ASME
SETTING THE STANDARD
1880 2005

Transactions of the ASME, Journal of Applied Mechanics (ISSN 0021-8936) is published bimonthly (Jan., Mar., May, July, Sept., Nov.) by The American Society of Mechanical Engineers, Three Park Avenue, New York, NY 10016. Periodicals postage paid at New York, NY and additional mailing offices. POSTMASTER: Send address changes to Transactions of the ASME, Journal of Applied Mechanics, c/o THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 22 Law Drive, Box 2300, Fairfield, NJ 07007-2300. **CHANGES OF ADDRESS** must be received at Society headquarters seven weeks before they are to be effective. Please send old label and new address.

Journal of Applied Mechanics

Published Bimonthly by ASME

VOLUME 72 • NUMBER 3 • MAY 2005

TECHNICAL PAPERS

BRIEF NOTE

(Contents continued on inside back cover)

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

Journal of Applied Mechanics Volume 72, Number 3 MAY 2005

DISCUSSION

ANNOUNCEMENTS AND SPECIAL NOTES

The ASME Journal of Applied Mechanics is abstracted and indexed in the following:

Alloys Index, Aluminum Industry Abstracts, Applied Science & Technology Index, Ceramic Abstracts, Chemical Abstracts, Civil Engineering Abstracts, Compendex (The electronic equivalent of Engineering Index), Computer & Information Systems Abstracts, Corrosion Abstracts, Current Contents, EEA (Earthquake Engineering Abstracts Database), Electronics & Communications Abstracts Journal, Engineered Materials Abstracts, Engineering Index, Environmental Engineering Abstracts, Environmental Science and Pollution Management, Fluidex, Fuel & Energy Abstracts, GeoRef, Geotechnical Abstracts, INSPEC, International Aerospace Abstracts, Journal of Ferrocement, Materials Science Citation Index, Mechanical Engineering Abstracts, METADEX (The electronic equivalent of Metals Abstracts and Alloys Index), Metals Abstracts, Nonferrous Metals Alert, Polymers Ceramics Composites Alert, Referativnyi Zhurnal, Science Citation Index, SciSearch (Electronic equivalent of Science Citation Index), Shock and Vibration Digest, Solid State and Superconductivity Abstracts, Steels Alert, Zentralblatt MATH

Ultimate Response of Composite Cylinders Under Flexural Load

Zheng-Ming Huang
Professor
School of Aeronautics,
Astronautics & Mechanics,
Tongji University,
1239 Siping Road,
Shanghai 200092,
People's Republic of China
e-mail: huangzm@mail.tongji.edu.cn

Composite cylinders are generally used as primary load carrying structures. Their constitutive behavior up to failure is crucial for a critical design. This paper focuses on the ultimate flexural strength of a polymer based composite cylinder subjected to bending. In such a case, the outmost filament of the cylinder subjected to the maximum bending stress fails the first. The complexity, however, lies in the fact that the failure of this outmost filament generally does not correspond to the ultimate failure. Additional loads can still be applied to the cylinder and a progressive failure process will result. To deal with such a problem in this paper, the cylinder is discretized into a number of lamina layers with different widths. The bridging micromechanics model [Huang, Z. M., Composites Part A, 2001] combined with the classical lamination theory has been applied to understand the progressive failure process generated in the cylinder. Only its constituent fiber and matrix properties under bending are necessary for this understanding and reasonably good accuracy has been achieved. However, the ultimate failure of the cylinder cannot be figured out only based on a stress failure criterion, as one cannot know a priori which ply failure corresponds to the ultimate failure. An additional critical deflection (curvature) condition must be employed also. By using both the stress and the deflection conditions, the estimated ultimate strength of the cylinder agreed well with an experimental measurement. [DOI: 10.1115/1.1867990]

1 Introduction

The use of composite cylinders as primary-load carrying structures has been tremendous in many engineering fields such as aerospace, automobile, chemical and energy, civil and infrastructure, sports and recreation, and even biomedical engineering. In biomedical applications, for example, composite cylinders have been proposed for spine rods, dental posts, total hip replacement stems, etc [1]. An orthodontic archwire can be best developed using a continuous fiber reinforced polymer matrix composite rod [2–4]. This is because the polymer matrix material, which is available from numerous candidates, can offer the aesthetic feature of the archwire comparable to the patient's teeth, whereas its various mechanical properties at different using stages can be achieved by choosing suitable fiber material, fiber content, and fiber arrangement pattern. The composite archwire is generally produced through pultrusion of a resin-impregnated fiber yarn bundle into a curing die [2–4]. An optimal design for the composite archwire will depend on a complete understanding for its structure-property relationship. As the composite archwires, and further most other composite cylinders, are mainly subjected to flexural loads, their bending behavior especially their ultimate bending load-carrying ability must be well understood.

Several attempts have been made to obtain the effective properties of composite cylinders. Bhattacharyya and Appiah considered a single (softer) fiber cylinder embedded in an annular matrix cylinder subjected to lateral load and obtained its exact elastoplastic response solution [5]. Their results, however, are not directly applicable to the present case. The reason is that in a real application many fiber filaments are gathered together and the failure of the outmost filament subjected to the maximum bending stress does not mean the ultimate failure of the composite cylinder. Starbuck performed a stress analysis for laminated composite cylinders under nonaxisymmetric loading and derived a closed-form solution for the cylinder responses up to the first-ply failure [6]. However, limited work on the simulation of the ultimate flexural failure of composite cylinders could be found in the literature. Other relevant references have been on the flexural analysis of laminated composites, but were mainly focused on the development of linear and geometrically nonlinear (large deflection) theories [7–12], also applicable before the laminate first-ply failure [13]. While favorable comparison in stiffness (elastic bending modulus) with experimental data with those theories was reported, less agreement in ultimate bending strength has been found in general [7,10–13]. A main reason is that the material nonlinearity has not been taken into account or has not been well addressed. Most studies used the elastic moduli of the laminae to define the laminated beam stiffness up to failure. The failure status of a lamina in the laminate was detected in terms of a phenomenological stress failure criterion such as the maximum stress–strain or the Tsai–Wu criterion based on the critical strength parameters measured from an individual lamina. It is known that the single-layer lamina during measurement of the strength parameters is generally statically determinate. However, the lamina becomes statically indeterminate in the laminate. As most composites can undergo nonlinear deformation before failure especially when subjected to bending, the use of the constant-stiffness elements up to failure would cause errors in the determination of the stresses shared by each lamina involved. Additional complexity lies in the fact that the failure of the outmost layer subjected initially to the maximum bending stress does not imply the ultimate failure of the whole laminate, nor the central ply failure implies such ultimate failure. Indeed, a laminated beam generally attains its ultimate bending strength at the failure of an intermediate ply [14]. After this intermediate ply failure, which is defined as the ultimate failure, the bending load sustained by the laminate lowers down. The bending stress of the laminate corresponding to the last-ply failure, if any, is lower than the ultimate bending strength. However, the stress failure criterion can only detect an individual ply failure. By incorporating the stress failure criterion with an incremental solution strategy, the predicted strength of a latter ply failure is always greater than or equal to the predicted strength of an earlier

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, April 21, 2002; final revision, October 28, 2003. Associate Editor: B. M. Moran. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California, Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

'ply failure. This means that the predicted progressive failure strengths after the ultimate failure cannot be correct if only a stress failure criterion is used. As one does not know a priori which intermediate ply failure corresponds to the ultimate failure, an additional controlling parameter, which should be a deflection related quantity, must be employed to characterize the ultimate failure. This is different from an in-plane load situation where the ultimate failure of the laminate generally occurs at its last-ply failure, thereby the ultimate failure position (i.e., the last ply) has already been characterized no matter how much load will be applied, and hence only a stress failure criterion is sufficient. This also suggests that the determination of the ultimate bending strength of the beam depends on an accurate calculation of its deflection.

In the present paper, the composite cylinder is discretized into a number of parallel laminae along its axis direction with different widths. In this way, the analysis of the cylinder is converted to that of a laminated composite beam. The classical lamination theory is applied to determine the stresses shared by each lamina in the laminate, whereas the lamina local analysis is performed by using the bridging model [15]. This is because the bridging model can provide the lamina instantaneous stiffness matrix up to failure and can explicitly give the internal stresses in the constituent fiber and matrix materials of the lamina at every load level. Thus, a stress failure criterion applied to the constituent rather than to the composite level will suffice: The lamina is considered to fail once a constituent has failed. This is important for applications, as only constituent properties are required, which are easily obtainable (either taken from a material data source such as for fiber material or measured from monolithic material specimens such as for matrix material).

As aforementioned, the laminate subjected to bending generally attains its ultimate strength after the first-ply and before the last-ply failures. This is also true for a composite cylinder. The load-deflection curve is downward after the ultimate failure. In order to figure out which discretized ply failure corresponds to the ultimate failure, we must provide an additional controlling parameter, i.e., the critical deflection or curvature condition. The critical deflection or curvature is that at which the ultimate bending load is measured.

2 Global Analysis

2.1 Discretization. For a specific purpose, the composite cylinder under consideration is circular, consisting of continuous fibers arranged in the same direction and the surrounding polymer matrix. However, the analysis procedure thus developed can be equally well applicable to other cross-sectional shapes. When the cylinder is subjected to an axial load, it can be regarded as a unidirectional composite the analysis of which is relatively easy [15,16]. However, when the cylinder is subjected to a lateral (bending) load, the problem becomes much more complicated. In such a case, the stress distribution on the cylinder cross section is not uniform. The outmost filament subjected to the maximum bending stress fails the first. The complexity is due to the fact that the failure of this outmost filament generally does not correspond to the ultimate failure. Additional higher load can still be applied. Thus, a progressive failure process, somewhat similar to that generated in a laminate, will result. This has to be understood before the ultimate bending strength of the composite cylinder can be obtained.

