((l + \delta)/l, 1, 1)$$

$$(i = 1, 2, 3; A = 1, 2, 3) \tag{10}$$

and Lee's decomposition $\mathbf{F} = \mathbf{F}_e \cdot \mathbf{F}_p$ with

$$\mathbf{F}_e = ((l + \delta)/(l + p), 1, 1),$$

$$\mathbf{F}_p = ((l + p)/l, 1, 1). \tag{11a, b}$$

So in terms of the definitions of Eqs. (2a, b, c) the Green strain, so-called the elastic strain and the plastic strain are, respectively,

$$\mathbf{E} = \tfrac{1}{2}(((l + \delta)/l)^2 - 1, 0, 0) \tag{12}$$

and

$$\mathbf{E}_e = \tfrac{1}{2}(((l + \delta)/(l + p))^2 - 1, 0, 0),$$

$$\mathbf{E}_p = \tfrac{1}{2}(((l + p)/l)^2 - 1, 0, 0). \tag{13a, b}$$

It is clear that $\mathbf{F}_e$ and $\mathbf{E}_e$ include the plastic elongation p and are not real elastic quantities. In the decomposition $\mathbf{F} = \bar{\mathbf{F}}_p \cdot \bar{\mathbf{F}}_e$ presented by Clifton (1972), $\bar{\mathbf{F}}_p$ and $\bar{\mathbf{E}}_p$ derived from $\bar{\mathbf{F}}_p$ include elastic deformation. In terms of the definitions of Eqs. (4a, b, c) the velocity gradient, the elastic and plastic velocity gradients are, respectively,

[1] Materials Science and Mechanics Research Center, Ningbo University, Ningbo, Zhenjiang 315211, China.

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, Nov. 11, 1997; final revision, Feb. 20, 1998. Associate Technical Editor: R. Abeyaratne.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\mathbf{L} = (\dot{\delta}/(l+\delta), 0, 0) \tag{14}$$

$$\mathbf{L}_p = (\dot{p}/(l+p), 0, 0),$$

$$\mathbf{L}_e = ((\dot{e}(l+\delta) - \dot{\delta}e)/((l+p)(l+\delta)), 0, 0) \tag{15a, b}$$

where $\dot{\delta} = \dot{e} + \dot{p}$. Therefore, $\mathbf{L}_e$ includes the plastic elongation rate $\dot{p}$. It is clear from the foregoing development that Lee's decomposition is not the separation of the total deformation into elastic and plastic parts, but implies that the process of deformation is separated into the processes of elastic and plastic deformations.

Next we decompose the displacement in the form

$$\mathbf{u} = \mathbf{u}_e + \mathbf{u}_p + \mathbf{u}_0 \tag{16}$$

where $\mathbf{u}_e$ and $\mathbf{u}_p$ are, respectively, elastic and plastic displacements and $\mathbf{u}_0$ is independent of the deformation. Let $\mathbf{u}_0 = \mathbf{Q}\cdot\mathbf{r}_0 - \mathbf{r}_0$, in which $\mathbf{r}_0$ is the radius vector at initial time. This decomposition is also obtained in terms of an intermediate stress-free configuration. The total strain, the elastic strain, and the plastic strain in terms of the Green definition are, respectively,

$$\mathbf{E} = \tfrac{1}{2}[\mathbf{u}\overset{<}{\nabla} + \overset{<}{\nabla}\mathbf{u} + (\overset{<}{\nabla}\mathbf{u})\cdot(\mathbf{u}\overset{<}{\nabla})] \tag{17}$$

$$\mathbf{E}_e = \tfrac{1}{2}[\mathbf{u}_e\overset{<}{\nabla} + \overset{<}{\nabla}\mathbf{u}_e + (\overset{<}{\nabla}\mathbf{u}_e)\cdot(\mathbf{u}_e\overset{<}{\nabla})],$$

$$\mathbf{E}_p = \tfrac{1}{2}[\mathbf{u}_p\overset{<}{\nabla} + \overset{<}{\nabla}\mathbf{u}_p + (\overset{<}{\nabla}\mathbf{u}_p)\cdot(\mathbf{u}_p\overset{<}{\nabla})]. \tag{18a, b}$$

We obtain

$$\mathbf{E} = \mathbf{E}_e + \mathbf{E}_p + (\overset{<}{\nabla}\mathbf{u}_e)\cdot(\mathbf{u}_p\overset{<}{\nabla}) \tag{19}$$

where $\overset{<}{\nabla}$ is Hamiltonian operator with respect to the initial (reference) configuration. The end term of Eq. (19) cannot be expressed as the sum of elastic and plastic parts since it is a product of elastic and plastic variables. If the Hamiltonian operator in Eq. (18*a*) is one with respect to the intermediate stress-free configuration, Eqs. (18*a, b*) are consistent with Eqs. (2*b, c*).

In the case of simple tension, we have the displacements in the direction of X_1,

$$u = \frac{\delta}{l}X_1, \quad u_e = \frac{e}{l}X_1, \quad u_p = \frac{p}{l}X_1, \tag{20}$$

and principal strains in the direction of X_1,

$$E = \frac{1}{2}\left[2\frac{\delta}{l} + \left(\frac{\delta}{l}\right)^2\right], \quad E_e = \frac{1}{2}\left[2\frac{e}{l} + \left(\frac{e}{l}\right)^2\right],$$

$$E_p = \frac{1}{2}\left[2\frac{p}{l} + \left(\frac{p}{l}\right)^2\right], \tag{21a, b, c}$$

which yield

$$E = E_e + E_p + \frac{ep}{l^2}. \tag{22}$$

Thus the strain cannot be decomposed into elastic and plastic parts owing to its nonlinear term.

However, the velocity can be decomposed in the form

$$\mathbf{v} = \mathbf{v}_e + \mathbf{v}_p + \mathbf{v}_0 \tag{23}$$

where $\mathbf{v}_0 = \mathbf{\Omega}\cdot\mathbf{r}$, in which $\mathbf{\Omega}$ is the rotational rate independent of the deformation, and $\mathbf{r}$ is the radius vector at current time. Equation (23) can be derived from Eq. (16) since $\mathbf{Q} \to \mathbf{I}$, $\dot{\mathbf{Q}} \to \mathbf{\Omega}$, $\mathbf{r} \to \mathbf{r}_0$, $\dot{\mathbf{u}}_0 = \dot{\mathbf{Q}}\cdot\mathbf{r}_0 = \mathbf{v}_0$, $\dot{\mathbf{u}}_e = \mathbf{v}_e$ and $\dot{\mathbf{u}}_p = \mathbf{v}_p$ when the current configuration acts as reference one. This scheme does not utilize any intermediate stress-free configuration. The velocity gradient $\mathbf{L}$, elastic and plastic velocity gradients $\mathbf{L}_e$ and $\mathbf{L}_p$ are, respectively, defined by

$$\mathbf{L} = \mathbf{v}\overset{>}{\nabla}, \quad \mathbf{L}_e = \mathbf{v}_e\overset{>}{\nabla}, \quad \mathbf{L}_p = \mathbf{v}_p\overset{>}{\nabla} \tag{24}$$

where the notation $\overset{>}{\nabla}$ denotes the Hamiltonian operator with respect to the current configuration. We get the decomposition of velocity gradient, spin, and deformation rate in the form

$$\mathbf{L} = \mathbf{L}_e + \mathbf{L}_p + \mathbf{\Omega}, \quad \mathbf{w} = \mathbf{w}_e + \mathbf{w}_p + \mathbf{\Omega},$$

$$\mathbf{D} = \mathbf{D}_e + \mathbf{D}_p. \tag{25a, b, c}$$

We further analyze the transformation under s.r.b.m of $\mathbf{L}$, $\mathbf{D}$, and $\mathbf{w}$. The transformation of the radius vector is $\mathbf{r} \to \mathbf{r}^+$ (super index "+" denotes the variable under s.r.b.m) with

$$\mathbf{r}^+ = \mathbf{b} + \mathbf{Q}\cdot\mathbf{r} \tag{26}$$

where $\mathbf{b}$ is a vector-valued function of time and $\mathbf{Q}$ is a proper orthogonal tensor-valued function of time, which represents rigid-body motion. The transformation of velocity is

$$\mathbf{v}^+ = \mathbf{Q}\cdot\mathbf{v} + \dot{\mathbf{Q}}\cdot\mathbf{r} + \dot{\mathbf{b}} \tag{27}$$

and the transformations of velocity gradient, spin, and deformation rate are, respectively,

$$\mathbf{L}^+ = \mathbf{Q}\cdot\mathbf{L}\cdot\mathbf{Q}^T + \mathbf{\Omega}^Q, \quad \mathbf{w}^+ = \mathbf{Q}\cdot\mathbf{w}\cdot\mathbf{Q}^T + \mathbf{\Omega}^Q,$$

$$\mathbf{D}^+ = \mathbf{Q}\cdot\mathbf{D}\cdot\mathbf{Q}^T \tag{28a, b, c}$$

where $\mathbf{\Omega}^Q = \dot{\mathbf{Q}}\cdot\mathbf{Q}^T$. Equation (28*c*) can be rewritten as

$$\mathbf{D}^+ = \mathbf{D}_e^+ + \mathbf{D}_p^+ \tag{29}$$

where

$$\mathbf{D}_e^+ = \mathbf{Q}\cdot\mathbf{D}_e\cdot\mathbf{Q}^T, \quad \mathbf{D}_p^+ = \mathbf{Q}\cdot\mathbf{D}_p\cdot\mathbf{Q}^T. \tag{30a, b}$$

The rigid-body rotation $\mathbf{Q}$ affects the elastic deformation and the plastic deformation simultaneously. From Eqs. (25*c*) and (29) we find that the deformation rate can be decomposed into the elastic and plastic parts. The spin may include a rotational rate independent of deformation.

In the case of simple tension, we obtain

$$\mathbf{L} = \mathbf{D} = (\dot{\delta}/(l+\delta), 0, 0) \tag{31}$$

$$\mathbf{L}_e = \mathbf{D}_e = (\dot{e}/(l+\delta), 0, 0),$$

$$\mathbf{L}_p = \mathbf{D}_p = (\dot{p}/(l+\delta), 0, 0) \tag{32}$$

$$\mathbf{w} = \mathbf{\Omega} = (0, 0, 0). \tag{33}$$

However, $\mathbf{D}_e$ and $\mathbf{D}_p$ here are different from elastic and plastic deformation rates derived from $\mathbf{F} = \mathbf{F}_e\cdot\mathbf{F}_p$ (see Eq. (15*a, b*)).

3. Discussion and Conclusion

The contradiction caused by $\mathbf{F} = \mathbf{F}_e\cdot\mathbf{F}_p$ had been better settled here. In this decomposition the elastic and plastic variables are not interchangeable, that is, $\mathbf{F}_e\cdot\mathbf{F}_p \neq \mathbf{F}_p\cdot\mathbf{F}_e$. The transformations under s.r.b.m of $\mathbf{F}_e$ and $\mathbf{F}_p$ are different, and the rigid-body rotation $\mathbf{Q}$ does not simultaneously affect the elastic deformation and the plastic deformation. Lee's decomposition implies that the elastic deformation and plastic deformation are not simultaneous but follow one another. We find from Eqs. (23) and (25) that the elastic and plastic variables play an equal role in the decomposition of the deformation and may be interchanged. These equations of decomposition imply that the elastic and plastic deformations are simultaneous. Therefore, these equations of decomposition are rational. It has been proved that the exact decomposition $\mathbf{D} = \mathbf{D}_e + \mathbf{D}_p$ exists. So the possibility is removed that error is introduced into the rate-type constitutive relation in the

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

process of separating the total deformation into elastic and plastic parts. It is found that the strain cannot generally be decomposed into the elastic and plastic strain in terms of the Green strain. The spin is, in general, not the sum of the elastic and plastic parts and possibly includes a rigid-body rotational rate.

Acknowledgment

I would like to thank the reviewers for their constructive and useful comments.

References

Casey, J., and Naghdi, P. M., 1980, "A Remark on the use of the Decomposition $\mathbf{F} = \mathbf{F}_e\mathbf{F}_p$ in Plasticity," ASME JOURNAL OF APPLIED MECHANICS, Vol. 47, pp. 672–657.

Clifton, R. J., 1972, "On the Equivalence of $\mathbf{F}^e\mathbf{F}^p$ and $\bar{\mathbf{F}}^p\bar{\mathbf{F}}^e$," ASME JOURNAL OF APPLIED MECHANICS, pp. 287–289.

Lee, E. H., 1969, "Elastic-plastic deformation at finite strains," ASME JOURNAL OF APPLIED MECHANICS, Vol. 36, pp. 1–6.

Metzger, D. R., and Dubey, R. N., 1987, "Corotational rates in constitutive modeling of elastic-plastic deformation," *Int. J. Plasticity*, Vol. 4, pp. 341–368.

Naghdi, P. M., 1990, "A critical review of the state of finite plasticity," *J. Appl. Math. Phys. (ZAMP)*, Vol. 41, pp. 315–393.

Some Comments on Rocking Motion on Rotating Disk/Spindle Systems

I. Y. Shen[1]

This note is a follow-up parametric study of the mathematical model developed by Shen and Ku (1997) and Shen (1997) addressing free and forced vibration of rotating disk/spindle systems used in computer hard-disk drives and other rotary machinery. The present study focuses on rocking modes of rotating disk/spindle systems, which are of great importance in computer disk drive applications. Additionally, this note examines the influence of disk damping on these modes.

For disk drives with increased track density and low-track misregistration budget, the presence of rocking modes is very undesirable because they cause relative in-plane motion between the spinning disks and the flying heads substantially contributing to the track misregistration. To suppress the rocking modes, the disk-drive industry is developing spindle motors with hydrodynamic bearings. Motivated by the trend in the disk-drive industry, this note studies, in the form of two numerical examples, natural frequencies and mode shapes of rocking modes and also the possibility to suppress the rocking modes through use of disk damping (e.g., laminated disks).

The first example demonstrates mode shapes and natural frequencies of six unbalanced (0, 1) modes governing rocking motion of a rotating disk/spindle system. The dimensions of the disk/spindle systems are shown in Table 1 of Shen and Ku (1997), and the disks are all undamped. The natural frequencies and mode shapes are obtained from (50) and (54) of Shen and Ku (1997), respectively. Figure 1 shows the natural frequencies with respect to the rotational speed and the corresponding mode shapes, where B and F represent backward and forward precession. According to Fig. 1, there are three pairs of unbalanced (0, 1) modes. For (1B, 1F), the spindle and the disks vibrate with the same phase and present little centroidal translation. For (2B, 2F) and (3B, 3F), the spindle and the disks vibrate with 180 deg out of phase. In addition, (2B, 2F) present little centroidal translation as opposed to (3B, 3F) with significant centroidal translation.

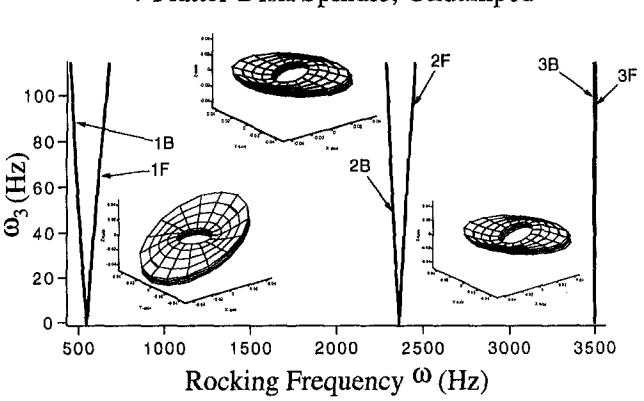


Fig. 1 Resonance plot and mode shapes of the unbalanced (0, 1) modes against rotational speed

The second example demonstrates how disk damping affects modal dampings of unbalanced (0, 1) modes. Consider the same disk/spindle system, except that the disk damping is $c = 54$ Ns/m^3 and bearing damping is $c_a = c_b = 116$ Ns/m. These numbers for disk and bearing dampings have been determined through a series of experiments (Shumway, 1997). They correspond to 0.32 percent damping of one-nodal-diameter stationary disk modes. Figure 2 shows the modal dampings of unbalanced (0, 1) modes with respect to the rotational speed. These modal dampings, defined as the ratio of the real part to the imaginary part of eigenvalues, are obtained from (54) of Shen (1997). There are several things worth noting in Fig. 2. First, (1B, 1F) have the least damping and (3B, 3F) have the largest damping.

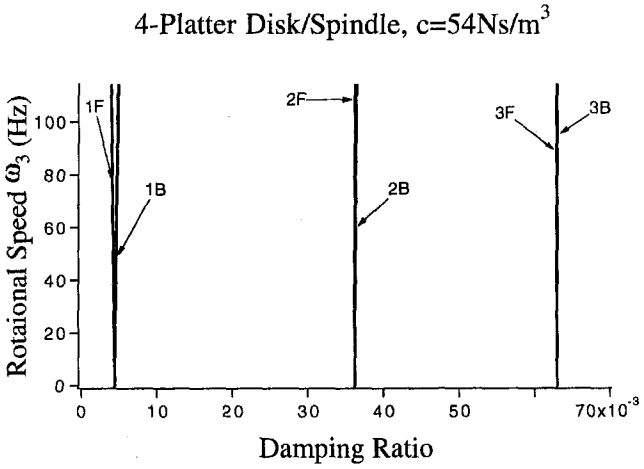


Fig. 2 Damping plot of the unbalanced (0, 1) modes against ω_3 when c = 54 Ns/m^3

[1] Associate Professor, Department of Mechanical Engineering, University of Washington, Seattle, WA 98195-2600. Mem. ASME.

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 22, 1997; final revision, June 18, 1998. Associate Technical Editor: A. A. Ferri.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

process of separating the total deformation into elastic and plastic parts. It is found that the strain cannot generally be decomposed into the elastic and plastic strain in terms of the Green strain. The spin is, in general, not the sum of the elastic and plastic parts and possibly includes a rigid-body rotational rate.

Acknowledgment

I would like to thank the reviewers for their constructive and useful comments.

References

Casey, J., and Naghdi, P. M., 1980, "A Remark on the use of the Decomposition $\mathbf{F} = \mathbf{F}_e\mathbf{F}_p$ in Plasticity," ASME JOURNAL OF APPLIED MECHANICS, Vol. 47, pp. 672–657.

Clifton, R. J., 1972, "On the Equivalence of $\mathbf{F}^e\mathbf{F}^p$ and $\bar{\mathbf{F}}^p\bar{\mathbf{F}}^e$," ASME JOURNAL OF APPLIED MECHANICS, pp. 287–289.

Lee, E. H., 1969, "Elastic-plastic deformation at finite strains," ASME JOURNAL OF APPLIED MECHANICS, Vol. 36, pp. 1–6.

Metzger, D. R., and Dubey, R. N., 1987, "Corotational rates in constitutive modeling of elastic-plastic deformation," *Int. J. Plasticity*, Vol. 4, pp. 341–368.

Naghdi, P. M., 1990, "A critical review of the state of finite plasticity," *J. Appl. Math. Phys. (ZAMP)*, Vol. 41, pp. 315–393.

Some Comments on Rocking Motion on Rotating Disk/Spindle Systems

I. Y. Shen[1]

This note is a follow-up parametric study of the mathematical model developed by Shen and Ku (1997) and Shen (1997) addressing free and forced vibration of rotating disk/spindle systems used in computer hard-disk drives and other rotary machinery. The present study focuses on rocking modes of rotating disk/spindle systems, which are of great importance in computer disk drive applications. Additionally, this note examines the influence of disk damping on these modes.

For disk drives with increased track density and low-track misregistration budget, the presence of rocking modes is very undesirable because they cause relative in-plane motion between the spinning disks and the flying heads substantially contributing to the track misregistration. To suppress the rocking modes, the disk-drive industry is developing spindle motors with hydrodynamic bearings. Motivated by the trend in the disk-drive industry, this note studies, in the form of two numerical examples, natural frequencies and mode shapes of rocking modes and also the possibility to suppress the rocking modes through use of disk damping (e.g., laminated disks).

The first example demonstrates mode shapes and natural frequencies of six unbalanced (0, 1) modes governing rocking motion of a rotating disk/spindle system. The dimensions of the disk/spindle systems are shown in Table 1 of Shen and Ku (1997), and the disks are all undamped. The natural frequencies and mode shapes are obtained from (50) and (54) of Shen and Ku (1997), respectively. Figure 1 shows the natural frequencies with respect to the rotational speed and the corresponding mode shapes, where B and F represent backward and forward precession. According to Fig. 1, there are three pairs of unbalanced (0, 1) modes. For (1B, 1F), the spindle and the disks vibrate with the same phase and present little centroidal translation. For (2B, 2F) and (3B, 3F), the spindle and the disks vibrate with 180 deg out of phase. In addition, (2B, 2F) present little centroidal translation as opposed to (3B, 3F) with significant centroidal translation.

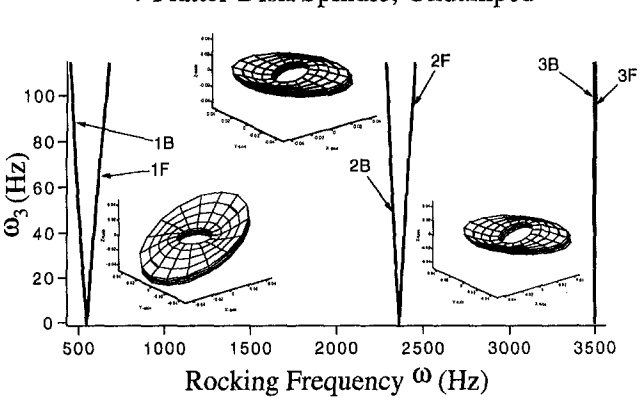


Fig. 1 Resonance plot and mode shapes of the unbalanced (0, 1) modes against rotational speed

The second example demonstrates how disk damping affects modal dampings of unbalanced (0, 1) modes. Consider the same disk/spindle system, except that the disk damping is $c = 54$ Ns/m^3 and bearing damping is $c_a = c_b = 116$ Ns/m. These numbers for disk and bearing dampings have been determined through a series of experiments (Shumway, 1997). They correspond to 0.32 percent damping of one-nodal-diameter stationary disk modes. Figure 2 shows the modal dampings of unbalanced (0, 1) modes with respect to the rotational speed. These modal dampings, defined as the ratio of the real part to the imaginary part of eigenvalues, are obtained from (54) of Shen (1997). There are several things worth noting in Fig. 2. First, (1B, 1F) have the least damping and (3B, 3F) have the largest damping.

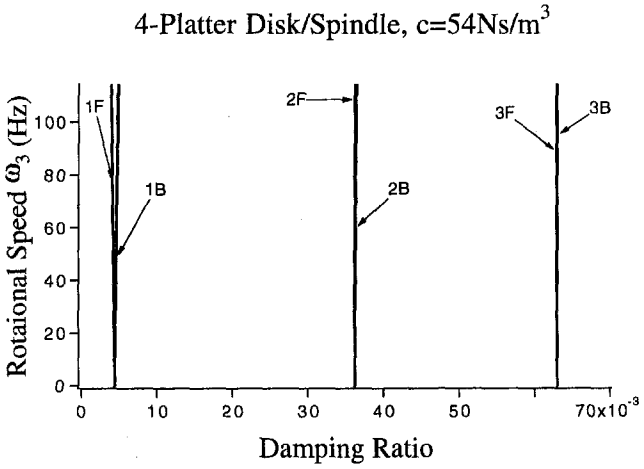


Fig. 2 Damping plot of the unbalanced (0, 1) modes against ω_3 when c = 54 Ns/m^3

[1] Associate Professor, Department of Mechanical Engineering, University of Washington, Seattle, WA 98195-2600. Mem. ASME.

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 22, 1997; final revision, June 18, 1998. Associate Technical Editor: A. A. Ferri.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

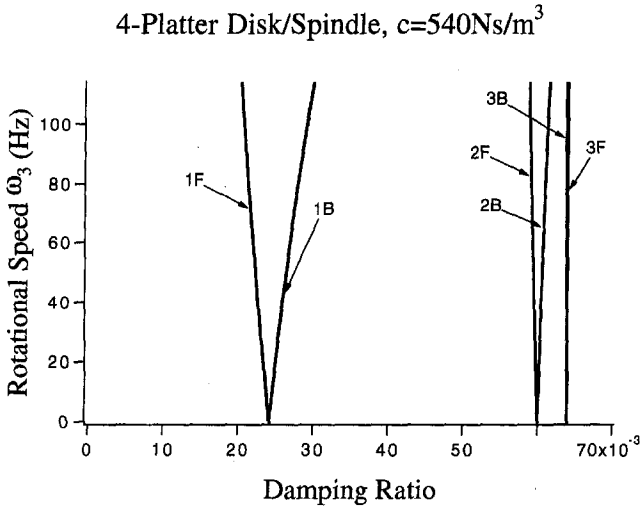


Fig. 3 Damping plot of the unbalanced (0, 1) modes against $\omega_3 c = 540$ Ns/m^3

Second, the backward precession presents slightly larger damping than the forward precession. Third, the damping of (1B, 1F) is about 0.44 percent, which is slightly higher than that of the disks (cf. 0.32 percent). Because of the small damping of (1B, 1F), disk/spindle systems usually present appreciable in-plane vibration.

Figure 3 shows the damping ratios of unbalanced (0, 1) modes, if the disk damping is increased 10 times to $c = 540$ Ns/m^3 while the bearing damping remains unchanged. The damping of (1B, 1F) is increased to 2.4 percent and (2B, 2F) to 6.0 percent. Note that the increase in damping for (3B, 3F) is very little. This may result from the fact that (3B, 3F) have strong in-plane translational components, whose damping is not increased because the bearing damping remains unchanged. This numerical example shows that damping of unbalanced (0, 1) modes can be increased by enhancing disk dampings.

To summarize, this parametric study shows natural frequencies, mode shapes, and damping ratios of three pairs of unbalanced (0, 1) modes. Each pair contains a backward and a forward precession mode. The first pair is in-phase rocking, the second pair is out-of-phase rocking, and the third pair has a strong in-plane translational component. As a result, the first pair has the least damping and the third pair has the greatest damping. Enhanced disk damping will increase damping of unbalanced (0, 1) modes and reduce in-plane disk/spindle vibration.

Acknowledgments

The material is based upon work supported by the National Science Foundation under Grant No. CMS-9634557.

References

Shen, I. Y., and Ku, C.-P. Roger, 1997, "A Non-Classical Vibration Analysis of Multiple Rotating Disks/Spindle Assembly," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 165–174.

Shen, I. Y., 1997, "Closed-Form Forced Response of a Damped, Rotating, Multiple Disk/Spindle System," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 343–352.

Shumway, M. L., 1997, *An Experimental Study of Damping in a Rotating Disk and Spindle System,* Master's thesis, University of Washington, Seattle, WA.

Stability of Gyroscopic Systems Near Vanishing Eigenvalues

A. A. Renshaw[1]

Renshaw and Mote (1996) proposed a conjecture concerning the growth of vibrating eigensolutions of gyroscopic systems in the neighborhood of a vanishing eigenvalue when the system operators depend on an independent system parameter. Although the conjecture was not proved, it was supported by several examples drawn from well-known continuous physical systems. Lancaster and Kliem (1997), however, recently presented three two-degree-of-freedom counter examples. Unlike the examples tested by Renshaw and Mote (1996), these counter examples lack a definiteness property that is usually found in models derived from physical systems which appears to be essential to the conjecture. This Brief Note revises the original conjecture to include this definiteness criterion and proves the conjecture for general two-degree-of-freedom systems.

1 Introduction

Renshaw and Mote (1996) proposed a conjecture for the growth of vibrating eigensolutions of gyroscopic systems in the neighborhood of a vanishing eigenvalue when the system operators depend on an independent system parameter. The conjecture used a solvability condition to predict whether or not the eigenvalue locus of an eigensolution with a zero eigenvalue has a positive real part as the system parameter is either increased or decreased. The primary value of the conjecture was that it often enabled system design changes to be assessed without explicitly calculating the eigensolutions. Although the conjecture was not proved, it was supported by several examples drawn from well-known continuous physical systems. Lancaster and Kliem (1997), however, recently presented three two-degree-of-freedom counter examples to the conjecture.

Unlike the examples tested by Renshaw and Mote (1996), these counter examples lack a definiteness property that is usually found in models derived from physical systems which appears to be essential to the conjecture. The purpose of this brief note is to revise the original conjecture to include this definiteness criterion. The brief note also proves the conjecture for general two-degree-of-freedom systems. The new conjecture retains the utility of the old conjecture since it can still often be applied without explicitly calculating the eigensolutions. The new definiteness criterion is satisfied by many gyroscopic systems, particularly those that derive their gyroscopic properties from a repeated application of the material derivative. Such systems include the translating and rotating systems on which the conjecture was originally tested.

2 The Stability Conjecture

Consider the general gyroscopic eigenvalue problem defined by

[1] Department of Mechanical Engineering, Columbia University, M/C 4703, New York, NY 10027. Mem. ASME.

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, Dec. 8, 1997; final revision, Apr. 25, 1998. Associate Technical Editor: N. C. Perkins.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

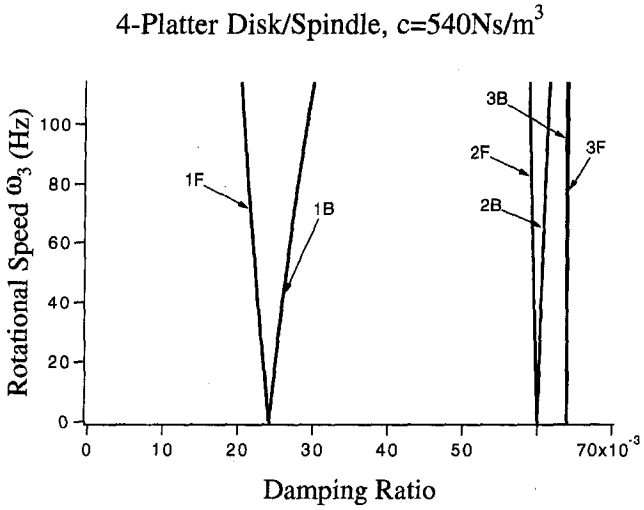


Fig. 3 Damping plot of the unbalanced (0, 1) modes against $\omega_3 c = 540$ Ns/m^3

Second, the backward precession presents slightly larger damping than the forward precession. Third, the damping of (1B, 1F) is about 0.44 percent, which is slightly higher than that of the disks (cf. 0.32 percent). Because of the small damping of (1B, 1F), disk/spindle systems usually present appreciable in-plane vibration.

Figure 3 shows the damping ratios of unbalanced (0, 1) modes, if the disk damping is increased 10 times to $c = 540$ Ns/m^3 while the bearing damping remains unchanged. The damping of (1B, 1F) is increased to 2.4 percent and (2B, 2F) to 6.0 percent. Note that the increase in damping for (3B, 3F) is very little. This may result from the fact that (3B, 3F) have strong in-plane translational components, whose damping is not increased because the bearing damping remains unchanged. This numerical example shows that damping of unbalanced (0, 1) modes can be increased by enhancing disk dampings.

To summarize, this parametric study shows natural frequencies, mode shapes, and damping ratios of three pairs of unbalanced (0, 1) modes. Each pair contains a backward and a forward precession mode. The first pair is in-phase rocking, the second pair is out-of-phase rocking, and the third pair has a strong in-plane translational component. As a result, the first pair has the least damping and the third pair has the greatest damping. Enhanced disk damping will increase damping of unbalanced (0, 1) modes and reduce in-plane disk/spindle vibration.