In order to track the progressive failure process in the unidirectional composite cylinder, let us imaginatively separate it into a number of lamina layers along the axial direction. A cross-sectional discretization is shown in Fig. 1. Take a global coordinate system, (X,Y,Z), where X is along the cylinder axis and X,Z is the plane on which the bending load is applied. Suppose that the cross section of the circular cylinder is discretized into $N=2n$ layers of an equal thickness, t, given by (see Fig. 1)

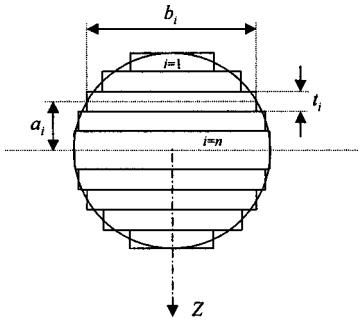


Fig. 1 A composite cylinder consisting of multilayers of laminas

$$t = d/N = d/(2n) \tag{1}$$

where d is the cylinder diameter. Referring to Fig. 1, the width of the ith layer, b_i, is determined from

$$b_i = \sqrt{d^2 - 4a_i^2} = \sqrt{d^2 - 4[(n-i)t + 0.5t]^2}, \quad i = 1, \ldots, n \tag{2}$$

Now, the composite cylinder can be regarded as a laminated composite subjected to an external (in-plane or out of plane) load. Note that in the present case, the longitudinal directions of all the laminae coincide with the global X direction, whereas the other two transverse directions of the laminas can be chosen along the Y and Z directions, respectively. Thus, the global coordinate system is coincident with the local systems of all the laminas.

2.2 Classical Lamination Theory. After the discretization, the classical lamination theory is applicable to the overall analysis. Corresponding to the present load condition, only the in-plane stress and strain increments, $\{d\sigma\}=\{d\sigma_{XX}, d\sigma_{YY}, d\sigma_{XY}\}^T$ and $\{d\varepsilon\}=\{d\varepsilon_{XX}, d\varepsilon_{YY}, 2d\varepsilon_{XY}\}^T$, are retained (refer to Fig. 2). The averaged stress increments on the kth lamina can be determined from [15]

$$\{d\sigma\}_k = [(C_{ij})_k]\{d\varepsilon\}_k \tag{3}$$

where

$$[(C_{ij})_k] = ([S]_k)^{-1} \tag{4.1}$$

and

$$\{d\varepsilon\}_k = \left\{ d\varepsilon_{XX}^0 + \frac{Z_k + Z_{k-1}}{2} d\kappa_{XX}^0, d\varepsilon_{YY}^0 + \frac{Z_k + Z_{k-1}}{2} d\kappa_{YY}^0, 2d\varepsilon_{XY}^0 + (Z_k + Z_{k-1}) d\kappa_{XY}^0 \right\}^T \tag{4.2}$$

$d\varepsilon_{XX}^0$, $d\varepsilon_{YY}^0$, and $d\varepsilon_{XY}^0$ and $d\kappa_{XX}^0$, $d\kappa_{YY}^0$, and $d\kappa_{XY}^0$ are the laminate in-plane strain and curvature increments. Z_k and Z_{k-1} are the Z coordinates of the top and the bottom surfaces of the kth lamina. $[S]_k$ is the lamina instantaneous compliance matrix in its local system as given subsequently. From the condition that the resultants of the internal stresses given by Eq. (3) should be balanced with the overall applied loads on the cylinder, we obtain the following equation [15]

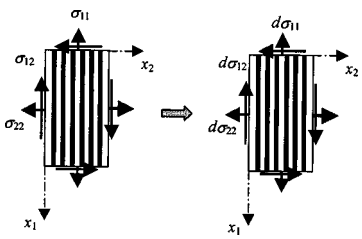


Fig. 2 Analysis of a lamina layer taken from the cylinder

$$\begin{Bmatrix} dN_{XX} \\ dN_{YY} \\ dN_{XY} \\ dM_{XX} \\ dM_{YY} \\ dM_{XY} \end{Bmatrix} = \begin{bmatrix} Q_{11}^I & Q_{12}^I & Q_{13}^I & Q_{11}^{II} & Q_{12}^{II} & Q_{13}^{II} \\ Q_{12}^I & Q_{22}^I & Q_{23}^I & Q_{12}^{II} & Q_{22}^{II} & Q_{23}^{II} \\ Q_{13}^I & Q_{23}^I & Q_{33}^I & Q_{13}^{II} & Q_{23}^{II} & Q_{33}^{II} \\ Q_{11}^{II} & Q_{12}^{II} & Q_{13}^{II} & Q_{11}^{III} & Q_{12}^{III} & Q_{13}^{III} \\ Q_{12}^{II} & Q_{22}^{II} & Q_{23}^{II} & Q_{12}^{III} & Q_{22}^{III} & Q_{23}^{III} \\ Q_{13}^{II} & Q_{23}^{II} & Q_{33}^{II} & Q_{13}^{III} & Q_{23}^{III} & Q_{33}^{III} \end{bmatrix} \begin{Bmatrix} d\varepsilon_{XX}^0 \\ d\varepsilon_{YY}^0 \\ 2d\varepsilon_{XY}^0 \\ d\kappa_{XX}^0 \\ d\kappa_{YY}^0 \\ 2d\kappa_{XY}^0 \end{Bmatrix} \tag{5}$$

$$Q_{ij}^I = \sum_{k=1}^{N} b_k (C_{ij})_k (Z_k - Z_{k-1}),$$

$$Q_{ij}^{II} = \frac{1}{2} \sum_{k=1}^{N} b_k (C_{ij})_k (Z_k^2 - Z_{k-1}^2),$$

$$Q_{ij}^{III} = \frac{1}{3} \sum_{k=1}^{N} b_k (C_{ij})_k (Z_k^3 - Z_{k-1}^3) \tag{6}$$

In Eq. (5), dN_{XX}, dN_{YY}, and dN_{XY} and dM_{XX}, dM_{YY}, and dM_{XY} are, respectively, the applied total in-plane force and moment increments on the laminate cross section.

Under a simple bending condition, the only nonzero quantity is the bending moment increment dM_{XX}. The middle plane deflection increment, dw^0, can be integrated from the equation

$$\frac{\partial^2 (dw^0)}{\partial X^2} = -d\kappa_{XX}^0, \tag{7}$$

together with proper boundary conditions. The total deflection is updated from $w^0 = w^0 + dw^0$.

2.3 Post-Failure Analysis. Apparently, different layers in the laminate carry different load shares. With the increase of the external load, some layers fail first before the others. Once some k_0th lamina layer fails, it can hardly sustain any additional load, in general. The additional external load must be shared by the remaining un-failed laminae. For example, we will have

$$dM_{XX} = \int_{-d/2}^{d/2} b d\sigma_{XX} Z dZ = \sum_{\substack{k=1 \\ k \neq k_0}}^{N} b_k \int_{Z_{k-1}}^{Z_k} (d\sigma_{XX})_k Z dZ \tag{8}$$

Thus, the post-failure analysis is still based on Eq. (5), but with reduced overall stiffness elements given by

$$Q_{ij}^I = \sum_{\substack{k=1 \\ k \notin \{k_0\}}}^{N} b_k (C_{ij}^G)_k (Z_k - Z_{k-1}),$$

$$Q_{ij}^{II} = \frac{1}{2} \sum_{\substack{k=1 \\ k \notin \{k_0\}}}^{N} b_k (C_{ij}^G)_k (Z_k^2 - Z_{k-1}^2),$$

$$Q_{ij}^{III} = \frac{1}{3} \sum_{\substack{k=1 \\ k \notin \{k_0\}}}^{N} b_k (C_{ij}^G)_k (Z_k^3 - Z_{k-1}^3) \tag{9}$$

In the above, $\{k_0\}$ represents all those laminae which have already failed. Continued in this way, the ultimate failure strength of the laminate can be determined incrementally. The ultimate tensile strength would correspond to the load level at which all of the layers fail.

It must be realized that the above stiffness reduction process, or, in other words, the incremental solution steps, should be stopped in general before reaching the last ply failure when the laminate is only subjected to a bending load. This is because the ultimate bending strength is attained generally before the last-ply failure [14]. After the ultimate bending load, which occurs at an intermediate ply failure, the predicted load-deflection using the above procedure is incorrect. In fact, under only the bending condition, the middle plane strain increments, $d\varepsilon_{XX}^0$ etc., are negligibly small. The remaining bending curvature will have very small, if any, stress contribution to the last ply failure, according to Eqs. (3) and (4). For instance, if the laminate consists of odd-number (e.g., 5, 7, 9…) of plies each of which has the same global property and the same thickness, the central ply will not carry any load no matter how much a pure bending is applied to the laminate. Thus, the last ply will not fail at all, but the deflection (curvature) can be increased unlimitedly as with the increase of the bending moment. The predicted deflection is, of course, not correct.

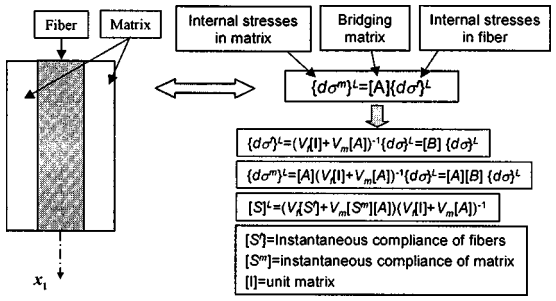


Fig. 3 Schematic of the bridging model for a UD lamina

3 Local Analysis

The local analysis deals with an individual UD (unidirectional) lamina, with three purposes. First, we need to provide the lamina instantaneous compliance matrix as required in Eq. (4.1). Second, the internal stresses in the constituent fiber, and third, in the matrix materials of the lamina must be identified, because without the knowledge of those stresses the lamina instantaneous compliance matrix cannot be defined. Furthermore, having known those internal stresses, the lamina ultimate load carrying ability (failure status) can be assessed by checking the ultimate strengths of the constituent materials. Thus, a pre-fabrication design can be achieved for the composite only based on the information of the constituent properties. All three purposes can be achieved by using the Bridging micromechanics Model, which is briefly summarized below (see Fig. 3). For more details refer to, e.g., Ref. [15].