Acknowledgments

The material is based upon work supported by the National Science Foundation under Grant No. CMS-9634557.

References

Shen, I. Y., and Ku, C.-P. Roger, 1997, "A Non-Classical Vibration Analysis of Multiple Rotating Disks/Spindle Assembly," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 165–174.

Shen, I. Y., 1997, "Closed-Form Forced Response of a Damped, Rotating, Multiple Disk/Spindle System," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 343–352.

Shumway, M. L., 1997, *An Experimental Study of Damping in a Rotating Disk and Spindle System,* Master's thesis, University of Washington, Seattle, WA.

Stability of Gyroscopic Systems Near Vanishing Eigenvalues

A. A. Renshaw[1]

Renshaw and Mote (1996) proposed a conjecture concerning the growth of vibrating eigensolutions of gyroscopic systems in the neighborhood of a vanishing eigenvalue when the system operators depend on an independent system parameter. Although the conjecture was not proved, it was supported by several examples drawn from well-known continuous physical systems. Lancaster and Kliem (1997), however, recently presented three two-degree-of-freedom counter examples. Unlike the examples tested by Renshaw and Mote (1996), these counter examples lack a definiteness property that is usually found in models derived from physical systems which appears to be essential to the conjecture. This Brief Note revises the original conjecture to include this definiteness criterion and proves the conjecture for general two-degree-of-freedom systems.

1 Introduction

Renshaw and Mote (1996) proposed a conjecture for the growth of vibrating eigensolutions of gyroscopic systems in the neighborhood of a vanishing eigenvalue when the system operators depend on an independent system parameter. The conjecture used a solvability condition to predict whether or not the eigenvalue locus of an eigensolution with a zero eigenvalue has a positive real part as the system parameter is either increased or decreased. The primary value of the conjecture was that it often enabled system design changes to be assessed without explicitly calculating the eigensolutions. Although the conjecture was not proved, it was supported by several examples drawn from well-known continuous physical systems. Lancaster and Kliem (1997), however, recently presented three two-degree-of-freedom counter examples to the conjecture.

Unlike the examples tested by Renshaw and Mote (1996), these counter examples lack a definiteness property that is usually found in models derived from physical systems which appears to be essential to the conjecture. The purpose of this brief note is to revise the original conjecture to include this definiteness criterion. The brief note also proves the conjecture for general two-degree-of-freedom systems. The new conjecture retains the utility of the old conjecture since it can still often be applied without explicitly calculating the eigensolutions. The new definiteness criterion is satisfied by many gyroscopic systems, particularly those that derive their gyroscopic properties from a repeated application of the material derivative. Such systems include the translating and rotating systems on which the conjecture was originally tested.

2 The Stability Conjecture

Consider the general gyroscopic eigenvalue problem defined by

[1] Department of Mechanical Engineering, Columbia University, M/C 4703, New York, NY 10027. Mem. ASME.

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, Dec. 8, 1997; final revision, Apr. 25, 1998. Associate Technical Editor: N. C. Perkins.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\lambda^2 Mu + \lambda Gu + Ku = 0 \quad (1)$$

where M, G, and K are real, linear differential (matrix) operators, λ is the eigenvalue and u is the eigenfunction. For all y and z satisfying appropriate boundary conditions, M, G, and K satisfy

$$(My, z) = (y, Mz) \quad (Gy, z) = -(y, Gz) \quad (Ky, z) = (y, Kz) \quad (2)$$

where $(,)$ is the inner product associated with (1). M is positive definite. The stability of separable vibrations of the form $e^{\lambda t}u$ where t is time ($t > 0$) depends on λ. For $\mathrm{Re}[\lambda] > 0$, the separable solution grows with time and is unstable, while for $\mathrm{Re}[\lambda] \leq 0$, $e^{\lambda t}u$ is bounded in time and is therefore stable.

Consider the case where G and K (but not M) depend analytically on an independent parameter α, and denote the eigenvalue loci by $\lambda(\alpha)$. Critical values $\alpha = \alpha_o$ are defined as those values of α at which an eigenvalue vanishes, i.e., $\lambda(\alpha_o) = 0$. The eigenfunctions associated with $\lambda = 0$ are defined by the set S

$$S = \{u: Ku = 0, u \neq 0, \text{plus boundary conditions}\}. \quad (3)$$

Critical α often demarcate the boundary between stable and unstable system response, and rules of thumb that can predict when $\mathrm{Re}[\lambda] > 0$ in the neighborhood of $\alpha = \alpha_o$ without extensive computation can be extremely helpful for design. In this spirit, Renshaw and Mote (1996) proposed the following conjecture:

Original Conjecture:

1 If there exists a $u \in S$ such that $(Gu, v) = 0$ for every $v \in S$, then $\mathrm{Re}[\lambda(\alpha)] > 0$ in the neighborhood of $\alpha = \alpha_o$ for the eigenvalue locus that contains $\lambda(\alpha_o) = 0$.

2 If for every $u \in S$ there exists a $v \in S$ such that $(Gu, v) \neq 0$, then $\mathrm{Re}[\lambda(\alpha)] = 0$ in the neighborhood of $\alpha = \alpha_o$ for the eigenvalue locus that contains $\lambda(\alpha_o) = 0$.

In the first part of the conjecture, $\mathrm{Re}[\lambda(\alpha)] > 0$ may occur when α is either increased or decreased, i.e., on only one side of $\alpha = \alpha_o$ (Seyranian, 1997). Although the original conjecture is true for the continuous gyroscopic systems tested by Renshaw and Mote (1996), the three two-degree-of-freedom counter examples given by Lancaster and Kliem (1997) demonstrate that the conjecture is not true in general.

The original conjecture can be reformulated to account for the counter examples by examining the quadratic equation obtained by taking the inner product of (1) with u (e.g., Lancaster, 1966; Shieh, 1971). The eigenvalues λ are the roots of this equation and are given by

$$\lambda = \frac{i}{2}\left[g \pm \sqrt{g^2 + 4k}\right] \quad (4)$$

where u is normalized by $(Mu, u) = 1$, $g = (iGu, u)$, $k = (Ku, u)$, and $i = \sqrt{-1}$. The symmetries (2) ensure that g and k are real. Vanishing eigenvalues occur when $k = 0$ which define α_o, and $\mathrm{Re}[\lambda(\alpha)] > 0$ occurs only when $g^2 + 4k < 0$. Since g and k are continuous in α, $\mathrm{Re}[\lambda(\alpha)] > 0$ in the neighborhood of $\alpha = \alpha_o$ only occurs when $g = 0$ and $\partial/\partial\alpha[g^2 + 4k] \neq 0$ at $\alpha = \alpha_o$, if that derivative exists. Since the eigenvalue loci $\lambda(\alpha)$ are not necessarily analytic functions of α (Lancaster and Kliem, 1997), this derivative may not exist. If it does, then $\mathrm{Re}[\lambda(\alpha)] > 0$ requires $g = 0$ and $\partial k/\partial\alpha \neq 0$ at $\alpha = \alpha_o$. It is this latter definiteness condition, $\partial k/\partial\alpha \neq 0$ at $\alpha = \alpha_o$, that is true for all the supporting examples used by Renshaw and Mote (1996) and is not true for the three counter examples given by Lancaster and Kliem (1997). Consequently, the original conjecture is reformulated as follows:

Reformulated Conjecture:

1 If there exists a $u \in S$ such that $(Gu, v) = 0$ for every $v \in S$ and $\partial(Ku, u)/\partial\alpha|_{\alpha=\alpha_o} \neq 0$ for every $u \in S$, then $\mathrm{Re}[\lambda(\alpha)] > 0$ in the neighborhood of $\alpha = \alpha_o$ for the eigenvalue locus that contains $\lambda(\alpha_o) = 0$.

2 If for every $u \in S$ there exists a $v \in S$ such that $(Gu, v) \neq 0$ and $\partial(Ku, u)/\partial\alpha|_{\alpha=\alpha_o} \neq 0$ for every $u \in S$, then $\mathrm{Re}[\lambda(\alpha)] = 0$ in the neighborhood of $\alpha = \alpha_o$ for the eigenvalue locus that contains $\lambda(\alpha_o) = 0$.

The reformulated conjecture is still valid for all the examples tested by Renshaw and Mote (1996)—i.e., the axially moving string and beam and the rotating disk with and without transverse spring loading—because $\partial(Ku, u)/\partial\alpha < 0$ for all of these systems. In addition, the conjecture can be proven explicitly for the general two-degree-of-freedom system. However, for general gyroscopic systems, the conjecture remains speculative. To the extent to which more complicated systems behave like the two-degree-of-freedom case, the conjecture will remain valid, but this cannot be expected to be true in general.

3 Proof for Two-Degree-of-Freedom Systems

Consider the general two-degree-of-freedom gyroscopic system

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad G = \begin{bmatrix} 0 & g \\ -g & 0 \end{bmatrix}$$

$$K = \begin{bmatrix} \pm a^2 & ab \\ ab & \pm b^2 \end{bmatrix} + \alpha \begin{bmatrix} a' & c' \\ c' & b' \end{bmatrix} \quad (5)$$

where α is the varying system parameter, the critical value of α is $\alpha_o = 0$, and one uses either the + or − sign only for all $\pm$. The characteristic equation associated with (5) is

$$\lambda^4 + (h + \alpha t)\lambda^2 + \alpha p + \alpha^2 d = 0 \quad (6)$$

where

$$h = g^2 \pm a^2 \pm b^2$$
$$t = a' + b'$$
$$p = \pm a'b^2 \pm b'a^2 - 2c'ab$$
$$d = a'b' - (c')^2. \quad (7)$$

We assume $\partial(Ku, u)/\partial\alpha \neq 0$ so that t is nonzero and $d > 0$ since t is the trace and d is the determinant of the varying part of K.

When $a = b = 0$, then $S = \Re^2$. $(Gu, v) = \pm g$ when u and v are linearly independent so the conjecture predicts $\mathrm{Re}[\lambda] = 0$ in the neighborhood of $\alpha = 0$ when $g \neq 0$ and $\mathrm{Re}[\lambda] > 0$ when $g = 0$. When either a or b is nonzero, $S = \{[\mp b\ a]^T\}$ (i.e., S is one dimensional) and $(Gu, u) = 0$ so the conjecture predicts $\mathrm{Re}[\lambda] > 0$ in the neighborhood of $\alpha = 0$.

In fact, when $a = b = 0$, $p = 0$ and $h = g^2$ so that the explicit solutions of (6) are

$$\lambda^2 = -d\alpha^2/g^2 + O(\alpha^3) \quad \text{and}$$
$$\lambda^2 = -g^2 - \alpha t + d\alpha^2/g^2 + O(\alpha^3) \quad (8)$$

when $g \neq 0$, both of which have $\mathrm{Re}[\lambda] = 0$ in the neighborhood of $\alpha = 0$, and

$$\lambda^2 = \alpha(-t \pm \sqrt{t^2 - 4d})/2 + O(\alpha^4) \quad (9)$$

when $g = 0$, for which $\mathrm{Re}[\lambda] > 0$ in the neighborhood of $\alpha = 0$.

When either a or b is nonzero, the definiteness criterion becomes

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\frac{\partial(Ku, u)}{\partial\alpha} = [\mp b \quad a]\begin{bmatrix} a' & c' \\ c' & b' \end{bmatrix}\begin{bmatrix} \mp b \\ a \end{bmatrix}$$

$$= a'b^2 + b'a^2 \mp 2c'ab = \pm p \qquad (10)$$

which is nonzero by assumption. Explicit solution of (6) then gives

$$\lambda^2 = -p\alpha/h + O(\alpha^2) \quad \text{and}$$

$$\lambda^2 = -h - \alpha(t - p/h) + O(\alpha^2) \qquad (11)$$

when $h \neq 0$ and

$$\lambda^2 = \pm\sqrt{-\alpha p} + O(\alpha^{3/2}) \qquad (12)$$

when $h = 0$, both of which have $\mathrm{Re}[\lambda] > 0$ for at least one eigensolution in the neighborhood of $\alpha = 0$. The preceding arguments remain valid under the substitution $g \to g + \alpha g'$, so the conjecture remains true when G is a linear function of α.

4 Concluding Comments

The original conjecture (Renshaw and Mote, 1996) related the conditions of the conjecture to a solvability condition for the existence of eigensolutions with a nonzero steady velocity. This interpretation remains valid for the reformulated conjecture and is used explicitly in the comments of Lancaster and Kleim (1997).

A number of systems derive their gyroscopic properties from a repeated application of the material derivative to evaluate the acceleration of the material particles comprising the system. For example, the axially moving string and the axially moving beam compute the acceleration of the string or beam as

$$\frac{D^2u}{Dt^2} = \left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial x}\right)^2 u \qquad (13)$$

where u is the displacement, x is the axial coordinate, t is time, and v is the axial velocity. The transverse acceleration of a rotating disk computed in the stationary frame of reference is

$$\frac{D^2u}{Dt^2} = \left(\frac{\partial}{\partial t} + \Omega\frac{\partial}{\partial\theta}\right)^2 u \qquad (14)$$

where Ω is the rotation speed of the disk. For these systems, $\partial(Ku, u)/\partial v < 0$ and $\partial(Ku, u)/\partial\Omega < 0$ because of the form of the material derivative. Consequently, one expects the conjecture to be generally applicable to systems such as these whose gyroscopic properties are derived from a repeated application of the material derivative.

References

Lancaster, P., 1966, *Lambda-matrices and Vibrating Systems,* Pergamon Press, Oxford, UK.

Lancaster, P., and Kliem, W., 1997, "Comments on Stability Properties of Conservative Gyroscopic Systems," ASME JOURNAL OF APPLIED MECHANICS, in press.

Renshaw, A. A., and Mote, Jr., C. D., 1996, "Local Stability of Gyroscopic Systems Near Vanishing Eigenvalues," ASME JOURNAL OF APPLIED MECHANICS, Vol. 63, pp. 116–120.

Seyranian, A. P., 1997, discussion on "Local Stability of Gyroscopic Systems Near Vanishing Eigenvalues," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 720–721.

Shieh, R. C., 1971, "Energy and Variational Principles for Generalized (Gyroscopic) Conservative Problems," *International Journal of Non-Linear Mechanics,* Vol. 5, pp. 495–509.

Linear Relationships Between the Stress Fields of ⟨111⟩ Line Defects

Y. T. Chou[1,2] and Sanboh Lee[1]

Introduction

Although the theory of two-dimensional anisotropic elasticity has been studied extensively (Eshelby et al., 1953; Stroh, 1958; 1962; Lekhnitskii, 1963; Steeds, 1973; Bacon et al., 1978; Ting, 1996), most applications are concerned with the orthotropic system in which the plane-strain and antiplane-strain fields are separable. Another special system for which the strain fields no longer decouple, but closed-form solutions are still obtainable, deals with the ⟨111⟩ line defects in cubic crystals (Head, 1964; Chou and Mitchell, 1967; Douthwaite and Evans, 1973; Yang and Chou, 1975; 1979; 1982). By ⟨111⟩ defect we mean that the defect line is oriented in the ⟨111⟩ direction. This system is important because the ⟨111⟩ screw dislocation plays a key role in plastic deformation of b.c.c. crystals (Foxall et al., 1967; Kroupa and Vitek, 1967).

Recently, Chou (1995) and Chou and Barnett (1997a, b) derived a set of reciprocal relationships between the stresses of a line force and the displacement gradients of a straight dislocation. These correlations provide a simple path for obtaining the solutions for one type of defect from the solutions for the other type without laborious calculations. A typical example has been illustrated for an orthotropic system (Chou and Barnett, 1997a).

In this note, we present a set of relationships between the stress fields of a ⟨111⟩ dislocation and a ⟨111⟩ line force, and we use these relationships to verify the stress expressions of ⟨111⟩ defects reported in the literature (Chou and Mitchell, 1967; Yang and Chou, 1979).