3.1 Lamina Compliance Matrix. Using the constituent compliance matrices and volume fractions, the instantaneous compliance of the UD lamina in its local coordinate system is given by [15]

$$[S]^{(L)} = (V_f[S^f] + V_m[S^m][A])[B] \tag{10}$$

where the superscript L stands for the local system and V_f and V_m are the volume fractions of the fiber and matrix materials in the composite, respectively. $[S^f]$ and $[S^m]$ are, respectively, the instantaneous compliances of the fiber and matrix materials, whose components are given subsequently. $[A]$ is a bridging matrix and $[B]=(V_f[I]+V_m[A])^{-1}$, whose elements are expressed as [15]

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} \text{ and } [B] = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ 0 & 0 & b_{33} \end{bmatrix} \tag{11}$$

where

$$a_{11} = E_m / E_{11}^f$$

$$a_{22} = 0.5(1 + E_m / E_{22}^f)$$

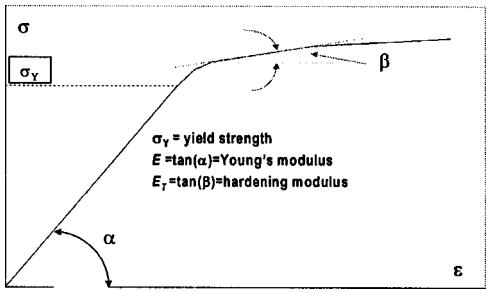


Fig. 4 An elastic–plastic stress–strain curve together with definition of material parameters

$$a_{33} = 0.5(1 + G_m/G_{12}^f)$$

$$E_m = \begin{cases} E^m, \text{ when } \sigma_e^m \leqslant \sigma_Y^m \\ E_T^m, \text{ when } \sigma_e^m > \sigma_Y^m \end{cases}$$

$$G_m = \begin{cases} 0.5E^m/(1+\nu^m), \text{ when } \sigma_e^m \leqslant \sigma_Y^m \\ E_T^m/3, \text{ when } \sigma_e^m > \sigma_Y^m \end{cases}$$

$$\sigma_e^m = \sqrt{(\sigma_{11}^m)^2 + (\sigma_{22}^m)^2 - (\sigma_{11}^m)(\sigma_{22}^m) + 3(\sigma_{12}^m)^2}$$

$$a_{12} = (S_{12}^f - S_{12}^m)(a_{11} - a_{22})/(S_{11}^f - S_{11}^m)$$

$$a_{13} = \frac{d_2\beta_{11} - d_1\beta_{21}}{\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}$$

$$a_{23} = \frac{d_1\beta_{22} - d_2\beta_{12}}{\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}$$

$$b_{11} = (V_f + V_m a_{22})(V_f + V_m a_{33})/c, \quad b_{12} = -(V_m a_{12})(V_f + V_m a_{33})/c$$

$$b_{13} = [(V_m a_{12})(V_m a_{23}) - (V_f + V_m a_{22})(V_m a_{13})]/c,$$

$$b_{22} = (V_f + V_m a_{11})(V_f + V_m a_{33})/c$$

$$b_{23} = -(V_m a_{23})(V_f + V_m a_{11})/c, \quad b_{33} = (V_f + V_m a_{22})(V_f + V_m a_{11})/c$$

$$c = (V_f + V_m a_{11})(V_f + V_m a_{22})(V_f + V_m a_{33})$$

$$d_1 = S_{13}^m(a_{11} - a_{33})$$

$$d_2 = S_{23}^m(V_f + V_m a_{11})(a_{22} - a_{33}) + S_{13}^m(V_f + V_m a_{33})a_{12}$$

$$\beta_{11} = S_{12}^m - S_{12}^f, \quad \beta_{12} = S_{11}^m - S_{11}^f, \quad \beta_{22} = (V_f + V_m a_{22})(S_{12}^m - S_{12}^f)$$

$$\beta_{21} = V_m(S_{12}^f - S_{12}^m)a_{12} - (V_f + V_m a_{11})(S_{22}^f - S_{22}^m)$$

Here E_{11}^f, E_{22}^f, and G_{12}^f are the longitudinal, transverse, and in-plane shear moduli of the fiber; E^m, G^m, and ν^m are Young's modulus, shear modulus, the Poisson's ratio of the matrix; σ_Y^m and E_T^m are the matrix's yield strength and hardening modulus (tangent to its stress–strain curve at the plastic region, see Fig. 4). S_{ij}^f and S_{ij}^m are the instantaneous compliance matrix elements of the fiber and matrix materials, respectively.

3.2 Constituent Internal Stresses. Suppose that arbitrary external stress increments, $\{d\sigma\}=\{d\sigma_{11}, d\sigma_{22}, d\sigma_{12}\}^T$, are applied to the UD lamina (Fig. 2). The internal stress increments generated in the fiber and matrix materials are calculated from [15]

$$\begin{Bmatrix} d\sigma_{11}^f \\ d\sigma_{22}^f \\ d\sigma_{12}^f \end{Bmatrix} = [B] \begin{Bmatrix} d\sigma_{11} \\ d\sigma_{22} \\ d\sigma_{12} \end{Bmatrix} \tag{12.1}$$

$$\begin{Bmatrix} d\sigma_{11}^m \\ d\sigma_{22}^m \\ d\sigma_{12}^m \end{Bmatrix} = [A][B] \begin{Bmatrix} d\sigma_{11} \\ d\sigma_{22} \\ d\sigma_{12} \end{Bmatrix} \tag{12.2}$$

where $[A]$ and $[B]$ are given by Eqs. (11). It is noted that in the present case, the local and the global coordinate systems coincide, i.e., "1" = "X" and "2" = "Y". The stress increments calculated from Eq. (3) can be directly substituted into the right hand sides of Eqs. (12.1) and (12.2).

The total stresses in the fiber and matrix are updated through

$$\{\sigma^f\} = \{\sigma^f\} + \{d\sigma^f\} \quad \text{and} \quad \{\sigma^m\} = \{\sigma^m\} + \{d\sigma^m\} \tag{13}$$

When the composite is fabricated at or near to room temperature, the thermal residual stresses can be neglected. Both of the initial internal stresses in the fiber and the matrix, $\{\sigma^f\}$ and $\{\sigma^m\}$, are set zero. If high thermal residual stresses are involved, the method provided in Ref. [17] can be used to calculate the initial $\{\sigma^f\}$ and $\{\sigma^m\}$.

3.3 Constituent Compliance Matrices. The fiber instantaneous compliance matrix can be defined using Hooke's law, whereas that of the matrix is specified using Prandtl–Reuss theory. Thus, the constituent instantaneous compliance matrices are given as [15]

$$[S_{ij}^f] = \begin{bmatrix} S_{11}^f & S_{12}^f & S_{13}^f \\ & S_{22}^f & S_{23}^f \\ \text{symmetric} & & S_{33}^f \end{bmatrix} = \begin{bmatrix} 1/E_{11}^f & -\upsilon_{12}^f/E_{11}^f & 0 \\ & 1/E_{22}^f & 0 \\ \text{symmetric} & & 1/G_{12}^f \end{bmatrix}$$

$$[S_{ij}^m] = \begin{bmatrix} S_{11}^m & S_{12}^m & S_{13}^m \\ & S_{22}^m & S_{23}^m \\ \text{symmetric} & & S_{33}^m \end{bmatrix} = \begin{cases} [S^m]^e, \text{ when } \sigma_e^m \leqslant \sigma_Y^m \\ [S^m]^e + [S^m]^p, \text{ when } \sigma_e^m > \sigma_Y^m \end{cases}$$

$$[S^m]^e = \begin{bmatrix} \dfrac{1}{E^m} & -\dfrac{\upsilon^m}{E^m} & 0 \\ & \dfrac{1}{E^m} & 0 \\ \text{symmetric} & & \dfrac{1}{G^m} \end{bmatrix}$$

$$[S^m]^p = \frac{9}{4M_T^m(\sigma_e^m)^2} \begin{bmatrix} \sigma'_{11}\sigma'_{11} & \sigma'_{22}\sigma'_{11} & 2\sigma'_{12}\sigma'_{11} \\ & \sigma'_{22}\sigma'_{22} & 2\sigma'_{12}\sigma'_{22} \\ \text{symmetry} & & 4\sigma'_{12}\sigma'_{12} \end{bmatrix}_{\sigma_{ij}=\sigma_{ij}^m}$$

$$M_T^m = \frac{E^m E_T^m}{E^m - E_T^m}$$

$$\sigma'_{ij} = \sigma_{ij} - \frac{1}{3}\sigma_{kk}\delta_{ij}, \quad i,j,k = 1,2.$$

3.4 Failure Criterion. As the composite consists of only two constituent materials, i.e., the fiber and the matrix (Fig. 3), its failure can be considered to occur as long as any of the constituents fails.