Analysis and Results

A ⟨111⟩ dislocation or line force in a cubic crystal has threefold rotational symmetry (Head, 1964), and can be embedded in a Cartesian coordinate system $0x_1x_2x_3$ with x_1 parallel to $[11\bar{2}]$, x_2 to $[\bar{1}10]$ and x_3 to $[111]$ and with the defect line along the x_3-axis. For this system, it is more convenient to divide the modified elastic compliances S_{MN} into two separate sets, i.e.,

$$\begin{pmatrix} e_{11} \\ e_{22} \\ e_{13} \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} & S_{15} \\ S_{12} & S_{11} & -S_{15} \\ S_{15} & -S_{15} & S_{44} \end{pmatrix}\begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{13} \end{pmatrix} \qquad (1a)$$

and

$$\begin{pmatrix} e_{23} \\ e_{12} \end{pmatrix} = \begin{pmatrix} S_{44} & -2S_{15} \\ -2S_{15} & S_{66} \end{pmatrix}\begin{pmatrix} \sigma_{23} \\ \sigma_{12} \end{pmatrix} \qquad (1b)$$

with

$$S_{11} = S_{22} = (s^0_{11} - \tfrac{1}{2}\delta) - (s^0_{12} + \tfrac{1}{3}\delta)^2/(s^0_{11} - \tfrac{2}{3}\delta)$$

$$S_{12} = (s^0_{12} + \tfrac{1}{6}\delta) - (s^0_{12} + \tfrac{1}{3}\delta)^2/(s^0_{11} - \tfrac{2}{3}\delta)$$

[1] Department of Materials Science and Engineering, National Tsing Hua University, Hsinchu, Taiwan.

[2] On leave from the Department of Materials Science and Engineering, Lehigh University, Bethlehem, PA 18015.

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. 27, 1998; final revision, May 28, 1998. Associate Technical Editor: R. C. Becker.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\frac{\partial(Ku, u)}{\partial\alpha} = [\mp b \quad a]\begin{bmatrix} a' & c' \\ c' & b' \end{bmatrix}\begin{bmatrix} \mp b \\ a \end{bmatrix}$$

$$= a'b^2 + b'a^2 \mp 2c'ab = \pm p \qquad (10)$$

which is nonzero by assumption. Explicit solution of (6) then gives

$$\lambda^2 = -p\alpha/h + O(\alpha^2) \quad \text{and}$$

$$\lambda^2 = -h - \alpha(t - p/h) + O(\alpha^2) \qquad (11)$$

when $h \neq 0$ and

$$\lambda^2 = \pm\sqrt{-\alpha p} + O(\alpha^{3/2}) \qquad (12)$$

when $h = 0$, both of which have $\mathrm{Re}[\lambda] > 0$ for at least one eigensolution in the neighborhood of $\alpha = 0$. The preceding arguments remain valid under the substitution $g \to g + \alpha g'$, so the conjecture remains true when G is a linear function of α.

4 Concluding Comments

The original conjecture (Renshaw and Mote, 1996) related the conditions of the conjecture to a solvability condition for the existence of eigensolutions with a nonzero steady velocity. This interpretation remains valid for the reformulated conjecture and is used explicitly in the comments of Lancaster and Kleim (1997).

A number of systems derive their gyroscopic properties from a repeated application of the material derivative to evaluate the acceleration of the material particles comprising the system. For example, the axially moving string and the axially moving beam compute the acceleration of the string or beam as

$$\frac{D^2u}{Dt^2} = \left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial x}\right)^2 u \qquad (13)$$

where u is the displacement, x is the axial coordinate, t is time, and v is the axial velocity. The transverse acceleration of a rotating disk computed in the stationary frame of reference is

$$\frac{D^2u}{Dt^2} = \left(\frac{\partial}{\partial t} + \Omega\frac{\partial}{\partial\theta}\right)^2 u \qquad (14)$$

where Ω is the rotation speed of the disk. For these systems, $\partial(Ku, u)/\partial v < 0$ and $\partial(Ku, u)/\partial\Omega < 0$ because of the form of the material derivative. Consequently, one expects the conjecture to be generally applicable to systems such as these whose gyroscopic properties are derived from a repeated application of the material derivative.

References

Lancaster, P., 1966, *Lambda-matrices and Vibrating Systems,* Pergamon Press, Oxford, UK.

Lancaster, P., and Kliem, W., 1997, "Comments on Stability Properties of Conservative Gyroscopic Systems," ASME JOURNAL OF APPLIED MECHANICS, in press.

Renshaw, A. A., and Mote, Jr., C. D., 1996, "Local Stability of Gyroscopic Systems Near Vanishing Eigenvalues," ASME JOURNAL OF APPLIED MECHANICS, Vol. 63, pp. 116–120.

Seyranian, A. P., 1997, discussion on "Local Stability of Gyroscopic Systems Near Vanishing Eigenvalues," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 720–721.

Shieh, R. C., 1971, "Energy and Variational Principles for Generalized (Gyroscopic) Conservative Problems," *International Journal of Non-Linear Mechanics,* Vol. 5, pp. 495–509.

Linear Relationships Between the Stress Fields of ⟨111⟩ Line Defects

Y. T. Chou[1,2] and Sanboh Lee[1]

Introduction

Although the theory of two-dimensional anisotropic elasticity has been studied extensively (Eshelby et al., 1953; Stroh, 1958; 1962; Lekhnitskii, 1963; Steeds, 1973; Bacon et al., 1978; Ting, 1996), most applications are concerned with the orthotropic system in which the plane-strain and antiplane-strain fields are separable. Another special system for which the strain fields no longer decouple, but closed-form solutions are still obtainable, deals with the ⟨111⟩ line defects in cubic crystals (Head, 1964; Chou and Mitchell, 1967; Douthwaite and Evans, 1973; Yang and Chou, 1975; 1979; 1982). By ⟨111⟩ defect we mean that the defect line is oriented in the ⟨111⟩ direction. This system is important because the ⟨111⟩ screw dislocation plays a key role in plastic deformation of b.c.c. crystals (Foxall et al., 1967; Kroupa and Vitek, 1967).

Recently, Chou (1995) and Chou and Barnett (1997a, b) derived a set of reciprocal relationships between the stresses of a line force and the displacement gradients of a straight dislocation. These correlations provide a simple path for obtaining the solutions for one type of defect from the solutions for the other type without laborious calculations. A typical example has been illustrated for an orthotropic system (Chou and Barnett, 1997a).

In this note, we present a set of relationships between the stress fields of a ⟨111⟩ dislocation and a ⟨111⟩ line force, and we use these relationships to verify the stress expressions of ⟨111⟩ defects reported in the literature (Chou and Mitchell, 1967; Yang and Chou, 1979).

Analysis and Results

A ⟨111⟩ dislocation or line force in a cubic crystal has threefold rotational symmetry (Head, 1964), and can be embedded in a Cartesian coordinate system $0x_1x_2x_3$ with x_1 parallel to $[11\bar{2}]$, x_2 to $[\bar{1}10]$ and x_3 to $[111]$ and with the defect line along the x_3-axis. For this system, it is more convenient to divide the modified elastic compliances S_{MN} into two separate sets, i.e.,

$$\begin{pmatrix} e_{11} \\ e_{22} \\ e_{13} \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} & S_{15} \\ S_{12} & S_{11} & -S_{15} \\ S_{15} & -S_{15} & S_{44} \end{pmatrix}\begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{13} \end{pmatrix} \qquad (1a)$$

and

$$\begin{pmatrix} e_{23} \\ e_{12} \end{pmatrix} = \begin{pmatrix} S_{44} & -2S_{15} \\ -2S_{15} & S_{66} \end{pmatrix}\begin{pmatrix} \sigma_{23} \\ \sigma_{12} \end{pmatrix} \qquad (1b)$$

with

$$S_{11} = S_{22} = (s^0_{11} - \tfrac{1}{2}\delta) - (s^0_{12} + \tfrac{1}{3}\delta)^2/(s^0_{11} - \tfrac{2}{3}\delta)$$

$$S_{12} = (s^0_{12} + \tfrac{1}{6}\delta) - (s^0_{12} + \tfrac{1}{3}\delta)^2/(s^0_{11} - \tfrac{2}{3}\delta)$$

[1] Department of Materials Science and Engineering, National Tsing Hua University, Hsinchu, Taiwan.

[2] On leave from the Department of Materials Science and Engineering, Lehigh University, Bethlehem, PA 18015.

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. 27, 1998; final revision, May 28, 1998. Associate Technical Editor: R. C. Becker.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\frac{1}{b_1}\{S_{15}\sigma_{11}^{(D1)} - S_{15}\sigma_{22}^{(D1)} + S_{44}\sigma_{13}^{(D1)}\} = \frac{1}{b_2}\{2S_{15}\sigma_{12}^{(D2)} - S_{44}\sigma_{23}^{(D2)}\}. \quad (8)$$

Note that there is no sum over the index "i" on the right-hand sides of Eqs. (7a), (7b), and (7c). Furthermore, Eqs. (7) and (8) can also be derived based on equations (1) and (5) and the fact that $\sigma_{ij}^{(F3)}$ can be expressed in two different ways using Eqs. (5a) and (5b).

A useful application of the above relationships is to verify the existing solutions of the ⟨111⟩ defects in cubic crystals. Using Eqs. (6) and the results of $\sigma_{ij}^{(F)}$ reported by Yang and Chou (1975), the calculated $\sigma_{ij}^{(D)}$ agree with those of Chou and Mitchell (1967). Equation (8) is also readily verified with the known $\sigma_{ij}^{(D)}$'s.

Acknowledgments

The authors thank Prof. D. M. Barnett of Stanford University for helpful discussions. The work was supported by the National Science Council, Taiwan, R.O.C.

References

Bacon, D. J., Barnett, D. M., and Scattergood, R. O., 1978, "The Anisotropic Continuum Theory of Lattice Defects," *Progress in Materials Science,* Vol. 23, pp. 51–262.

Chou, Y. T., 1995, "Moutier Theorem and Dislocation Field," *Micromechanics of Advanced Materials,* S. N. G. Chu et al., eds., Minerals, Metals and Materials Society Press, Warrendale, PA, pp. 25–28.

Chou, Y. T., and Barnett, D. M., 1997a, "Reciprocal Relationships between the Stress fields of a Line Force and a Dislocation in Anisotropic Media," *Philosophical Magazine A,* Vol. 76, pp. 221–229.

Chou, Y. T., and Barnett, D. M., 1997b, "On a Reciprocal Relationship between the Deformation Fields of a line Force and a Straight Dislocation," *Philosophical Magazine Letters,* Vol. 76, pp. 139–141.

Chou, Y. T., and Mitchell, T. E., 1967, "Stress and Dilatation Fields of the ⟨111⟩ Dislocation in Cubic Crystals," *Journal of Applied Physics,* Vol. 38, pp. 1535–1540.

Douthwaite, R. W., and Evans, J. T., 1973, "Interaction between a Tetragonal Distortion and a ⟨111⟩ Screw Dislocation in an Anisotropic Cubic Crystal," *Scripta Metallurgica,* Vol. 7, pp. 1019–1026.

Eshelby, J. D., Read, W. T., and Shockley, W., 1953, "Anisotropic Elasticity with Applications to Dislocation Theory," *Acta Metallurgica,* Vol. 1, pp. 251–259.

Foxall, R. A., Duesbery, M. S., and Hirsch; P. B., 1967, "The Deformation of Niobium Single Crystals," *Canadian Journal of Physics,* Vol. 45, pp. 607–629.

Head, A. K., 1964, "The [111] Dislocation in a Cubic Crystal," *Physica Status Solidi,* Vol. 6, pp. 461–465.

Kroupa, F., and Vitek, V., 1967, "Slip and the Conception of Splitting of Dislocations in b.c.c. Metals," *Canadian Journal of Physics,* Vol. 45, pp. 945–971.

Lekhnitskii, S. G., 1963, *Theory of Elasticity of an Anisotropic Body,* Holden-Day, San Francisco, CA.

Steeds, J. W., 1973, *Anisotropic Elasticity Theory of Dislocations,* Oxford University Press, New York.

Stroh, A. N., 1958, "Dislocations and Cracks in Anisotropic Elasticity," *Philosophical Magazine,* Series 8, Vol. 3, pp. 625–646.

Stroh, A. N., 1962, "Steady State Problems in Anisotropic Elasticity," *Journal of Mathematics and Physics,* Vol. 41, pp. 77–103.

Ting, T. C. T., 1996, *Anisotropic Elasticity: Theory and Applications,* Oxford University Press, New York.

Yang, H. C., and Chou, Y. T., 1975, "Strain Energy of ⟨111⟩ Slit Cracks in Cubic Crystals," *Materials Science and Engineering,* Vol. 17, pp. 57–62.

Yang, H. C., and Chou, Y. T., 1979, "The ⟨111⟩ Line Force in a Cubic Crystal," *Journal of Applied Physics,* Vol. 50, pp. 773–776; 6570.

Yang, H. C., and Chou, Y. T., 1982, "The ⟨111⟩ Elliptic Inclusion in a Cubic Crystal," ASME JOURNAL OF APPLIED MECHANICS, Vol. 49, pp. 353–360.

Mohr-Coulomb Failure Criterion Expressed in Terms of Stress Invariants

G. S. Schajer[1]

A polynomial expression for the Mohr-Coulomb criterion for failure of brittle or frictional materials in terms of stress invariants is derived. The form of the polynomial expression corresponds to the pyramidal shape of the failure surface in principal stress space. The pressure-sensitive character of the failure of a Mohr-Coulomb material is indicated by a linear influence of the first stress invariant I_1 in the polynomial expression.

Introduction

An isotropic material has physical properties that are independent of spatial orientation. Failure criteria for such materials must therefore have the same property. In this context, the word "failure" can have a general meaning encompassing both yielding and fracture. An effective mathematical way of achieving spatial independence is to express failure criteria in terms of the Cauchy and deviator stress invariants (Gurtin, 1991)

$$I_1 = \sigma_1 + \sigma_2 + \sigma_3 \quad (1)$$

$$J_2 = ((\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2)/6 \quad (2)$$

$$J_3 = (2\sigma_1 - \sigma_2 - \sigma_3)(2\sigma_2 - \sigma_3 - \sigma_1) \times (2\sigma_3 - \sigma_1 - \sigma_2)/27 \quad (3)$$

where σ_1, σ_2, and σ_3 are the principal stresses. These invariants are coefficients of the characteristic equations of the Cauchy and deviator stress tensors. This mixed group of invariants, which is mathematically convenient, is an acceptable choice because they uniquely define the three Cauchy stress invariants I_1, I_2, and I_3.

The Mises and Tresca criteria for yielding of ductile metals have well-known expressions in terms of the stress invariants. Both these criteria refer to "pressure-insensitive" materials, i.e., materials whose yielding failure is independent of hydrostatic stress. This implies an independence of the first invariant I_1. However, the concept of describing isotropic material properties in terms of the stress invariants is not limited to such materials. It is the purpose of this Note to describe an example of using the above stress invariants to describe the failure of a pressure-sensitive material. An invariant expression for the Mohr-Coulomb failure criterion for brittle (or frictional) materials is derived explicitly in terms of I_1, J_2, and J_3.

Tresca Criterion

The Tresca (maximum shear stress) criterion can be considered as a special case of the Mohr-Coulomb criterion. The simpler Tresca criterion is considered first here to give some insight into working with the Mohr-Coulomb criterion. In terms of the principal stresses, the Tresca criterion for yielding of ductile metals is usually written as

[1] Professor, Department of Mechanical Engineering, University of Britist Columbia, Vancouver, BC, Canada. Mem. ASME.

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, Dec. 12, 1997; final revision, July 9, 1998. Associate Technical Editor: M. M. Carroll.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\frac{1}{b_1}\{S_{15}\sigma_{11}^{(D1)} - S_{15}\sigma_{22}^{(D1)} + S_{44}\sigma_{13}^{(D1)}\} = \frac{1}{b_2}\{2S_{15}\sigma_{12}^{(D2)} - S_{44}\sigma_{23}^{(D2)}\}. \quad (8)$$

Note that there is no sum over the index "i" on the right-hand sides of Eqs. (7a), (7b), and (7c). Furthermore, Eqs. (7) and (8) can also be derived based on equations (1) and (5) and the fact that $\sigma_{ij}^{(F3)}$ can be expressed in two different ways using Eqs. (5a) and (5b).

A useful application of the above relationships is to verify the existing solutions of the ⟨111⟩ defects in cubic crystals. Using Eqs. (6) and the results of $\sigma_{ij}^{(F)}$ reported by Yang and Chou (1975), the calculated $\sigma_{ij}^{(D)}$ agree with those of Chou and Mitchell (1967). Equation (8) is also readily verified with the known $\sigma_{ij}^{(D)}$'s.

Acknowledgments

The authors thank Prof. D. M. Barnett of Stanford University for helpful discussions. The work was supported by the National Science Council, Taiwan, R.O.C.

References

Bacon, D. J., Barnett, D. M., and Scattergood, R. O., 1978, "The Anisotropic Continuum Theory of Lattice Defects," *Progress in Materials Science,* Vol. 23, pp. 51–262.

Chou, Y. T., 1995, "Moutier Theorem and Dislocation Field," *Micromechanics of Advanced Materials,* S. N. G. Chu et al., eds., Minerals, Metals and Materials Society Press, Warrendale, PA, pp. 25–28.

Chou, Y. T., and Barnett, D. M., 1997a, "Reciprocal Relationships between the Stress fields of a Line Force and a Dislocation in Anisotropic Media," *Philosophical Magazine A,* Vol. 76, pp. 221–229.

Chou, Y. T., and Barnett, D. M., 1997b, "On a Reciprocal Relationship between the Deformation Fields of a line Force and a Straight Dislocation," *Philosophical Magazine Letters,* Vol. 76, pp. 139–141.

Chou, Y. T., and Mitchell, T. E., 1967, "Stress and Dilatation Fields of the ⟨111⟩ Dislocation in Cubic Crystals," *Journal of Applied Physics,* Vol. 38, pp. 1535–1540.

Douthwaite, R. W., and Evans, J. T., 1973, "Interaction between a Tetragonal Distortion and a ⟨111⟩ Screw Dislocation in an Anisotropic Cubic Crystal," *Scripta Metallurgica,* Vol. 7, pp. 1019–1026.

Eshelby, J. D., Read, W. T., and Shockley, W., 1953, "Anisotropic Elasticity with Applications to Dislocation Theory," *Acta Metallurgica,* Vol. 1, pp. 251–259.

Foxall, R. A., Duesbery, M. S., and Hirsch; P. B., 1967, "The Deformation of Niobium Single Crystals," *Canadian Journal of Physics,* Vol. 45, pp. 607–629.

Head, A. K., 1964, "The [111] Dislocation in a Cubic Crystal," *Physica Status Solidi,* Vol. 6, pp. 461–465.

Kroupa, F., and Vitek, V., 1967, "Slip and the Conception of Splitting of Dislocations in b.c.c. Metals," *Canadian Journal of Physics,* Vol. 45, pp. 945–971.

Lekhnitskii, S. G., 1963, *Theory of Elasticity of an Anisotropic Body,* Holden-Day, San Francisco, CA.

Steeds, J. W., 1973, *Anisotropic Elasticity Theory of Dislocations,* Oxford University Press, New York.

Stroh, A. N., 1958, "Dislocations and Cracks in Anisotropic Elasticity," *Philosophical Magazine,* Series 8, Vol. 3, pp. 625–646.

Stroh, A. N., 1962, "Steady State Problems in Anisotropic Elasticity," *Journal of Mathematics and Physics,* Vol. 41, pp. 77–103.

Ting, T. C. T., 1996, *Anisotropic Elasticity: Theory and Applications,* Oxford University Press, New York.

Yang, H. C., and Chou, Y. T., 1975, "Strain Energy of ⟨111⟩ Slit Cracks in Cubic Crystals," *Materials Science and Engineering,* Vol. 17, pp. 57–62.

Yang, H. C., and Chou, Y. T., 1979, "The ⟨111⟩ Line Force in a Cubic Crystal," *Journal of Applied Physics,* Vol. 50, pp. 773–776; 6570.

Yang, H. C., and Chou, Y. T., 1982, "The ⟨111⟩ Elliptic Inclusion in a Cubic Crystal," ASME JOURNAL OF APPLIED MECHANICS, Vol. 49, pp. 353–360.

Mohr-Coulomb Failure Criterion Expressed in Terms of Stress Invariants

G. S. Schajer[1]

A polynomial expression for the Mohr-Coulomb criterion for failure of brittle or frictional materials in terms of stress invariants is derived. The form of the polynomial expression corresponds to the pyramidal shape of the failure surface in principal stress space. The pressure-sensitive character of the failure of a Mohr-Coulomb material is indicated by a linear influence of the first stress invariant I_1 in the polynomial expression.

Introduction

An isotropic material has physical properties that are independent of spatial orientation. Failure criteria for such materials must therefore have the same property. In this context, the word "failure" can have a general meaning encompassing both yielding and fracture. An effective mathematical way of achieving spatial independence is to express failure criteria in terms of the Cauchy and deviator stress invariants (Gurtin, 1991)

$$I_1 = \sigma_1 + \sigma_2 + \sigma_3 \quad (1)$$

$$J_2 = ((\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2)/6 \quad (2)$$

$$J_3 = (2\sigma_1 - \sigma_2 - \sigma_3)(2\sigma_2 - \sigma_3 - \sigma_1) \times (2\sigma_3 - \sigma_1 - \sigma_2)/27 \quad (3)$$

where σ_1, σ_2, and σ_3 are the principal stresses. These invariants are coefficients of the characteristic equations of the Cauchy and deviator stress tensors. This mixed group of invariants, which is mathematically convenient, is an acceptable choice because they uniquely define the three Cauchy stress invariants I_1, I_2, and I_3.

The Mises and Tresca criteria for yielding of ductile metals have well-known expressions in terms of the stress invariants. Both these criteria refer to "pressure-insensitive" materials, i.e., materials whose yielding failure is independent of hydrostatic stress. This implies an independence of the first invariant I_1. However, the concept of describing isotropic material properties in terms of the stress invariants is not limited to such materials. It is the purpose of this Note to describe an example of using the above stress invariants to describe the failure of a pressure-sensitive material. An invariant expression for the Mohr-Coulomb failure criterion for brittle (or frictional) materials is derived explicitly in terms of I_1, J_2, and J_3.

Tresca Criterion

The Tresca (maximum shear stress) criterion can be considered as a special case of the Mohr-Coulomb criterion. The simpler Tresca criterion is considered first here to give some insight into working with the Mohr-Coulomb criterion. In terms of the principal stresses, the Tresca criterion for yielding of ductile metals is usually written as

[1] Professor, Department of Mechanical Engineering, University of Britist Columbia, Vancouver, BC, Canada. Mem. ASME.

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, Dec. 12, 1997; final revision, July 9, 1998. Associate Technical Editor: M. M. Carroll.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

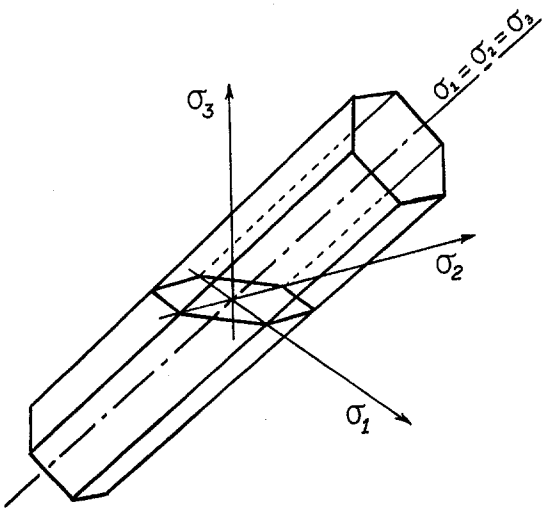


Fig. 1 Yield surface for the Tresca criterion

$$\max\left[|\sigma_1 - \sigma_2|, |\sigma_2 - \sigma_3|, |\sigma_3 - \sigma_1|\right] = \sigma_Y. \quad (4)$$

The three terms in Eq. (4) can be interpreted as the circle diameters in a three-dimensional Mohr's circle representation of the applied stress state. The "max" operator chooses the largest circle. An equivalent procedure is to choose the circle containing the most and least tensile (or least and most compressive) principal stresses. Equation (4) then becomes

$$\frac{\max\left[\sigma_1, \sigma_2, \sigma_3\right]}{\sigma_Y} - \frac{\min\left[\sigma_1, \sigma_2, \sigma_3\right]}{\sigma_Y} - 1 = 0. \quad (5)$$

Equations (4) and (5) each contain the same six possible yield conditions. They are $\sigma_1 - \sigma_2 - \sigma_Y = 0$, $\sigma_2 - \sigma_1 - \sigma_Y = 0$, etc. Figure 1 shows the Tresca yield surface in σ_1, σ_2, σ_3 space. The six faces of the regular hexagonal tube correspond to the six possible yield conditions.

One interpretation of the stress invariants defined in Eqs. (1)–(3) is as axes of a cylindrical coordinate system in the σ_1, σ_2, σ_3 space shown in Fig. 1. The first stress invariant describes the axial distance $a = I_1/\sqrt{3}$ along the line $\sigma_1 = \sigma_2 = \sigma_3$. This is a measure of the hydrostatic stress. The second deviator stress invariant describes the radial distance $r = \sqrt{2J_2}$ perpendicular to the $\sigma_1 = \sigma_2 = \sigma_3$ line. The third deviator stress invariant (together with J_2) describes the Lode or "similarity" angle $\theta = \frac{1}{3}\cos^{-1}\left(\frac{1}{2}J_3(3/J_2)^{3/2}\right)$ around the line $\sigma_1 = \sigma_2 = \sigma_3$ relative to the projection of any of the principal stress axes (Novozhilov, 1952; Chen and Han, 1988).

The Tresca criterion can be expressed in terms of the stress invariants by multiplying together all six yield conditions from Eqs. (4) and (5). After some algebraic manipulation, the resulting expression in terms of the stress invariants is (adapted from Prager and Hodge (1963))

$$4J_2^3 - 27J_3^2 - 9\sigma_y^2 J_2^2 + 6\sigma_Y^4 J_2 - \sigma_Y^6 = 0 \quad (6)$$

Equation (6) shows dependence on J_2 and J_3, but not on I_1. This confirms that the Tresca yield criterion is independent of hydrostatic stress.

Mohr-Coulomb Criterion

The Mohr-Coulomb criterion describes the failure of many brittle (or frictional) materials. The behavior of such materials can be specified by the failure stresses in uniaxial tension and compression, σ_T and σ_C, respectively. Both these quantities are taken as positive numbers. The Mohr-Coulomb criterion can be considered as a generalization of the Tresca criterion. Thus, equations for the Mohr-Coulomb criterion should have similar forms to the Tresca equations, both in terms of principal stresses and stress invariants.

Equation (5) can be generalized to provide a compact principal stress equation for a Mohr-Coulomb material. The result is (Chen and Han, 1988; Schajer, 1994)

$$\frac{\max\left[\sigma_1, \sigma_2, \sigma_3\right]}{\sigma_T} - \frac{\min\left[\sigma_1, \sigma_2, \sigma_3\right]}{\sigma_C} - 1 = 0. \quad (7)$$

Like Eq. (5), Eq. (7) also contains six possible failure conditions. These consist of the six possible permutations of the most and least tensile principal stresses, e.g., $\sigma_1/\sigma_T - \sigma_2/\sigma_C - 1 = 0$, $\sigma_2/\sigma_T - \sigma_1/\sigma_C - 1 = 0$, etc. Figure 2 illustrates the failure surface for the Mohr-Coulomb criterion. It is a six-sided pyramid with an apex at the point $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_C\sigma_T/(\sigma_C - \sigma_T)$. The six faces of the pyramid correspond to the six possible failure conditions. Following the Tresca procedure, these six possible failure conditions are multiplied together. The result is a sixth-order polynomial in terms of the principal stresses σ_1, σ_2, and σ_3. In turn, this polynomial corresponds to another sixth-order polynomial whose 22 terms contain all possible combinations of the three stress invariants I_1, J_2, and J_3. Using an inverse approach, the coefficient of each term in the invariant polynomial can be identified one by one from the principal stress polynomial. After some further algebraic manipulation, a relatively compact result can be achieved:

$$p^2J_2^3 - v^3J_3^2 - (s - qI_1)qvwJ_2J_3 - (s - qI_1)^2v^2J_2^2 + (s - qI_1)^3qwJ_3 + 2(s - qI_1)^4vJ_2 - (s - qI_1)^6 = 0 \quad (8)$$

where

$$s = \sqrt{\sigma_C\sigma_T} \quad p = (\sigma_C + \sigma_T)/s \quad q = (\sigma_C - \sigma_T)/3s$$

$$v = (\sigma_C^2 + \sigma_C\sigma_T + \sigma_T^2)/s^2 \quad w = 3(2\sigma_C + \sigma_T)(\sigma_C + 2\sigma_T)/s^2.$$

As expected, Eq. (8) has a similar algebraic form to Eq. (6). It reduces to Eq. (6) when σ_T and σ_C are both set equal to σ_Y. A Mohr-Coulomb material is "pressure-sensitive," i.e., its failure depends on the first stress invariant I_1. The pyramidal shape of the Mohr-Coulomb failure surface shows that the influence of I_1 is linear. Correspondingly, each term containing σ_Y in Eq. (6) is replaced by a term containing $(s - qI_1)$ in Eq. (8). The two extra terms containing odd powers of $(s - qI_1)$ occur because the cross section of the pyramid is not a regular hexagon. It only has threefold symmetry, not six. The fifth-order $(s - qI_1)$ term is absent in Eq. (8). The condition $(s - qI_1) = 0$ corresponds to the stress state at the apex of the pyramid, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_C\sigma_T/(\sigma_C - \sigma_T)$.

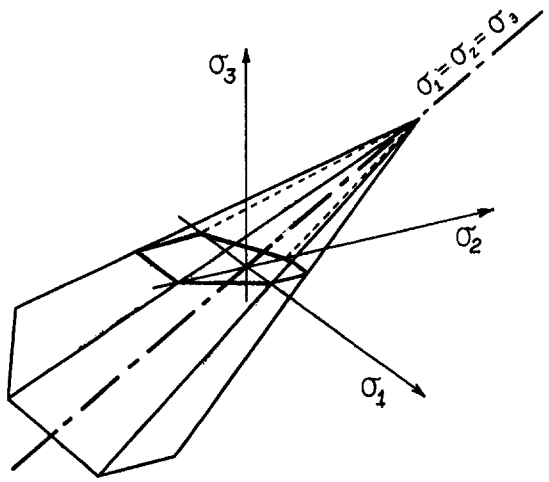


Fig. 2 Failure surface for the Mohr-Coulomb criterion

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Previous authors have formulated compact expressions for the Mohr-Coulomb criterion in terms of derived invariant quantities. For example, Chen (Chen and Han, 1988) presented a expression in terms of the quantities I_1, J_2, and θ. However, to the author's knowledge, Eq. (8) is the first presentation of an expression for the Mohr-Coulomb criterion that involves the stress invariants I_1, J_2, and J_3 explicitly.

Conclusions

A polynomial expression for the Mohr-Coulomb criterion for failure of brittle (or frictional) materials has been formulated in terms of the stress invariants I_1, J_2, and J_3. It is a generalization of the corresponding expression for yielding of a ductile material according to the Tresca criterion. The form of the polynomial expression corresponds to the pyramidal shape of the failure surface in principal stress space. The pressure-sensitive character of the failure of a Mohr-Coulomb material is indicated by a linear influence of the first stress invariant I_1 in the polynomial expression.