To detect the constituent failure, one of the most successful and simplest criteria is the maximum normal stress criterion of homogeneous materials. According to this criterion, the fiber or the matrix failure is attained if either of the following conditions,

$$\frac{\sigma_{11}^f + \sigma_{22}^f}{2} + \frac{1}{2}\sqrt{(\sigma_{11}^f - \sigma_{22}^f)^2 + 4(\sigma_{12}^f)^2} \geqslant \sigma_u^f, \tag{14.1}$$

$$\frac{\sigma_{11}^f + \sigma_{22}^f}{2} - \frac{1}{2}\sqrt{(\sigma_{11}^f - \sigma_{22}^f)^2 + 4(\sigma_{12}^f)^2} \leqslant -\sigma_{u,c}^f, \tag{14.2}$$

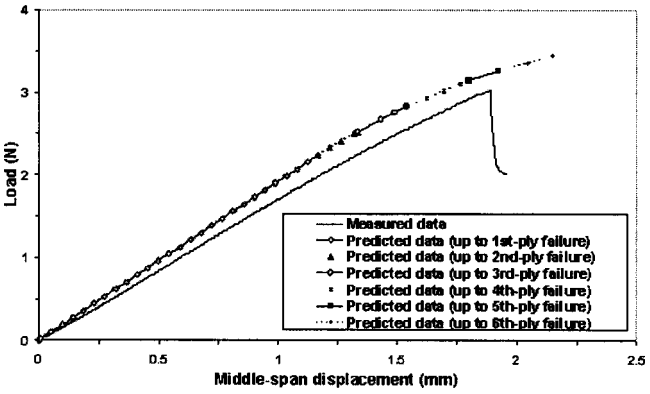


Fig. 5 Measured and predicted load-deflections of a composite cylinder (d=0.5 mm and V_f=0.45)

$$\frac{\sigma_{11}^m + \sigma_{22}^m}{2} + \frac{1}{2}\sqrt{(\sigma_{11}^m - \sigma_{22}^m)^2 + 4(\sigma_{12}^m)^2} \geqslant \sigma_u^m, \qquad (14.3)$$

$$\frac{\sigma_{11}^m + \sigma_{22}^m}{2} - \frac{1}{2}\sqrt{(\sigma_{11}^m - \sigma_{22}^m)^2 + 4(\sigma_{12}^m)^2} \leqslant -\sigma_{u,c}^m, \qquad (14.4)$$

is satisfied, where σ_u^f, $\sigma_{u,c}^f$ and σ_u^m, $\sigma_{u,c}^m$ are the ultimate tensile and compressive strengths of the fiber and the matrix materials, respectively, which can be obtained, e.g., through uniaxial tension and compression tests. It is noted that σ_u^f and $\sigma_{u,c}^f$ are the quantities along the fiber axis direction.

4 Simulation Example

4.1 Experimental work. Experiments have been made to fabricate a UD composite cylinder. The reinforcement was a bundle of five E-glass fiber yarns, each containing 200 fiber filaments (the filament diameter=9 μm, Unitica Glass Fiber, Japan). A mixture consisting of 68 wt % of an epoxy resin, R50, and 32 wt % of a hardener, H64, provided by Chemicrete Enterprise Pte, Ltd (Singapore) was used as the matrix material in the present study. A tube-shrinkage method was applied. Namely, the resin-impregnated fiber bundle was introduced into a polyolefin tube which has an inner diameter larger than that of the fiber bundle. Once heated with an electronic soldering iron, the tube shrank and pushed the extra resin out of the tube ends. The tube was then put into an oven of 100°C for complete curing. After removing the tube carefully, the composite cylinder resulted, with a diameter of about 0.5 mm. The fiber volume content of the cylinder was measured through a combustion method and an averaged value of 45% was obtained.

Three-point bending test was performed for the fabricated cylinder, with a span size of 14 mm following the load condition of an archwire [3]. A total number of seven specimens were tested. Typical load-middle span deflection is plotted in Fig. 5, whereas averaged measured mechanical parameters are summarized in Table 1. Standard deviations are also shown in the table. These results can be used as the benchmark to check the correctness of the developed theory.

4.2 Constituent Properties. In order to apply the theory summarized in the previous sections to simulate the 3-point bending response of the composite cylinder, its constituent fiber and matrix properties must be specified. Both the E-glass fiber and the epoxy matrix used in the present example can be considered as isotropic materials. Further, the fiber is taken as linearly elastic until rupture.

In general, it is difficult to measure directly the mechanical properties of the fiber material. From the material data sheet provided by the supplier, the fiber used has some comparable mechanical parameters to those of the Silenka E-glass 1200tex given in Ref. [18]. Thus, the elastic properties of the fiber were taken from Ref. [18] and are listed in Table 2. These parameters are considered to be the same until rupture at both tension and compression. The fiber tensile and compressive strengths, however, were retrieved from the uniaxial tensile and compressive strengths of a UD composite provided in Ref. [18]. The retrievals have been done in Ref. [19], and are given in Table 2.

In contrast to the fiber properties, the monolithic matrix properties are easily measurable. Pure matrix panels of 6 mm (nominal) thickness were made through resin-casting method and were cured at the 100°C oven as in curing the composite cylinder. The panels were then cut to required specimens according to the relevant ASTM standards using a water-cooled diamond saw. As in the present case the composite cylinder is subjected to a flexural load, the bending behavior of the pure matrix must be understood. Four-point bending tests were carried out, and a typical load-deflection curve of one loading point is plotted in Fig. 6. In order to differentiate the mechanical properties of the pure matrix at tension from those at compression, strain gauges have also been used and the stress–strain data from their measurements are shown in Fig. 7. Unfortunately, the bending strains of the matrix specimen during the test exceeded the measurable range limitation of the strain gauges used and the data shown in Fig. 7 are not complete.

Four linear segments were used to approximate the stress–strain (or load-deflection) responses of the matrix material (Fig. 6 as well as Fig. 7). Based on these representations, the matrix hardening modulus at any load level was found to be

$$E_m = (E_T^m)_i, \quad \text{when } (\sigma_Y^m)_{i-1} \leqslant \sigma_e^m \leqslant (\sigma_Y^m)_i, \quad \text{with } (\sigma_Y^m)_0 = 0$$

$$E_m \equiv (E_T^m)_4, \quad \text{when } \sigma_e^m \geqslant (\sigma_Y^m)_4$$

where the tangential moduli and the critical stresses corresponding to tension and compression are summarized in Table 3. The remaining question is how to define the matrix tensile and compressive strengths under bending. The bending test only determined one of these two strengths. To resolve the problem, uniaxial tensile and compressive tests have also been performed for the pure

Table 1 Measured properties of composite cylinder under 3-point bending test

Materials	Diameter, d (mm)	Fiber volume fraction, V_f	Bending modulus E_b(GPa)	Maximum load(N)
GF/epoxy	0.5	0.45	31.78 (2.1)	3.03 (0.25)*

(.)* = standard deviation

Table 2 Mechanical properties of E-glass fibers [18,19]

E_{11}^f (GPa)	ν_{12}^f	E_{22}^f (GPa)	G_{12}^f (GPa)	ν_{23}^f	σ_u^f (MPa)	$\sigma_{u,c}^f$ (MPa)
74	0.2	74	30.8	0.2	2093	1312

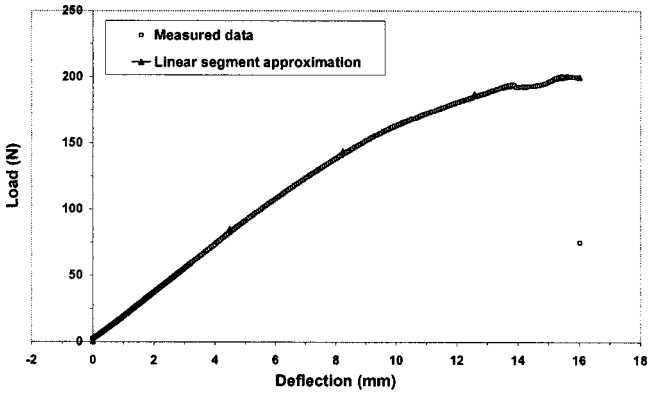


Fig. 6 Load-deflection of R50/H64 pure matrix material under 4-point bending

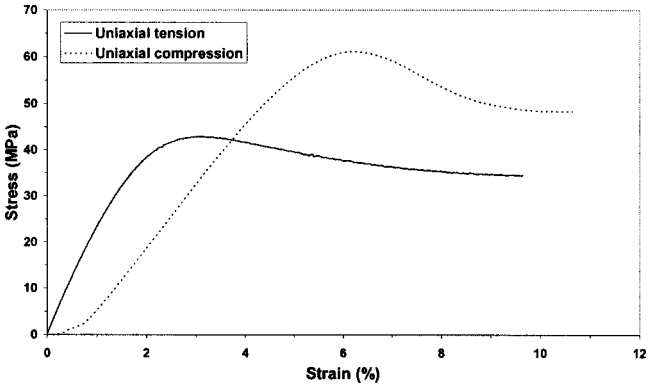


Fig. 8 Stress–strain responses of R50/H64 under uniaxial loads

matrix specimens and typical testing curves are shown in Fig. 8. It was found that the matrix has a uniaxial tensile strength of 42.6 MPa and a uniaxial compressive strength of 63 MPa. In other words, the uniaxial compressive strength of the matrix material is about 20 MPa higher than its uniaxial tensile counterpart. Thus, the bending strength of the matrix, 67.8 MPa, measured from the 4-point bending test must be the matrix tensile strength under bending. The matrix compressive strength under bending was simply set to: 67.8+20=87.8 MPa, as indicated in Table 3. The Poisson's ratio of the matrix, 0.414, was obtained from measured longitudinal and transverse strains under the uniaxial tension.

4.3 Number of Layers in Discretization. A suitable number of lamina layers should be chosen to discretize the cross section of the composite cylinder. This can be done through comparison on the discretization dependent predictions. The predicted load (maximum moment)—deflection curves by using different numbers of discretized layers are plotted in Fig. 9. In the predictions, the cylinder diameter was assumed to be 0.5 mm and a 45% fiber volume fraction was used. From this figure, it is seen that the predictions for the initial parts of the load-deflection curves with differently discretized layers were essentially the same. Discrepancies among them existed only for the later parts of the curves. It seems that the prediction with 16-layers [$N=16$ in Eq. (1)] of discretization was close to those with even more layers of discretizations. Thus, the following results are all based on the 16-layers discretization for the cylinder cross sections.