Acknowledgment

This work was financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Dr. R. Vaziri kindly reviewed the manuscript.

References

Chen, W. F., and Han, D. J., 1988, *Plasticity for Structural Engineers,* Springer-Verlag, New York.

Gurtin, M. E., 1981, *Introduction to Continuum Mechanics,* Academic, New York.

Novozhilov, V. V., 1952, "On the Physical Interpretation of the Stress Invariants used in the Theory of Plasticity," *Prikladnaya Mathematika i Mekhanika,* (Applied Mathematics and Mechanics), Vol. 16, No. 5, pp. 617–619 (in Russian).

Prager, W., and Hodge, P. G., 1963, *Theory of Perfectly Plastic Solids,* John Wiley and Sons, New York.

Schajer, G. S., 1994, "A Teaching Note on failure criteria and failure surfaces for ductile and brittle materials," *International Journal of Mechanical Engineering Education,* Vol. 22, No. 1, pp. 1–13.

ERRATUM

"A Volume Integral Equation Technique for Multiple Inclusion and Crack Interaction Problems," by Jungki Lee and Ajit Mal, published in the ASME JOURNAL OF APPLIED MECHANICS, Mar. 1997, Vol. 64, pp. 23–31.

The material properties used in the numerical results given in Table 2 (p. 25) and Fig. 4 (p. 26) are different from those listed in Table 1 (p. 25). The correct values are given in the table below:

Table 1 Material that was used for the calculation

	E(GPa)	ν
SiC	440.15	0.17
Ti	120.83	0.32
Carbon	41.0	0.25

The authors apologize for any confusion caused by this error.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Previous authors have formulated compact expressions for the Mohr-Coulomb criterion in terms of derived invariant quantities. For example, Chen (Chen and Han, 1988) presented a expression in terms of the quantities I_1, J_2, and θ. However, to the author's knowledge, Eq. (8) is the first presentation of an expression for the Mohr-Coulomb criterion that involves the stress invariants I_1, J_2, and J_3 explicitly.

Conclusions

A polynomial expression for the Mohr-Coulomb criterion for failure of brittle (or frictional) materials has been formulated in terms of the stress invariants I_1, J_2, and J_3. It is a generalization of the corresponding expression for yielding of a ductile material according to the Tresca criterion. The form of the polynomial expression corresponds to the pyramidal shape of the failure surface in principal stress space. The pressure-sensitive character of the failure of a Mohr-Coulomb material is indicated by a linear influence of the first stress invariant I_1 in the polynomial expression.

Acknowledgment

This work was financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Dr. R. Vaziri kindly reviewed the manuscript.

References

Chen, W. F., and Han, D. J., 1988, *Plasticity for Structural Engineers,* Springer-Verlag, New York.

Gurtin, M. E., 1981, *Introduction to Continuum Mechanics,* Academic, New York.

Novozhilov, V. V., 1952, "On the Physical Interpretation of the Stress Invariants used in the Theory of Plasticity," *Prikladnaya Mathematika i Mekhanika,* (Applied Mathematics and Mechanics), Vol. 16, No. 5, pp. 617–619 (in Russian).

Prager, W., and Hodge, P. G., 1963, *Theory of Perfectly Plastic Solids,* John Wiley and Sons, New York.

Schajer, G. S., 1994, "A Teaching Note on failure criteria and failure surfaces for ductile and brittle materials," *International Journal of Mechanical Engineering Education,* Vol. 22, No. 1, pp. 1–13.

ERRATUM

"A Volume Integral Equation Technique for Multiple Inclusion and Crack Interaction Problems," by Jungki Lee and Ajit Mal, published in the ASME JOURNAL OF APPLIED MECHANICS, Mar. 1997, Vol. 64, pp. 23–31.

The material properties used in the numerical results given in Table 2 (p. 25) and Fig. 4 (p. 26) are different from those listed in Table 1 (p. 25). The correct values are given in the table below:

Table 1 Material that was used for the calculation

	E(GPa)	ν
SiC	440.15	0.17
Ti	120.83	0.32
Carbon	41.0	0.25

The authors apologize for any confusion caused by this error.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Journal of Applied Mechanics

Discussion

Time-Integral Variational Principles for Nonlinear Nonholonomic Systems[1]

B. Ravindra[2].

1 Introduction

The author presents a concise account of the time-integral variational principles for nonholonomic systems and also gives a geometrical interpretation of the discrepancies found between the equations of motion obtained using the variational approach and those derived from the d'Alembert-Lagrange principle. He rightly points out the lack of appreciation of these nuances in the engineering community, as has also been observed recently by Hagedorn (1997). The author follows Hertz (1894) in distinguishing the straightest (mechanical) and the shortest (variational equations for the geodesics or "mathematical" as termed by the author) and arrives at the well-known conclusion that the lack of commutativity of the variational and integral operators is the source of the discrepancy. While the author rightly emphasizes the significance of the geometric view point, some of the exciting, recent developments on the new geometry arising out of the problems of "nonholonomic" mechanics and related connections to control theory are not mentioned. Since these aspects are of considerable theoretical and practical value to the practitioners of mechanics, a brief description with relevant references is provided here.

Riemann's Habilitationsvortrag in 1854 was a watershed in the history of geometry. Since then, deep relations between geometry and mechanics have been discovered by Levi-Civita, Ricci, Weyl, among others. The problems of nonholonomic mechanics also served as a source of inspiration for the creation of a new language and new geometric objects. A simple example of nonholonomic structure is the contact structure. It is interesting to note that this arose from the work of Gibbs on his graphical methods in thermodynamics of fluids (1873). However, it is the pioneering work of Caratheodory in 1909 on foundations of thermodynamics that provided the necessary impetus for further work. In his analysis of second law of thermodynamics and entropy, Caratheodory needed a theorem that any two points on a contact manifold may be connected by an admissible curve. An extension of Caratheodory's work is due to Chow (1939) and Rashevsky (1938). As pointed by Hermann (1963), this theorem, often referred to as Chow theorem, provides the differential geometric picture of the accessibility problem in control theory. It is a remarkable fact that, though Caratheodory's theorem has a kinematic nature, it can be used to define a nonholonomic metric known as Carnot-Caratheodory metric (Gromov, 1995). The geodesic problems arising out of such a metric have many interesting features (Montgomery, 1994). The geodesics, which do not satisfy the usual geodesic equations are termed as the "abnormal minimizers" and there exists a natural interpretation of these via Pontrjagin's maximum principle. It may be noted that one already encounters singular extremals in Caratheodory's calculus of variations. The history of nonholonomic mechanics and associated controversies reminds one the need for a suitable language and notation. It appears that E. Cartan has already advocated this in 1935. As Laplace said, "Such is the advantage of a well-constructed language that its simplified notation often becomes the source of profound theories." The controversy between Leibniz and Newton over the invention of calculus is well known. Historians point out that despite the claims of priority, the notation adopted by Leibniz was superior and this fact played a major role in the advancement of mathematics in the Continent as opposed to England (Babbage, 1830). Nonholonomic geometry might provide a suitable notation and language to resolve the controversies in nonholonomic mechanics.

It is indeed remarkable that Hertz already cited the theorem due to Caratheodory without proof (Hertz, 1894). Hertz's *Principles of Mechanics* was written with the idea of eliminating force as a fundamental concept in Mechanics and with a program of giving an interpretation of Ether in terms of "hidden motions" and "hidden masses." While this book was superseded by the development of relativity and quantum mechanics, Hertz's own introduction to his mechanics is philosophically profound and influenced Boltzmann and Wittgenstein. In this book, Hertz clearly mentioned the possibility of using higher dimensional geometry in mechanics and introduced a clear distinction between holonomic and nonholonomic constraints. He was also aware of the fact that application of Hamilton's principle to nonintegrable equations is problematic, in particular he mentioned that when three-dimensional bodies roll on one another without slipping, nonintegrable equations necessarily occur. He wrote "For rolling without slipping does not contradict either the principle of energy or any other generally accepted law of physics. The process is one which is so nearly realised in the visible world that even integration machines are constructed on the assumption that it strictly takes place." While this reference to integrating machines is quite curious, one finds an echo in Roger Brockett's foreword to a book on dynamics and control of multibody systems (Marsden et al., 1989):

"In my case, I had been fascinated for some time with the idea of nonholonomic constraints, partly because of the role they play in the planar integration scheme found in Vannevar Bush's differential analyzer and partly because they seemed mysterious. As I attempted to find a comfortable way of thinking about them, and the related matter of the Caratheodory statement of the second law of thermodynamics, I eventually realized that there was a wide variety of ways in which geometrical ideas could be helpful in forming a unified view of control."

[1] By J. G. Papastravridis and published in the Dec. 1997 issue of the ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 985–991.

[2] Institute fur Mechanik II, Darmstadt University of Technology, Darmstadt, D-64289, Germany.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

It is hoped that the control theoretic perspective sheds light on the pathological aspects surrounding the variational problems of the nonholonomic systems.

Acknowledgments

The author thanks the Alexander von Humboldt foundation for their hospitality and Prof. P. Hagedorn and Prof. C. P. Kwong for useful discussions.

References

Babbage, C., 1830, *Reflections on the Decline of Science in England and Some of its Causes,* London.

Gromov, M., 1995, "Carnot-Caratheodory Spaces Seen From Within," IHES preprint.

Hagedorn, P., 1997, "Discussion on Investigation of a New Formulation of the Lagrange Method for Constrained Dynamic Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 1024–1025.

Hermann, Robert, 1963, *Nonlinear Differential Equations and Nonlinear Mechanics,* J. P. LaSalle and S. Lefschetz, eds., Academic Press, New York, pp. 325–332.

Hertz, H., 1894, *Die Prinzipien der Mechanik,* J. A. Barth, Leipzig.

Marsden, J. E., et al., eds., 1989, *Dynamics and Control of Multibody Systems,* American Mathematical Society, Providence, RI.

Montgomery, R., 1994, "Abnormal Minimizers," *SIAM Journal of Control and Optimization,* Vol. 32, pp. 1605–1620.

Author's Closure[3]

I would like to thank B. Ravindra for his scholarly and constructive comments. The reasons that I did not include the modern geometrical developments he is referring to are that, first, for all that one would need a monograph or an extensive review article (à la ASME *Applied Mechanics Reviews*), not a modest size paper like the one discussed; and, second, I am not that well familiar with those developments so as to be able to present them in an accessible fashion. My impression is that, outside of some mathematicians and control theorists, most other engineers are quite uncomfortable with the mathematical jargon currently used to express them; i.e., the kind utilized by such authors as (alphabetically): Arnold, Cartan, Gallissot, Griffiths, Hermann, Lewis, Marsden, Ratiu, Simo, Smale, et al. All these formalisms, in addition to constituting an intimidating "moat" that prevents the rest of us from getting enlightened and participating on those modern developments, they also seem to constitute an expensive and low-yield investment of effort and time. My sincere hope is that someone more knowledgeable and competent than I would present the subject in simple and readable form (i.e., without epsilonics and/or set theoretic straightjackets)—*that* would be a real contribution! What us nonmathematicians need today is (inverting the title of F. Klein's famous book) *advanced mathematics, and mechanics, from an elementary viewpoint,* even at the expense of rigor; otherwise we will keep drifting towards a dynamical tower of Babel. Perhaps B. Ravindra would choose to carry out such a useful task!

Enhanced Elastic Buckling Loads of Composite Plates With Tailored Thermal Residual Stresses[1]

Ch. Sarath Babu and T. Kant[2]. The authors present an innovative idea of tailoring the manufacturing-introduced thermal residual stresses in order to enhance the buckling strength of laminated composite plates and have given some interesting results. To demonstrate the effect of the tailored thermal residual stresses on the buckling load of symmetric composites, the authors have considered a square graphite/epoxy plates with two types (types I and II) of reinforcement that introduce nonzero thermal residual stress resultants when the plates cool to room temperature after the cure process. The basic laminate considered is a $[90/0]_s$ cross-ply having a/h ratio equal to 600 (a = 360 mm; thickness of each ply, t = 0.15 mm) where a and h denote width and thickness of the laminate, respectively.

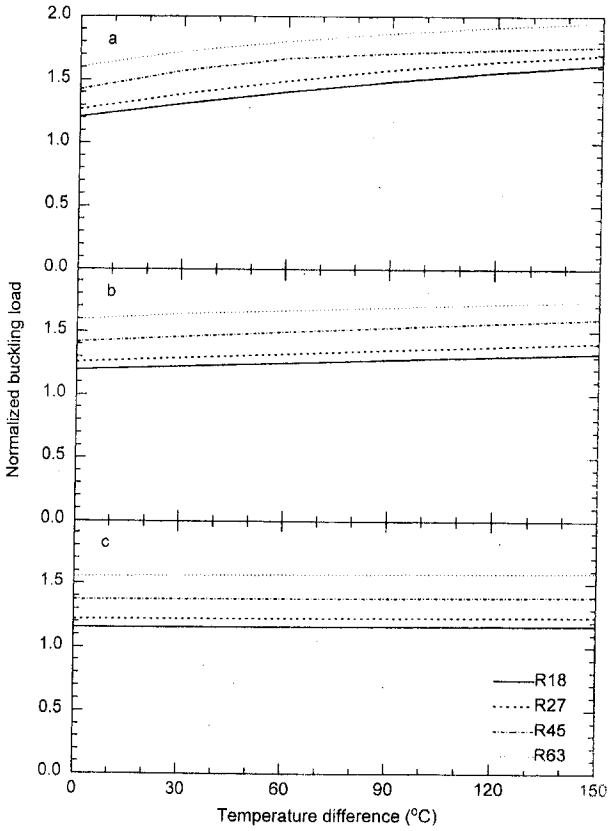


Fig. 1 Normalized buckling loads of plates (type I) R18, R27, R45, and R63 as a function of temperature difference; (*a*) a/h = 600, (*b*) a/h = 300, and (*c*) a/h = 100

Buckling loads of these plates are determined using finite element formulation within the context of linear stability analysis and the eigenvalue problem for λ is obtained as

$$\{[K] + [K_G^R] - \lambda[K_G^Q]\}\{\delta\} = 0 \qquad (1)$$

where $[K]$, $[K_G^R]$, and $[K_G^Q]$ are, respectively, the global stiffness matrix, global geometric stiffness matrix due to thermal residual stress resultants, and global geometric stiffness matrix due to stress resultants caused by displacement loading.

The buckling loads of reinforced plates are normalized with respect to a uniform $[90/0]_s$ plate having the same mass as that of the reinforced plate. The authors have shown that the normalized buckling loads of plates with type I reinforcement increase by about 30 percent and that of plates with type II reinforcement increase by about 300 percent, as the temperature difference is increased from 0°C to 150°C. Though the results presented are interesting, the effect of the thickness of the laminate on enhancement of the buckling load is not presented. Since the authors have considered a very thin laminate (a/h = 600), the increase in normalized buckling load due to increase in temperature difference is very high. However, in case of laminates having higher thickness than the one considered, the

[3] J. G. Papastavridis, School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405.

[1] By S. F. Muller de Almeida and J. S. Hansen and published in the Dec. 1997 issue of the ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 772–780.

[2] Department of Civil Engineering, ITT Bombay, Powai, Mumbai-400 076, India. e-mail: tkant@civil.iitb.ernet.in

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

It is hoped that the control theoretic perspective sheds light on the pathological aspects surrounding the variational problems of the nonholonomic systems.

Acknowledgments

The author thanks the Alexander von Humboldt foundation for their hospitality and Prof. P. Hagedorn and Prof. C. P. Kwong for useful discussions.

References

Babbage, C., 1830, *Reflections on the Decline of Science in England and Some of its Causes,* London.

Gromov, M., 1995, "Carnot-Caratheodory Spaces Seen From Within," IHES preprint.

Hagedorn, P., 1997, "Discussion on Investigation of a New Formulation of the Lagrange Method for Constrained Dynamic Systems," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 1024–1025.

Hermann, Robert, 1963, *Nonlinear Differential Equations and Nonlinear Mechanics,* J. P. LaSalle and S. Lefschetz, eds., Academic Press, New York, pp. 325–332.

Hertz, H., 1894, *Die Prinzipien der Mechanik,* J. A. Barth, Leipzig.

Marsden, J. E., et al., eds., 1989, *Dynamics and Control of Multibody Systems,* American Mathematical Society, Providence, RI.

Montgomery, R., 1994, "Abnormal Minimizers," *SIAM Journal of Control and Optimization,* Vol. 32, pp. 1605–1620.

Author's Closure[3]

I would like to thank B. Ravindra for his scholarly and constructive comments. The reasons that I did not include the modern geometrical developments he is referring to are that, first, for all that one would need a monograph or an extensive review article (à la ASME *Applied Mechanics Reviews*), not a modest size paper like the one discussed; and, second, I am not that well familiar with those developments so as to be able to present them in an accessible fashion. My impression is that, outside of some mathematicians and control theorists, most other engineers are quite uncomfortable with the mathematical jargon currently used to express them; i.e., the kind utilized by such authors as (alphabetically): Arnold, Cartan, Gallissot, Griffiths, Hermann, Lewis, Marsden, Ratiu, Simo, Smale, et al. All these formalisms, in addition to constituting an intimidating "moat" that prevents the rest of us from getting enlightened and participating on those modern developments, they also seem to constitute an expensive and low-yield investment of effort and time. My sincere hope is that someone more knowledgeable and competent than I would present the subject in simple and readable form (i.e., without epsilonics and/or set theoretic straightjackets)—*that* would be a real contribution! What us nonmathematicians need today is (inverting the title of F. Klein's famous book) *advanced mathematics, and mechanics, from an elementary viewpoint,* even at the expense of rigor; otherwise we will keep drifting towards a dynamical tower of Babel. Perhaps B. Ravindra would choose to carry out such a useful task!

Enhanced Elastic Buckling Loads of Composite Plates With Tailored Thermal Residual Stresses[1]

Ch. Sarath Babu and T. Kant[2]. The authors present an innovative idea of tailoring the manufacturing-introduced thermal residual stresses in order to enhance the buckling strength of laminated composite plates and have given some interesting results. To demonstrate the effect of the tailored thermal residual stresses on the buckling load of symmetric composites, the authors have considered a square graphite/epoxy plates with two types (types I and II) of reinforcement that introduce nonzero thermal residual stress resultants when the plates cool to room temperature after the cure process. The basic laminate considered is a $[90/0]_s$ cross-ply having a/h ratio equal to 600 (a = 360 mm; thickness of each ply, t = 0.15 mm) where a and h denote width and thickness of the laminate, respectively.

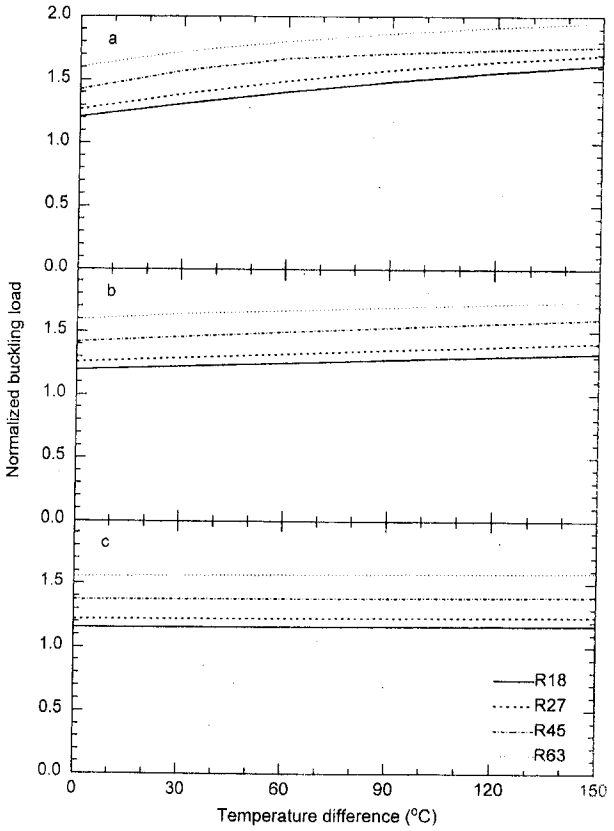


Fig. 1 Normalized buckling loads of plates (type I) R18, R27, R45, and R63 as a function of temperature difference; (*a*) a/h = 600, (*b*) a/h = 300, and (*c*) a/h = 100

Buckling loads of these plates are determined using finite element formulation within the context of linear stability analysis and the eigenvalue problem for λ is obtained as

$$\{[K] + [K_G^R] - \lambda[K_G^Q]\}\{\delta\} = 0 \qquad (1)$$

where $[K]$, $[K_G^R]$, and $[K_G^Q]$ are, respectively, the global stiffness matrix, global geometric stiffness matrix due to thermal residual stress resultants, and global geometric stiffness matrix due to stress resultants caused by displacement loading.

The buckling loads of reinforced plates are normalized with respect to a uniform $[90/0]_s$ plate having the same mass as that of the reinforced plate. The authors have shown that the normalized buckling loads of plates with type I reinforcement increase by about 30 percent and that of plates with type II reinforcement increase by about 300 percent, as the temperature difference is increased from 0°C to 150°C. Though the results presented are interesting, the effect of the thickness of the laminate on enhancement of the buckling load is not presented. Since the authors have considered a very thin laminate (a/h = 600), the increase in normalized buckling load due to increase in temperature difference is very high. However, in case of laminates having higher thickness than the one considered, the

[3] J. G. Papastavridis, School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405.

[1] By S. F. Muller de Almeida and J. S. Hansen and published in the Dec. 1997 issue of the ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 772–780.

[2] Department of Civil Engineering, ITT Bombay, Powai, Mumbai-400 076, India. e-mail: tkant@civil.iitb.ernet.in

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

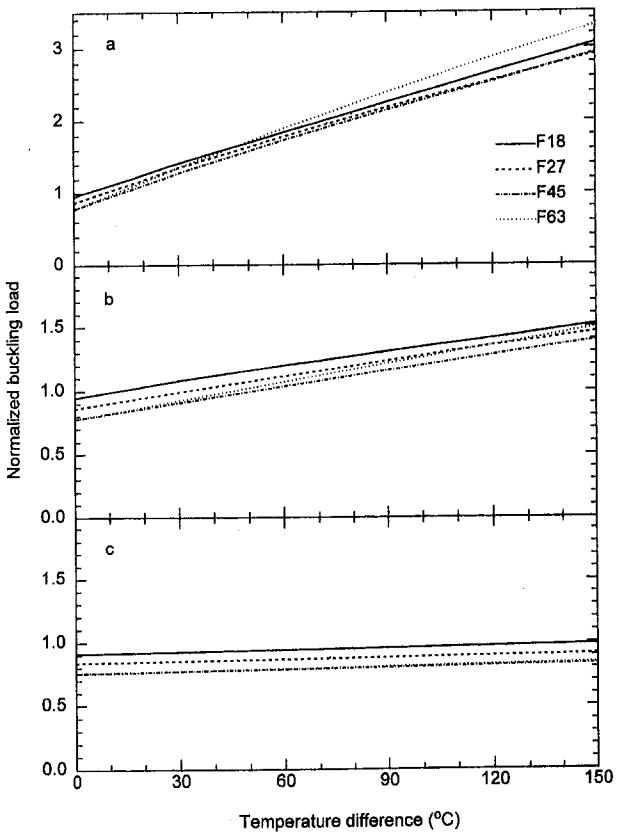


Fig. 2 Normalized buckling loads of plates (type II) F18, F27, F45, and F63 as a function of temperature difference; (*a*) a/h = 600, (*b*) a/h = 300, and (*c*) a/h = 100

increase in normalized buckling load shall not be as much as the authors have demonstrated. The reason for this is as the thickness of plate increases, the contribution of $[K_G^R]$ to buckling strength becomes smaller and smaller when compared to that of $[K]$, because thermal residual stress resultants are proportional to thickness h of the laminate, whereas bending stiffness is proportional to h^3. So in case of thick laminates, the increase in values of $[K_G^R]$ due to increase in temperature difference shall not result in enhancement of the normalized buckling load to the extent demonstrated by the authors.

We illustrate here the thickness effect on the buckling load in the presence of thermal residual stresses. The same first-order shear deformation theory considered by the authors is utilized to develop a finite element code but with a nine-node Lagrangian element. A selective integration scheme (3×3 for membrane and bending energy terms, 2×2 for shear energy terms) is used to evaluate stiffness matrices. Plates with both types (I and II) of reinforcement are analyzed with three different values of ply thicknesses t: 0.15 mm (a/h = 600), 0.30 mm (a/h = 300), and 0.90 mm (a/h = 100), keeping the width of the plate (a = 360 mm) as constant. The complete plate is discretized into 36 (6×6 grid) elements. The boundary conditions, the material properties, and the method of loading considered here are the same as used by the authors.

The normalized buckling loads for plates with reinforcement type I is given in Fig. 1. Figure 1(a) shows the normalized buckling loads for plates having a/h equal to 600, and the plots are exactly the same as obtained by the authors, thus demonstrating the accuracy of the code developed. Figures 1(b) and 1(c) show the normalized buckling loads for plates having a/h ratios 300 and 100, respectively. It can be seen that the normalized buckling loads of plates having a/h values 600, 300, and 100 increase by about 30 percent, 10 percent, and 2 percent, respectively, as the temperature difference is increased from 0°C to 150°C.

Figure 2 shows the normalized buckling loads for plates with type II reinforcement. Figures 2(a), 2(b), and 2(c) show the normalized buckling loads for plates having a/h ratios 600, 300, and 100, respectively. It can be observed that the normalized buckling loads for plates having a/h ratios 600, 300, and 100 increase by about 300 percent, 60 percent, and 8 percent, respectively, as the temperature difference is increased from 0°C to 150°C. It is to be noted that as the temperature difference increases, the reinforced plates having a/h ratios equal to 600 and 300 become more efficient when compared to uniform plates because of the effect of residual stresses. However, the normalized buckling loads of plates having a/h ratio equal to 100 are always less than unity for all the values of temperature difference, thus indicating the ineffectiveness of such reinforced plates in spite of the presence of thermal residual stresses.

The results presented here demonstrate that the enhancement of buckling load due to the presence of thermal residual stresses is significant in case of very thin laminates. However, as the thickness of the plate increases, the increase in buckling strength is not significant. In fact, for plates with a/h equal to 100, the enhancement in buckling load is almost negligible. So the concept of tailoring the thermal residual stresses for improving the buckling strength of composite laminates gives encouraging results for very thin laminates, but does not do so for plates as the thickness increases. We nevertheless congratulate the authors for suggesting an interesting idea of tailoring the thermal residual stresses for enhancing the buckling strength of composite laminates.

Reference

Muller de Almeida, S. F., and Hansen, J. S., 1997, "Enhanced Elastic Buckling Loads of Composite Plates With Tailored Thermal Residual Stresses," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 772–780.

Author's Closure[3]

The authors appreciate the interest shown by Professors Babu and Kant in our work. We completely agree with their point that the thermal effects discussed in our work are reduced as a plate gets thicker. It is important, however, to consider what realistic a/h ratios are. It is our experience, related to aerospace structures, that a ratio of a/h = 600 is considered very thin but in the practical range, particularly when space applications are considered. On the other hand, an a/h = 100 is considered to be quite thick for use in control surfaces for example. Therefore, practical values of a/h are in a range where our and their calculations indicate thermal effects are significant. Another point which should be noted is that our illustrative example is a panel with stiffeners only on the edges. The problem of analyzing an actual plate or shell structure which is reinforced by a gridwork of stiffeners is more complex.

In such a situation and when considering global buckling, for example, the dimensions of interest are the overall dimensions of the structure and not the dimensions of a single panel. The effects we discuss will, of course, come into play in such a situation as well. Thus we feel that for many aerospace applications the manufacturing induced thermal effects we discuss can be potentially very significant, either in a positive or negative sense, and it is important these effects be taken into account when appropriate.

[3] Jorn S. Hansen, University of Toronto, Institute for Aerospace Studies, 4925 Dufferin Street, Downsview, Ontario M3H 5T6, Canada.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

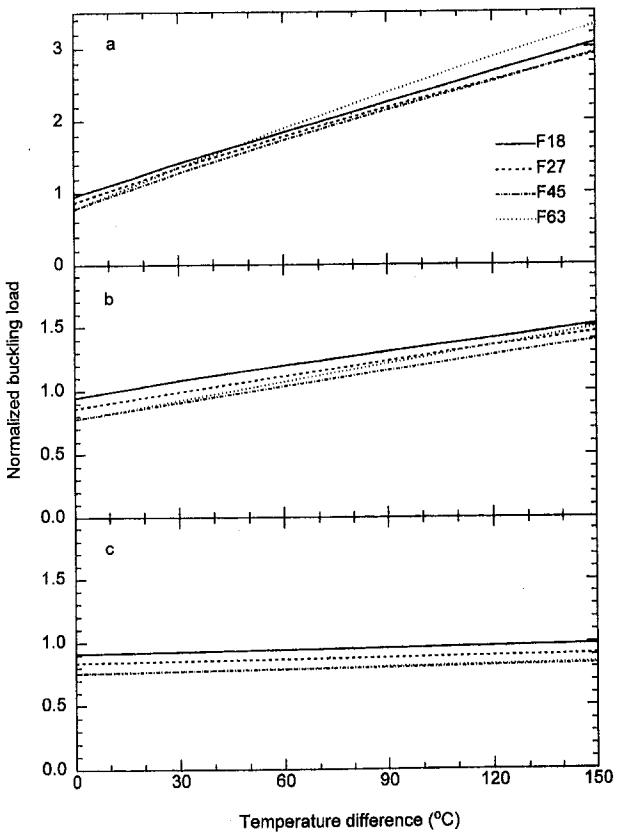


Fig. 2 Normalized buckling loads of plates (type II) F18, F27, F45, and F63 as a function of temperature difference; (*a*) a/h = 600, (*b*) a/h = 300, and (*c*) a/h = 100

increase in normalized buckling load shall not be as much as the authors have demonstrated. The reason for this is as the thickness of plate increases, the contribution of $[K_G^R]$ to buckling strength becomes smaller and smaller when compared to that of $[K]$, because thermal residual stress resultants are proportional to thickness h of the laminate, whereas bending stiffness is proportional to h^3. So in case of thick laminates, the increase in values of $[K_G^R]$ due to increase in temperature difference shall not result in enhancement of the normalized buckling load to the extent demonstrated by the authors.

We illustrate here the thickness effect on the buckling load in the presence of thermal residual stresses. The same first-order shear deformation theory considered by the authors is utilized to develop a finite element code but with a nine-node Lagrangian element. A selective integration scheme (3×3 for membrane and bending energy terms, 2×2 for shear energy terms) is used to evaluate stiffness matrices. Plates with both types (I and II) of reinforcement are analyzed with three different values of ply thicknesses t: 0.15 mm (a/h = 600), 0.30 mm (a/h = 300), and 0.90 mm (a/h = 100), keeping the width of the plate (a = 360 mm) as constant. The complete plate is discretized into 36 (6×6 grid) elements. The boundary conditions, the material properties, and the method of loading considered here are the same as used by the authors.

The normalized buckling loads for plates with reinforcement type I is given in Fig. 1. Figure 1(a) shows the normalized buckling loads for plates having a/h equal to 600, and the plots are exactly the same as obtained by the authors, thus demonstrating the accuracy of the code developed. Figures 1(b) and 1(c) show the normalized buckling loads for plates having a/h ratios 300 and 100, respectively. It can be seen that the normalized buckling loads of plates having a/h values 600, 300, and 100 increase by about 30 percent, 10 percent, and 2 percent, respectively, as the temperature difference is increased from 0°C to 150°C.