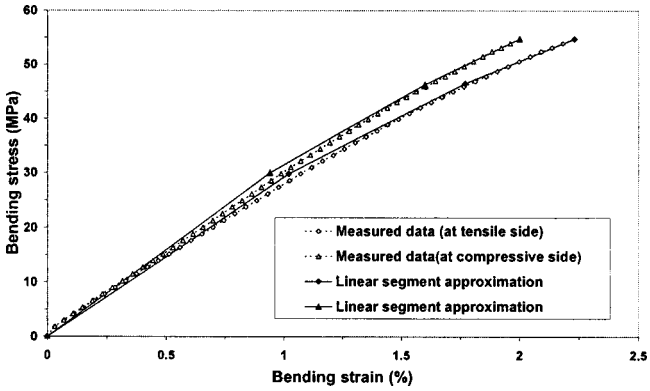


Fig. 7 Stress–strain data of R50/H64 pure matrix under 4-point bending (not complete due to limitation in the strain gauge measurement range)

Some additional words deserve mentioning regarding the predicted load-deflection curves shown in Fig. 9. As mentioned previously, the deflection will be increased unlimitedly before the last-ply failure if Eq. (4.2) is used to calculate the strain increments of each lamina. In order to resolve this problem, the strain increments were calculated in the present paper not necessarily using Eq. (4.2), but according to the following rule.

$$\text{Let } \{d\varepsilon\}_k^{(0)} = \left\{ d\varepsilon_{XX}^0 + \frac{Z_k + Z_{k-1}}{2} d\kappa_{XX}^0, d\varepsilon_{YY}^0 + \frac{Z_k + Z_{k-1}}{2} d\kappa_{YY}^0, 2d\varepsilon_{XY}^0 + (Z_k + Z_{k-1}) d\kappa_{XY}^0 \right\}^T \tag{15.1}$$

$$\{d\varepsilon\}_k^{(1)} = \{d\varepsilon_{XX}^0 + Z_{k-1} d\kappa_{XX}^0, d\varepsilon_{YY}^0 + Z_{k-1} d\kappa_{YY}^0, 2d\varepsilon_{XY}^0 + 0.5 Z_{k-1} d\kappa_{XY}^0\}^T \tag{15.2}$$

$$\text{and } \{d\varepsilon\}_k^{(2)} = \{d\varepsilon_{XX}^0 + Z_k d\kappa_{XX}^0, d\varepsilon_{YY}^0 + Z_k d\kappa_{YY}^0, 2d\varepsilon_{XY}^0 + 0.5 Z_k d\kappa_{XY}^0\}^T \tag{15.3}$$

The absolute values of the sums of the first two linear strain increments from Eqs. (15.1)–(15.3) were calculated. Suppose that amongst the largest absolute value corresponds to i (which is either 0, or 1, or 2). Then, the strain $\{d\varepsilon\}_k^{(i)}$ was set to $\{d\varepsilon\}_k$ and used

Table 3 Mechanical properties of R50 epoxy matrix under bending ($\nu^m=0.414$)

	Tensile properties (strength=67.8 MPa)				Compressive properties (strength=87.8 MPa)			
i	1	2	3	4	1	2	3	4
$(\sigma_Y^m)_i$ (MPa)	28.8	48.9	63.4	67.8	35.8	52.9	68.4	87.8
$(E_T^m)_i$ (GPa)	2.98	2.48	1.45	0.56	3.31	2.73	1.72	0.66

Note: $E_m=(E_T^m)_i$, when $(\sigma_Y^m)_{i-1} \leqslant \sigma_e^m \leqslant (\sigma_Y^m)_i$, with $(\sigma_Y^m)_0=0$, $E_m \equiv (E_T^m)_4$, when $\sigma_e^m \geqslant (\sigma_Y^m)_4$

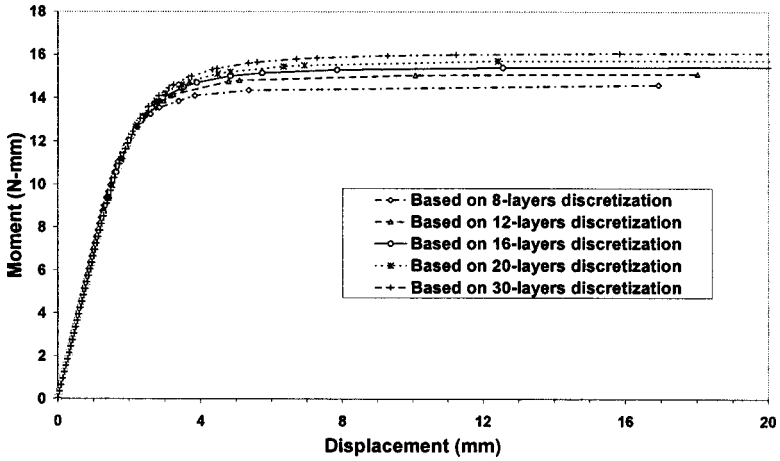


Fig. 9 Comparison between the predicted results of different lamina layers used in discretizing the cross-section of the composite cylinder (d=0.5 mm & V_f=0.45)

in Eq. (3). For instance, if $|d\varepsilon^0_{XX}+0.5(Z_k+Z_{k-1})(d\kappa^0_{XX}+d\kappa^0_{YY})+d\varepsilon^0_{YY}|$ is greater than both $|d\varepsilon^0_{XX}+Z_{k-1}(d\kappa^0_{XX}+d\kappa^0_{YY})+d\varepsilon^0_{YY}|$ and $|d\varepsilon^0_{XX}+Z_k(d\kappa^0_{XX}+d\kappa^0_{YY})+d\varepsilon^0_{YY}|$, then $\{d\varepsilon\}_k=\{d\varepsilon\}^{(0)}_k$. The load-deflection curves thus obtained were slightly different from those throughout using Eq. (4.2) only before the last few plies failed. However, the predicted deflection in such way was finite even at the last-ply failure.

4.4 Simulation Results. Using the material parameters given in Tables 2 and 3 as input data, the load-deflection curve of the composite cylinder, with a diameter of 0.5 mm and a 45% fiber volume fraction, was estimated up to the sixth-ply failure, and is graphed in Fig. 5 for comparison. It is noted that the cylinder has been descritized into 16 lamina layers of equal thickness. From the figure, we can clearly see that the failure load corresponding to the fifth-ply failure should be regarded as the maximum load sustainable by the composite cylinder under consideration. This is because the predicted deflection at the fifth-ply failure is in the nearest close to but greater than the critical deflection, which occurred when the cylinder was three-point bending loaded to the maximum, on the measured curve. As mentioned previously, the predicted even larger deflection after the fifth-ply failure would be incorrect, and is in fact not necessary as we are only interested in the cylinder structure-property behavior until its ultimate failure. After the ultimate failure, the load-carrying ability of the cylinder lowers down quickly as shown in Fig. 5 by the measured curve although the cylinder may have not completely fractured (been broken into segments). The figure also shows that the theoretical modeling by using independent constituent mechanical properties is reasonably accurate. Some discrepancy between the predicted and the measured load-deflection curves may be attributed to the use of inaccurate fiber modulus. It is seen that the predicted curve was slightly stiffer than the measured one. In the present paper, the glass fiber modulus was directly taken from Ref. [18], which may be somewhat different from that of the fibers used in situ. Another source of the error may come from the neglection of the thermal residual stresses involved. The present composite was fabricated at an elevated temperature (100°C), whereas the measurement was performed at room temperature (25°C), although the moderate temperature variation from the fabrication to the room temperatures does not bring significant influence on the predicted results [20], and indeed has not.

Further study was carried out with respect to some different geometric parameters. As the same constituent materials and the same number of discretization layers are used, it is reasonable to assume that the ultimate bending failure of a UD composite cyl-

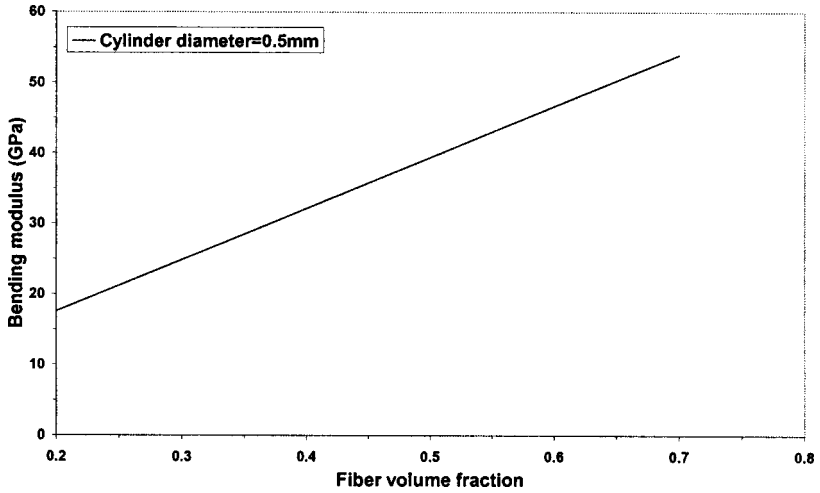


Fig. 10 Bending modulus of GF/R50 composite cylinder versus fiber volume fraction

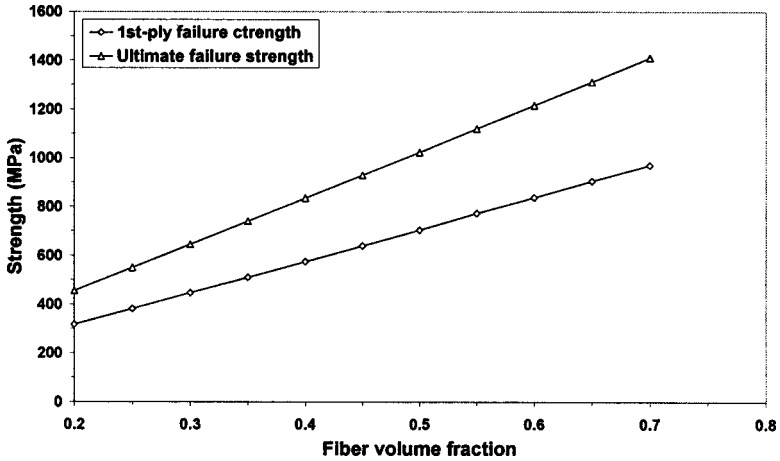


Fig. 11 Strength of GF/R50 composite cylinder versus fiber volume fraction under bending

inder occurs at its fifth-ply failure even if a different fiber volume fraction is assumed. The predicted bending moduli and the ultimate bending strengths of circular composite cylinders made of GF/R50 material system varied with fiber volume fractions are plotted in Figs. 10 and 11, respectively. The first-ply failure strengths of the cylinders are also shown in Fig. (11). From these figures, it is seen that the bending stiffness and strength are essentially linearly dependent on the fiber volume fraction of the cylinder.

4.5 Addition Remark. In comparison with an in-plane load situation where a stress failure criterion is generally sufficient, the ultimate bending failure analysis of a composite requires not only the stress failure criterion but also an extra controlling parameter, i.e., a critical deflection related condition. This extra condition must be obtained based on a bending measurement. The deflection corresponding to which the maximum bending load occurs on the measured load-deflection curve is defined as the critical deflection. Then, the critical failed layer code (which was the fifth failed ply in the present paper) or the critical curvature (not dealt with in the present paper) can be retrieved using the analyzing scheme (i.e., the bridging model combined with the classical lamination theory) provided in the paper. Thus, the critical failed layer code or the critical curvature condition can be used for design purpose. It should be noted that compared with the tested sample the composite to be designed should have the same constituent materials, the same fiber arrangement pattern (e.g., uniaxial arrangement in the present case), and a similar structure (e.g., a circular cross section in the present case), but different fiber content as well as some different dimension (e.g., different cylinder diameter). With these conditions, the new composite (cylinder) will most probably have the same critical failed layer code or the same critical curvature.

5 Conclusion

Theoretical modeling for the flexural response of composite cylinders up to their ultimate failure has been described in the present paper. By discretizing the cylinder into a number of parallel laminae, the analysis of the cylinder is transformed into that of a laminated beam. The bridging model combined with the classical lamination theory is then used to estimate the flexural progressive failure process of the laminated beam. However, the determination of the ultimate bending strength of the cylinder requires an addition critical deflection (curvature) condition. By using both the stress failure and the critical deflection conditions and by using independent fiber and matrix properties, the predicted ultimate bending strength of a cylinder agreed well with the measurement. It has been found that the bending stiffness and the ultimate bending strength of the composite cylinder is essentially proportional to its fiber volume fraction.

References

[1] Ramakrishna, S., Mayer, J., Wintermantel, E., and Leong, K. W., 2001, "Biomedical Applications of Polymer-Composite Materials: A Review," Compos. Sci. Technol., **61**, pp. 1189–1224.
[2] Watari, F., Yamagata, S., Imai, T., and Nakamura, S., 1998, "The Fabrication and Properties of Aesthetic FRP Wires for use in Orthodontics," J. Mater. Sci., **33**, pp. 5661–5664.
[3] Toyoizumi, H., Watari, F., Imai, T., Yamagata, S., and Kobayashi, M., 1999, "Fabrication of Aesthetic Wires with Flexural and Torsional Stiffness by Photo Curing Method," The Journal of the Japanese Society for Dental Materials and Devices, **18**, pp. 429–440.
[4] Fallis, D. W., and Kusy, R. P., 2000, "Variation in Flexural Properties of Photo-pultruded Composite Archwires: Analyses of Round and Rectangular Profiles," J. Mater. Sci.: Mater. Med., **11**, pp. 683–693.
[5] Bhattacharyya, A., and Appiah, E. J., 2000, "On the Exact Solution of Elastoplastic Response of an Infinitely Long Composite Cylinder during Cyclic Radial Loading," J. Mech. Phys. Solids, **48**, pp. 1065–1092.
[6] Starbuck, J. M., 1999, "Stress Analysis of Laminated Composite Cylinders under NonAxisymmetric Loading," *International SAMPE technical Conference*, 31, pp. 604–615.
[7] Whitney, J. M., Browning, C. E., and Mair, A., 1974, "Analysis of the Flexure Test for Laminated Composite Materials," *Composite Materials: Testing and Design, ASTM STP 5 46*, American Society for Testing and Materials, pp. 30–45.
[8] Turvey, G. J., 1982, "Uniformly Loaded, Antisymmetric Cross-Ply Laminated, Rectangular Plates: An Initial Flextural Failure Analysis," Fibre Sci. Technol., **16**, pp. 1–10.
[9] Dvorak, G. J., and Laws, N., 1987, "Analysis of Progressive Matrix Cracking in Composite Laminates II: First Ply Failure," J. Compos. Mater., **21**, pp. 309–329.
[10] Grief, R., and Chapon, E., 1993, "Investigation of Successive Failure modes in Graphite/Epoxy Laminated Composite Beams," J. Reinf. Plast. Compos., **12**, pp. 602–621.
[11] Kam, T. Y., and Sher, H. F., 1995, "Nonlinear and First-Ply Failure Analyses of Laminated Composite Cross-Ply Plates," J. Compos. Mater., **29**, pp. 463–482.
[12] Echaabi, J., Trochu, F., Pham, X. T., and Ouellet, M., 1996, "Theoretical and Experimental Investigation of Failure and Damage Progression of Graphite-Epoxy Composites in Flexural Bending Test," J. Reinf. Plast. Compos., **15**, pp. 740–755.
[13] Smith, P. A., and Ogin, S. L., 1999, "On Transverse Matrix Cracking in Cross-Ply Laminates Loaded in Simple Bending," Composites, Part A, **30**, pp. 1003–1008.
[14] Huang, Z. M., Teng, X. C., and Ramakrishna, S., 2001, "Bending Behavior of Laminated Knitted Fabric Reinforced Beams," Advanced Composites Lett., **10**, pp. 211–218.
[15] Huang, Z. M., 2001, "Simulation of the Mechanical Properties of Fibrous Composites by the Bridging Micromechanics Model," Composites, Part A, **32**, pp. 143–172.
[16] Huang, Z. M., 1999, "Micromechanical Strength Formulae of Unidirectional Composites," Mater. Lett., **40**, pp. 164–169.

[17] Huang, Z. M., 2001, "Modeling Strength of Multidirectional Laminates under Thermo-Mechanical Loads," J. Compos. Mater., **35**, pp. 281–315.
[18] Soden, P. D., Hinton, M. J., and Kaddour, A. S., 1998, "Lamina Properties, Lay-up Configurations and Loading Conditions for a Range of Fiber-Reinforced Composite Laminates," Compos. Sci. Technol., **58**, pp. 1011–1022.
[19] Huang, Z. M., 2004, "A Bridging Model Prediction of the Ultimate Strength of Composite Laminates Subjected to Biaxial Loads," Compos. Sci. Technol., **64**, pp. 395–448.
[20] Huang, Z. M., 2004, "Correlation of the Bridging Model Predictions of the Biaxial Failure Strengths of Fibrous Laminates With Experiments," Compos. Sci. Technol., **64**, pp. 529–548.

Q. Yang[1]
Department of Hydraulic Engineering,
Tsinghua University,
Beijing 100084, P. R. China.
e-mail: yangq@tsinghua.edu.cn

L. G. Tham
Rick Engineering Research Center,
The University of Hong Kong,
Hong Kong, China

G. Swoboda
Faculty of Civil Engineering and Architecture,
The University of Innsbruck,
Innsbruck, Austria

Normality Structures With Homogeneous Kinetic Rate Laws

In this paper, a homogeneous type of kinetic rate laws of local internal variables and its corresponding macroscopic behaviors, are explored within the framework of "normality structures" by Rice. Rice's kinetic rate laws of local internal variables, with each rate being stress dependent only via its conjugate thermodynamic force, are corner stones of the normality structure. It is revealed in this paper that nonlinear phenomenological equations and Onsager reciprocal relations emerge naturally if each rate is a homogeneous function of degree q in its conjugate force. Furthermore, the nonlinear phenomenological coefficient matrix is identical to the Hessian matrix of the flow potential function in conjugate forces only scaled by q. It is further shown that the refined version of Griffith criterion proposed by Rice, $(G-2\gamma)\dot{a}\geqslant 0$, can be derived from the normality structure with the homogeneous rate laws. Finally, some issues related to damage evolution laws have been discussed based on the remarkable properties.
[DOI: 10.1115/1.1867991]

1 Introduction

As an internal variable approach, the normality structure proposed by Rice [1,2] has been an appealing constitutive framework for solids undergoing irreversible thermodynamic processes. Rice's kinetic rate laws of local internal variables, with each rate being stress dependent only via its conjugate thermodynamic force, are corner stones of the normality structure and represent a wide class of inelastic behaviors.

In this paper, we are interested in a special class of Rice's kinetic rate laws and its remarkable properties. It is revealed in Sec. 3 that nonlinear phenomenological equations and Onsager reciprocal relations emerge naturally from the normality structure if each rate is a homogeneous function of degree q in its conjugate force. Furthermore, the phenomenological coefficient matrix is identical to the Hessian matrix of the flow potential function in conjugate forces only scaled by q, and the homogeneous property transfers exactly from local internal variables to global average internal variables.

Rather than the more usually cited condition that $G=2\gamma$ for the onset of crack extension, Rice [4] proposed the restriction on quasi-static extension or healing of Griffith cracks, $(G-2\gamma)\dot{a}\geqslant 0$, at any local crack front, where G is the Irwin energy release rate and γ is the surface free energy. Although inspired by the requirement of the second law of thermodynamics, the restriction is not an essential thermodynamic requirement. It is shown in Sec. 5 that the restriction becomes essential within the normality structure with the homogeneous kinetic rate laws. Furthermore, it is indicated that the widely used power laws for cracking are just the simplest forms of homogeneous kinetic rate laws.

Linear irreversible thermodynamics or phenomenological equations along with Onsager [3] reciprocal relations where thermodynamic fluxes and forces are assumed to be linear dependent and related by a symmetric phenomenological coefficient matrix, have provided access to both the understanding and the analysis of a wide range of physical phenomena see, e.g., De Groot and Mazur [5]. The linear theory is restricted to linear thermodynamic processes or else the derivation from equilibrium have to be sufficient small that a linear approximation is valid, and can not be applied to phenomena which represent large nonlinear deviation from equilibrium states. The nonlinear generalization of the linear theory includes the normality structure of Rice [1,2] briefed in Sec. 2, the maximum dissipation rate of Ziegler [6] and the Onsager fluxes of Edelen [7,8] briefed in Sec. 6. These generalizations all lead to certain normality structures. Especially, significant progress has to be achieved following the line of Ziegler [6]; see, e.g., Ziegler and Wehrli [9], Rajagopal and Srinivasa [10–12], Puzrin and Houlsby [13].

In Sec. 6, it is revealed that the homogenous property of Rice's kinetic rate laws is consistent with the principle of maximum dissipation rate of Ziegler [6], and the kinetic rate laws of Rice [1,2] is just certain specific Onsager fluxes of Edelen [7,8].

In continuum damage mechanics, damage evolution laws have been the most elusive parts owing to their complex tensorial and high-degree nonlinear properties see, e.g., Krajcinovic [14] and Lemaitre et al. [15]. In fact, it is the main drive force behind this research. In Sec. 7, the revealed remarkable properties help us gain a deep insight into the structures of anisotropic damage evolution laws.

2 Normality Structure

Consider a material sample of size V. Introduce the specific free energy ϕ and its Legendre transform ψ with respect to strain

$$\phi=\phi(\boldsymbol{\varepsilon},\vartheta,H),\quad \psi=\psi(\boldsymbol{\sigma},\vartheta,H)=\boldsymbol{\varepsilon}:\frac{\partial\phi}{\partial\varepsilon}-\phi \tag{1}$$

where ϑ denotes temperature; ε denotes any strain tensor, objective and symmetric, that measures deformation from an arbitrary reference state; $\boldsymbol{\sigma}$ denotes the symmetric conjugate stress such that $\boldsymbol{\sigma}:\mathrm{d}\boldsymbol{\varepsilon}$ is the work per unit volume of the adopted reference state in any virtual deformation $\mathrm{d}\boldsymbol{\varepsilon}$; H denotes symbolically the current pattern of microstructural rearrangement of constituent elements of the materials. At fixed H, variations of $\boldsymbol{\sigma}$ and ϑ necessarily induce a purely elastic response. Then the first law of thermodynamics leads to the stress–strain relations,

[1]Author to whom correspondence should be addressed. Telephone: 8610-62794874; Fax: 8610-62782159.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, August 13, 2002; final revision, July 21, 2004. Associate Editor: K. R. Rajagopal. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California, Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

$$\boldsymbol{\sigma} = \frac{\partial \phi(\boldsymbol{\varepsilon}, \vartheta, H)}{\partial \boldsymbol{\varepsilon}}, \quad \varepsilon = \frac{\partial \psi(\boldsymbol{\sigma}, \vartheta, H)}{\partial \boldsymbol{\sigma}}. \tag{2}$$

Consider two neighboring patterns of microstructural rearrangement denoted by H, $H+\mathrm{d}H$. It is assumed that a set of incremental scalar internal variables $\mathrm{d}\xi_1, \mathrm{d}\xi_2 \ldots, \mathrm{d}\xi_n$ characterize the specific local rearrangements, which are represented collectively by $\mathrm{d}H$, at sites throughout the material sample. The $\mathrm{d}\xi$'s and $\mathrm{d}H$ are related by

$$\frac{1}{V} f_\alpha \, \mathrm{d}\xi_\alpha = -\, \mathrm{d}^p \phi = \mathrm{d}^p \psi \tag{3}$$

where[1]

$$\mathrm{d}^p \phi = \phi(\boldsymbol{\varepsilon}, \vartheta, H + \mathrm{d}H) - \phi(\boldsymbol{\varepsilon}, \vartheta, H)$$

$$\mathrm{d}^p \psi = \psi(\boldsymbol{\sigma}, \vartheta, H + \mathrm{d}H) - \psi(\boldsymbol{\sigma}, \vartheta, H) \tag{4}$$

Equation (3) also defines the scalar thermodynamic forces $f_1, f_2, \ldots, f_n$ (collectively $\mathbf{f}$) conjugate to the variables,

$$\mathbf{f} = \mathbf{f}(\boldsymbol{\sigma}, \vartheta, H) \quad \text{or} \quad \mathbf{f} = \mathbf{f}(\boldsymbol{\varepsilon}, \vartheta, H) \tag{5}$$

The corresponding set of total internal variables,

$$\boldsymbol{\xi} = \{\xi_1, \xi_2, \ldots, \xi_n\} \tag{6}$$

generally are not state variables in the sense that thermodynamic state functions are not direct functions of $\boldsymbol{\xi}$, but instead depend on the path history of $\boldsymbol{\xi}$. Only if the $\boldsymbol{\xi}$ is one set of explicit state variables, the conjugate forces can be determined as

$$f_\alpha = V \frac{\partial \psi}{\partial \xi_\alpha} = -\, V \frac{\partial \phi}{\partial \xi_\alpha}, \quad \phi = \phi(\boldsymbol{\varepsilon}, \vartheta, \boldsymbol{\xi}), \quad \psi = \psi(\boldsymbol{\sigma}, \vartheta, \boldsymbol{\xi}) \tag{7}$$

Following the second law of thermodynamics, the entropy production rate should be always non-negative,

$$\sigma = \frac{1}{\vartheta V} f_\alpha \dot{\xi}_\alpha \geqslant 0 \tag{8}$$

In the normality structure, a key assumption is that the kinetic rate laws of the internal variables take the form

$$\dot{\xi}_\alpha = \dot{\xi}_\alpha(f_\alpha, \vartheta, H), \quad (\alpha = 1, 2, \ldots, n) \tag{9}$$

Therefore, the kinetic rate laws can be related to a flow potential Q and be recast as

$$\dot{\xi}_\alpha = V \frac{\partial Q}{\partial f_\alpha}, \quad Q = Q(\mathbf{f}, \vartheta, H) = \frac{1}{V} \int_0^{\mathbf{f}} \dot{\xi}_\alpha(f_\alpha, \vartheta, H) \mathrm{d}f_\alpha \tag{10}$$

where the integration is carried out at fixed ϑ and H, and defines a direct function Q of $\mathbf{f}$ since each term in the integrand is a total differential. The inelastic part of a strain increment is, due to Eqs. (2)–(4),

$$\mathrm{d}^p \boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\boldsymbol{\sigma}, \vartheta, H + \mathrm{d}H) - \boldsymbol{\varepsilon}(\boldsymbol{\sigma}, \vartheta, H) = \frac{\partial (\mathrm{d}^p \psi)}{\partial \boldsymbol{\sigma}} = \frac{1}{V} \frac{\partial f_\alpha}{\partial \sigma} \mathrm{d}\xi_\alpha \tag{11}$$

Therefore, the following normality structure holds, noting $\mathbf{f} = \mathbf{f}(\boldsymbol{\sigma}, \vartheta, H)$,

$$\frac{\mathrm{d}^p \boldsymbol{\varepsilon}}{\mathrm{d}t} = \frac{\partial Q}{\partial \boldsymbol{\sigma}}, \quad Q = Q(\boldsymbol{\sigma}, \vartheta, H) = Q(\mathbf{f}, \vartheta, H) \tag{12}$$

where t denotes time, since, due to Eq. (10)

$$\frac{\partial Q}{\partial \boldsymbol{\sigma}} = \frac{1}{V} \frac{\partial f_\alpha}{\partial \boldsymbol{\sigma}} \dot{\xi}_\alpha. \tag{13}$$

[1]In this paper, Einstein's summation convention is adopted for repeated indexes. However, if an index range is listed like α in Eq. (9), the index is considered as a free index without the summation convention.

2.1 Introduction of Averaging Internal Variables $\boldsymbol{\zeta}$. The set $\boldsymbol{\xi}$ generally contains numerous elements. One set of much reduced internal variables $\boldsymbol{\zeta}$ can be introduced as the average measurements of $\boldsymbol{\xi}$

$$\boldsymbol{\zeta} = \{\zeta_1, \zeta_2, \ldots, \zeta_m\}, \; \zeta_\mu = \zeta_\mu(\xi_1, \xi_2, \ldots, \xi_n; V) \; (\mu = 1, 2, \ldots, m \ll n) \tag{14}$$

where V indicates averaging over the volume. The thermodynamic forces acting on the averaging variables $\boldsymbol{\zeta}$ are $g_1, g_2, \ldots, g_m$ (collectively $\mathbf{g}$). Only if the $\boldsymbol{\zeta}$ is one set of explicit state variables, the conjugate forces can be determined as

$$g_\mu = \frac{\partial \psi}{\partial \zeta_\mu} = -\, \frac{\partial \phi}{\partial \zeta_\mu}, \quad \phi = \phi(\boldsymbol{\varepsilon}, \vartheta, \boldsymbol{\zeta}), \quad \psi = \psi(\boldsymbol{\sigma}, \vartheta, \boldsymbol{\zeta}). \tag{15}$$

The equivalence that the averaging variables $\boldsymbol{\zeta}$ can describe the thermodynamic system characterized by $\boldsymbol{\xi}$, is achieved by requiring the equality for all $\delta\boldsymbol{\xi}$

$$g_\mu \delta \zeta_\mu = \frac{1}{V} f_\alpha \delta \xi_\alpha, \tag{16}$$

which just the requirement of equal entropy production rate at micro- and macro-levels. Due to Eq. (14), one obtains

$$\delta \zeta_\mu = \frac{\partial \zeta_\mu}{\partial \xi_\alpha} \delta \xi_\alpha, \quad \zeta_\mu = \zeta_\mu(\boldsymbol{\xi}, V). \tag{17}$$

Substituting Eq. (17) into Eq. (16) leads to

$$f_\alpha = V g_\mu \frac{\partial \zeta_\mu(\boldsymbol{\xi}, V)}{\partial \xi_\alpha} = f_\alpha(\mathbf{g}, \vartheta, H) \Rightarrow \frac{\partial \zeta_\mu(\boldsymbol{\xi}, V)}{\partial \xi_\alpha} = \frac{1}{V} \frac{\partial f_\alpha(\mathbf{g}, \vartheta, H)}{\partial g_\mu} \tag{18}$$

Therefore,

$$\dot{\zeta}_\mu = \frac{\partial \zeta_\mu}{\partial \xi_\alpha} \dot{\xi}_\alpha = \frac{1}{V} \dot{\xi}_\alpha \frac{\partial f_\alpha(\mathbf{g}, \vartheta, H)}{\partial g_\mu} = \frac{\partial Q(\mathbf{g}, \vartheta, H)}{\partial g_\mu} \tag{19}$$

where

$$Q(\mathbf{g}, \vartheta, H) = \frac{1}{V} \int_0^{\mathbf{f}(\mathbf{g}, \vartheta, H)} \dot{\xi}_\alpha(f_\alpha, \vartheta, H) \mathrm{d}f_\alpha = Q(\mathbf{f}, \vartheta, H) \tag{20}$$

2.2 Incremental Dependence of $\boldsymbol{\zeta}$ on $\boldsymbol{\xi}$. Direct relations do not always exist between $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ like in Eq. (14). However, since the set of incremental internal variables $\mathrm{d}\boldsymbol{\xi}$ determines fully the internal rearrangement $\mathrm{d}H$, a proper set of incremental averaging internal variables $\mathrm{d}\boldsymbol{\zeta}$ can also describe $\mathrm{d}H$ with sufficient accuracy. Thus, it is reasonable to assume such a linear dependence between $\mathrm{d}\boldsymbol{\zeta}$ and $\mathrm{d}\boldsymbol{\xi}$,

$$\mathrm{d}\boldsymbol{\zeta} = \mathbf{R} \cdot \mathrm{d}\boldsymbol{\xi} \quad \text{or} \quad \mathrm{d}\zeta_\mu = R_{\mu\alpha} \mathrm{d}\xi_\alpha \tag{21}$$

which implies that the set $\boldsymbol{\zeta}$ depends not only on the set $\boldsymbol{\xi}$ but also its path history. Here it is only assumed that the $\mathbf{R}$ exists uniquely for a given internal rearrangement H. Evidently, the direct relations Eq. (14) can also be written as Eq. (21) with

$$R_{\mu\alpha} = \frac{\partial \zeta_\mu}{\partial \xi_\alpha} \tag{22}$$

No matter what the relation between $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ are, the dissipation relation Eq. (16) should always hold. Therefore, the following relations are obtained, similar to Eq. (18)

$$f_\alpha = V g_\mu R_{\mu\alpha} \Rightarrow R_{\mu\alpha} = \frac{1}{V} \frac{\partial f_\alpha}{\partial g_\mu} \tag{23}$$

Therefore, with the potential function $Q(\mathbf{g}, \vartheta, H)$ defined in Eq. (20), the normality condition similar to Eq. (19) still holds,

$$\dot{\zeta}_\mu = R_{\mu\alpha}\dot{\xi}_\alpha = \frac{1}{V}\dot{\xi}_\alpha \frac{\partial f_\alpha(\mathbf{g},H)}{\partial g_\mu} = \frac{\partial Q(\mathbf{g},\vartheta,H)}{\partial g_\mu} \tag{24}$$

In the following discussion, the incremental relations like Eq. (21) are generally assumed between $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$, and the direct relations like Eq. (14) are just considered as a special case of the incremental relations with Eq. (22).

3 Normality Structures With Homogeneous Kinetic Rate Laws

Let us define the dissipation functions at both the microscopic and macroscopic levels,

$$\Phi(\mathbf{f},\vartheta,H) = \frac{1}{V} f_\alpha \dot{\xi}_\alpha, \quad \Phi(\mathbf{g},\vartheta,H) = g_\mu \dot{\zeta}_\mu \tag{25}$$

In fact, Φ/ϑ is just the entropy production rate. The introduction of the Rice [1,2] kinetic rate laws implies that the dissipation function $\Phi(\mathbf{f},\vartheta,H)$ is well-defined. On the other hand, Eq. (16) is just the requirement that the microscopic dissipation function should be equal to the macroscopic one, i.e.,

$$\Phi(\mathbf{f},\vartheta,H) = \Phi(\mathbf{g},\vartheta,H) \tag{26}$$

Due to Eqs. (10) and (19), the dissipation and flow potential functions are related by

$$\Phi(\mathbf{f},\vartheta,H) = f_\alpha \frac{\partial Q(\mathbf{f},\vartheta,H)}{\partial f_\alpha}, \quad \Phi(\mathbf{g},\vartheta,H) = g_\mu \frac{\partial Q(\mathbf{g},\vartheta,H)}{\partial g_\mu} \tag{27}$$

In view of Eqs. (3), (11), and (13), we have the following relations,

$$\Phi = -\frac{\mathrm{d}^p \phi}{\mathrm{d}t} = \frac{\mathrm{d}^p \psi}{\mathrm{d}t} \tag{28}$$

The Rice [1,2] kinetic rate laws of the internal variables, Eq. (9), with each rate being stress dependent only via its conjugate thermodynamic force, are corner stones of the normality structure. However, they should be thought of only as an approximation and not as a physical law, as remarked by Rice [2]. In this paper, we are interested in a special type of kinetic rate laws that each rate $\dot{\xi}_\alpha$ is a homogeneous function of degree q in its conjugate force f_α,

$$\frac{\partial \dot{\xi}_\alpha(f_\alpha,\vartheta,H)}{\partial f_\alpha} f_\alpha = q\dot{\xi}_\alpha(f_\alpha,\vartheta,H) \quad (\alpha = 1,2,\ldots,n) \tag{29}$$

where Euler's Theorem on homogeneous functions is adopted as a definition. It is emphasized that all kinetic rate laws possess the same homogeneous degree q. With the homogeneous property Eq. (29) and integration by parts, the flow potential Q defined in Eq. (10) can be recast as

$$\begin{aligned} Q(\mathbf{f},\vartheta,H) &= \frac{1}{V}\int_0^{\mathbf{f}} \dot{\xi}_\alpha \,\mathrm{d}f_\alpha = \frac{1}{V}\dot{\xi}_\alpha f_\alpha - \frac{1}{V}\int_0^{\mathbf{f}} q\dot{\xi}_\alpha \,\mathrm{d}f_\alpha \\ &= \Phi(\mathbf{f},\vartheta,H) - qQ(\mathbf{f},\vartheta,H) \end{aligned} \tag{30}$$

which leads to

$$\Phi(\mathbf{f},\vartheta,H) = (q+1)Q(\mathbf{f},\vartheta,H) \Rightarrow \Phi(\mathbf{g},\vartheta,H) = (q+1)Q(\mathbf{g},\vartheta,H), \tag{31}$$

due to Eqs. (20) and (26). This equation along with Eq. (27) indicate that both $\Phi(\mathbf{f},\vartheta,H)$, $Q(\mathbf{f},\vartheta,H)$ and $\Phi(\mathbf{g},\vartheta,H)$, $Q(\mathbf{g},\vartheta,H)$ are homogeneous functions of degree $q+1$ in $\mathbf{f}$ and $\mathbf{g}$, respectively,

$$\frac{\partial \Phi}{\partial f_\alpha} f_\alpha = \frac{\partial \Phi}{\partial g_\mu} g_\mu = (q+1)\Phi, \quad \frac{\partial Q}{\partial f_\alpha} f_\alpha = \frac{\partial Q}{\partial g_\mu} g_\mu = (q+1)Q. \tag{32}$$

3.1 Nonlinear Phenomenological Equations. Differentiating Eq. (27) by g_μ, it follows that

$$\frac{\partial \Phi(\mathbf{g},\vartheta,H)}{\partial g_\mu} = \frac{\partial Q(\mathbf{g},\vartheta,H)}{\partial g_\mu} + g_\kappa \frac{\partial^2 Q(\mathbf{g},\vartheta,H)}{\partial g_\kappa \partial g_\mu} \tag{33}$$

with the summation convention for κ ($\kappa = 1,2,\ldots,m$). Using Eqs. (19) and (31), one obtains

$$\dot{\zeta}_\mu = J_{\mu\kappa} g_\kappa, \quad J_{\mu\kappa} = \frac{1}{q}\frac{\partial^2 Q(\mathbf{g},\vartheta,H)}{\partial g_\mu \partial g_\kappa} = J_{\kappa\mu} \tag{34}$$

which are exactly the phenomenological equations and Onsager reciprocal relations. The phenomenological equations can be written in matrix form

$$\dot{\zeta} = Jg, \quad \dot{\zeta} = \{\dot{\zeta}_1, \dot{\zeta}_2, \ldots, \dot{\zeta}_m\}^T, \quad g = \{g_1, g_2, \ldots, g_m\}^T \tag{35}$$

where the nonlinear phenomenological coefficient matrix J is a $m \times m$ square matrix, and its element at μth row and κth column is $J_{\mu\kappa}$. Note that the Hessian matrix of the flow potential Q in g is denoted by $H(Q,g)$ and defined as

$$H(Q,g) = \begin{bmatrix} \dfrac{\partial^2 Q}{\partial g_1^2} & \dfrac{\partial^2 Q}{\partial g_1 \partial g_2} & \cdots & \dfrac{\partial^2 Q}{\partial g_1 \partial g_m} \\ \dfrac{\partial^2 Q}{\partial g_2 \partial g_1} & \dfrac{\partial^2 Q}{\partial g_2^2} & \cdots & \dfrac{\partial^2 Q}{\partial g_2 \partial g_m} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 Q}{\partial g_m \partial g_1} & \dfrac{\partial^2 Q}{\partial g_m \partial g_2} & \cdots & \dfrac{\partial^2 Q}{\partial g_m^2} \end{bmatrix}. \tag{36}$$

Evidently, the matrix J is identical to the Hessian matrix scaled by $1/q$, i.e.,

$$J = \frac{1}{q} H(Q,g) \text{ or } J = \frac{1}{q(q+1)} H(\Phi,g) \tag{37}$$

due to Eq. (31). Since Hessian matrices are always symmetric, the Onsager reciprocal relations are incorporated implicitly.