Figure 2 shows the normalized buckling loads for plates with type II reinforcement. Figures 2(a), 2(b), and 2(c) show the normalized buckling loads for plates having a/h ratios 600, 300, and 100, respectively. It can be observed that the normalized buckling loads for plates having a/h ratios 600, 300, and 100 increase by about 300 percent, 60 percent, and 8 percent, respectively, as the temperature difference is increased from 0°C to 150°C. It is to be noted that as the temperature difference increases, the reinforced plates having a/h ratios equal to 600 and 300 become more efficient when compared to uniform plates because of the effect of residual stresses. However, the normalized buckling loads of plates having a/h ratio equal to 100 are always less than unity for all the values of temperature difference, thus indicating the ineffectiveness of such reinforced plates in spite of the presence of thermal residual stresses.

The results presented here demonstrate that the enhancement of buckling load due to the presence of thermal residual stresses is significant in case of very thin laminates. However, as the thickness of the plate increases, the increase in buckling strength is not significant. In fact, for plates with a/h equal to 100, the enhancement in buckling load is almost negligible. So the concept of tailoring the thermal residual stresses for improving the buckling strength of composite laminates gives encouraging results for very thin laminates, but does not do so for plates as the thickness increases. We nevertheless congratulate the authors for suggesting an interesting idea of tailoring the thermal residual stresses for enhancing the buckling strength of composite laminates.

Reference

Muller de Almeida, S. F., and Hansen, J. S., 1997, "Enhanced Elastic Buckling Loads of Composite Plates With Tailored Thermal Residual Stresses," ASME JOURNAL OF APPLIED MECHANICS, Vol. 64, pp. 772–780.

Author's Closure[3]

The authors appreciate the interest shown by Professors Babu and Kant in our work. We completely agree with their point that the thermal effects discussed in our work are reduced as a plate gets thicker. It is important, however, to consider what realistic a/h ratios are. It is our experience, related to aerospace structures, that a ratio of a/h = 600 is considered very thin but in the practical range, particularly when space applications are considered. On the other hand, an a/h = 100 is considered to be quite thick for use in control surfaces for example. Therefore, practical values of a/h are in a range where our and their calculations indicate thermal effects are significant. Another point which should be noted is that our illustrative example is a panel with stiffeners only on the edges. The problem of analyzing an actual plate or shell structure which is reinforced by a gridwork of stiffeners is more complex.

In such a situation and when considering global buckling, for example, the dimensions of interest are the overall dimensions of the structure and not the dimensions of a single panel. The effects we discuss will, of course, come into play in such a situation as well. Thus we feel that for many aerospace applications the manufacturing induced thermal effects we discuss can be potentially very significant, either in a positive or negative sense, and it is important these effects be taken into account when appropriate.

[3] Jorn S. Hansen, University of Toronto, Institute for Aerospace Studies, 4925 Dufferin Street, Downsview, Ontario M3H 5T6, Canada.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Toroidal Vibrations of Anisotropic Spheres With Spherical Isotropy[1]

W. Q. Chen[2]. Professor K. T. Chau recently extended Hu's theory (Hu, 1954) by introducing two wave potentials and derived an exact solution for the free toroidal vibrations of anisotropic spheres with spherical isotropy; some interesting and valuable numerical results were also presented. However, the description in the Introduction that there is no general theory for dynamics problems of spherically isotropic materials (p. 60, lines 4 and 5) is not true. In fact, as early as 1972, Cohen and Shah (1972) presented an exact analysis on the general free vibrations of spherically isotropic hollow sphere by introducing two auxiliary variables. They obtained the toroidal frequency equation of a hollow sphere and from which the one of a solid sphere (Eq. (33) as in Chau (1998)) is obvious. The remained unsolved equations, Eqs. (35) and (36) in Chau (1998), were also obtained by Cohen and Shah. For the special case that $n = 0$, a Bessel function solution is available and, for $n > 0$, the equations are coupled and the solutions were derived by the well-known Frobenius power series (FPS) method (Cohen and Shah, 1972). However, the traditional FPS method seems to be a little confusing in giving complete solutions of such equations. A more effective solution method based on Matrix formulae was thus developed recently by Ding et al. (1995), and the complete solutions of the coupled system (Eqs. (35) and (36) in Chau (1998)) for all possible cases were easily derived. Because of the general property of the FPS method, the solution can be applied in the analyses of vibrations, not only for spherical shells but also for solid spheres (Chen, 1996).

The Lame separation technique (Eq. (12)) employed by Chau (1998) is not unknown; in fact, Ding and Chen have used an improved formula in the analyses of the coupled vibration problems (Ding and Chen, 1994; Ding and Chen, 1996a, b) as follows:

$$u_\theta = -\frac{1}{\sin\theta}\frac{\partial\psi}{\partial\phi} - \frac{\partial G}{\partial\theta}, \quad u_\phi = \frac{\partial\psi}{\partial\theta} - \frac{1}{\sin\theta}\frac{\partial G}{\partial\phi},$$

$$u_r = w. \tag{1}$$

The superiority of the above formula over Eq. (12) in Chau (1998) is obvious since any further involved differential manipulations can be greatly simplified. Shu'lga et al. (Shu'lga, 1980; Shu'lga et al., 1988) also have used similar formulae to simplify the basic dynamic equations of a spherically isotropic medium and have considered the general nonaxisymmetric vibrations of spherical shells. In contrast to using the FPS method, Shu'lga et al. expressed the solutions of the coupled system in terms of polynomial expansions. However, their solutions cannot be applied in the case of a solid sphere.

Ding and Chen have employed formula (1) to analyze various coupled free vibration problems of spherically isotropic hollow spheres, for example, when submerged in incompressible fluid (Ding and Chen, 1994) and compressible fluid (Ding and Chen, 1996b) and when embedded in an elastic medium (Ding and Chen, 1996a). The free vibrations of multilayered spherical shells have also been considered in Ding et al. (1994). It is found that the vibrations of these cases can all be divided into two classes as in the case of a vacuum. More details and discussions can be found in the writer's Ph.D. dissertation (Chen, 1996).

[1] By K. T. Chau and published in the Mar. 1998 issue of the ASME JOURNAL OF APPLIED MECHANICS, Vol. 65, pp. 59–65.

[2] Department of Civil Engineering, Zhejiang University, Hangzhou 310027, P. R. China. (Current address: Department of Aeronautics and Astronautics, Faculty of Engineering, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan. e-mail: tt79326@hongo.ecc.u-tokyo.ac.jp)

References

Chau, K. T. 1998, "Toroidal Vibrations of Anisotropic Spheres With Spherical Isotropy," ASME JOURNAL OF APPLIED MECHANICS, Vol. 65, pp. 59–65.

Chen, W. Q., 1996, "Coupled free vibrations of spherically isotropic hollow spheres," Ph.D. dissertation, Zhejiang University, Hangzhou, China (in Chinese).

Cohen, H., and Shah, A. H., 1972, "Free vibrations of a spherically isotropic hollow sphere," *Acustica,* Vol. 26, pp. 329–333.

Ding, H. J., and Chen, W. Q., 1994, "Free vibrations of submerged spherically isotropic spherical shells," *Acta Solida Mechanica Sinica,* Vol. 15, pp. 39–48 (in Chinese).

Ding, H. J., Ren, Y. J., Zou, D. Q., and Chen, W. Q., 1994, "Displacement solution method of spherically isotropic elasticity," *Acta Mechanica Sinica,* Vol. 26, pp. 186–197 (in Chinese).

Ding, H. J., and Chen, W. Q., 1996a, "Nonaxisymmetric free vibrations of a spherically isotropic spherical shell embedded in an elastic medium," *International Journal of Solids and Structures,* Vol. 33, pp. 2575–2590.

Ding, H. J., and Chen, W. Q., 1996b, "Natural frequencies of an elastic spherically isotropic hollow sphere submerged in a compressible fluid medium," *Journal of Sound and Vibration,* Vol. 192, pp. 173–198.

Ding, H. J., Chen, W. Q., and Liu, Z., 1995, "Solutions to equations of vibrations of spherical and cylindrical shells," *Applied Mathematics and Mechanics,* Vol. 16, pp. 1–15.

Hu, H. C., 1954, "On the general theory of elasticity for a spherically isotropic medium," *Acta Scientia Sinica,* Vol. 3, pp. 247–260.

Shu'lga, N. A., 1980, "Natural vibrations of a hollow transversely isotropic sphere," *Prikladnaya Meknanika,* Vol. 16, No. 12, pp. 124–127.

Shu'lga, N. A., Grigorenko, A. Y., and Efimova, T. L., 1988, "Free non-axisymmetric oscillations of a thick-walled, nonhomogeneous, transversely isotropic, hollow sphere," *Prikladnaya Meknanika,* Vol. 24, No. 5, pp. 12–17.

Author's Closure[3]

I would like to thank Prof. W. Q. Chen (hereafter referred as the discusser) for expressing his/her interest in my publication and for indicating to me that an analytic solution on a coupled system equivalent to the coupled system of (39) – (40) given in Chau (1998) has, in fact, been obtained by Cohen et al. (1972) using Frobenius power series and reconsidered by Ding et al. (1995) using a matrix formulation. Note, however, that the reference of Cohen et al. (1972) given by the discusser is incorrect, as the third author C. V. Ramakrishnan is missing in the Reference section of the discussion. Thus, the discusser incorrectly cited the reference as Cohen and Shah (1972), instead of Cohen et al. (1972). The solution given by Ding et al. (1995) appears to be more comprehensive than those given by Cohen et al. (1972). In particular, they identified seven possible scenarios of the solution form for the coupled system, depending on the combinations of the elastic constants and the frequency of vibrations.

However, the discusser's statement that "Eqs. (35) and (36) in Chau (1998) have also been obtained by Cohen and Shah" is not precise as the coupled system (35) – (36) cannot be found anywhere in Cohen et al. (1972), although the coupled system given by (6) of Cohen et al. (1972) can be shown to be equivalent to the system (39) – (40) of Chau (1998) (see later discussion). In addition, the discusser claimed that the displacement functions given in (1) of the discussion, which is also given in (4) of Ding et al. (1994), (2) of Ding and Chen (1996a) and in (2) of Ding and Chen (1996b), are obviously superior to the wave potentials used in (12) of Chau (1998). However, no substantiation was given to support the discusser's claim. In this closure I will first show that the system of coupled Eqs. (39) – (40) derived independently by Chau (1998) is in fact equivalent to the system (6) of Cohen et al. (1972). Secondly, I will demonstrate that the displacement functions given in (1) of the discussion is *not* superior to the wave potential functions given in (12) of Chau (1998).

In particular, for the "vibrations of the second class" Cohen et al. (1972) obtained the following coupled system:

[3] K. T. Chau, Associate Professor, Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Yuk Choi Road, Hung Hom, Hong Kong, China. e-mail: cektchau@polyu.edu.hk

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Journal of Applied Mechanics

Book Review

7R31. AMR Infobase of Journal Literature on CD-ROM, 1989–1997. ASME, New York. 1997. CD-ROM, Windows Format. ISBN 1088-8535. ASME Book No AMR98C. $499.00. (ASME members $249.00). Combination price for journal and CD-ROM: $913.00; (ASME members $375.00).[1]

REVIEWED BY M. S. QATU[2] AND C. R. STEELE[3]

This newly released Compact Disk (CD) from ASME contains all the journal literature in mechanics that was abstracted by *Applied Mechanics Reviews* (*AMR*) for the last nine years from over 450 international journals. The CD is in its third release. The price for ASME members is $249.00 with $100.00 annual renewal. Institution price is $499.00, with a $250.00 annual renewal. The software is available for PCs running *Windows*, but can also be run satisfactorily on MACs using Windows simulation software. This release contains all data from the previous CD as well as that of the past year (1989–1997). The intent is to release a new CD annually in March, each year adding another year's worth of data.

The CD can be used on a local area network (LAN) for access by ten simultaneous users. Additional licenses can be purchased for $67.50 each to bring that capability up to as many as 127 simultaneous users.

This is an excellent source of information for all engineers whose work relates to mechanics. In minutes, one can search the literature on just about any subject related to mechanics and obtain a list of articles, their citing, authors, author addresses as well as abstracts. This took several weeks if not months just a few years ago.

The database presents all the bibliographic information in the same format of the published journal. The full review articles usually published in the journal itself are not included on the CD. However, the abstracts of those *AMR* articles are included. One can use two major methods to search for articles. The first is to use the *Query* button. This will open a window in which a keyword can be entered. These keywords can relate to a subject being investigated, an author, a country, a university, or other information needed. Several keywords can be used at the same time using the logical expressions "and," "not" as well as "or" to expand or narrow the search results. The second method to do a search is by using the templates. A typical template contains the *AMR* issue and year, article ID, author, keyword in the abstract, as well as other information. One may fill in the information partially or completely to obtain all the articles related to the criteria on which he or she is searching. Researchers familiar with the *AMR* Classification Scheme designations can also make a useful search of those codes to target journal articles on a specific subject area.

The software also comes with other features that will make the search easier. For example, one can choose to view all of the search results (hits) in one document. This document can be printed directly from within the software. Other features include *backtrack* which allows the user to navigate through previous queries made in one search. Other features like *copy* and *paste*, will allow one to attach a certain component to your word processing document. Features like *Bookmark*, *Highlighter*, and *Note*, will allow one to add a note to a special article, or highlight it. To make full use of these features, the user can create a *Shadow File* which functions like an invisible overlay. One can create a *Note* over a certain bibliographic record which can later be saved to diskette. The original CD remains untouched. In the user's next session, he can open up the database along with his *Shadow File*, and all of his notes will still be there for referral.

The software is of great value to all engineers, researchers, educators, and journal editors. An engineer can use it to find if any research has been done on a problem of interest. For example, one author of this review was recently involved in a project about the vibration of electric motors.

Using the software, a query made with the keywords "vibration," "electric," and "motor" revealed 17 papers on the subject.

One must be careful and should enter possible combinations of a certain word. A civil engineer may be interested in finding out if any research has been done on vibration of concrete beams. He can enter the keywords *vibration* or *vibrations* and *concrete* and *beam* or *beams*, and press the *apply* button to obtain a list of the 13 papers which appeared in the past nine years on the subject. Such a task could have taken weeks if the same engineer was to search every issue of *AMR* for the subject. As shown before, the searcher should be careful and input all the possible combinations of a word to achieve his objective. With the use of the logical expressions "and" and "or," this is indeed a simple task.

Literature reviews should be easier with this software. The last address of an author can also be found relatively easily using this software. This information is important for researchers, as well as journal editors who may be looking for appropriate people to review or referee articles on a particular subject. The editors could do a search for authors of recent articles on the same subject who may make a good reviewer or referee. Complete author addresses are included in the *AMR* bibliographic citations, making it easy to contact the prospective reviewer/referee. The advantages of the *AMR Infobase* are particularly noticed when returning to other literature data bases, such as Engineering Index and InSpec, for topics not covered in *AMR*.

If issues of *AMR* prior to 1989 were also included, it would make the search more comprehensive. With current computer technology like Optical Character Recognition (OCR), this may not be a difficult task. Finally, issuing an Apple version would

[1] Reprinted for ASME *Applied Mechanics Reviews*, Vol. 51, No. 7, July 1998.
[2] Echlin Automotive, 1900 Opdyke Court, Auburn Hills, MI 48326.
[3] Applied Mechanics Division, Stanford University, Durang Building, Stanford, CA 94305.

 Copyright © 1998 by ASME

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

be of considerable help to the Apple/Macintosh users, since it would be a pleasure not dealing with additional simulation software and the idiosyncrasies of Windows.

These reviewers recommend that libraries and engineering departments purchase the *AMR Infobase of Journal Literature* on CD-ROM for use on their LAN. Individual researchers may also want to purchase the CD. The cost of the CD is minimal when considering the time saved doing searches. In addition, having a personal copy of the CD gives much quicker results when compared to online searches.

Downloaded 04 May 2010 to 171.66.16.26. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm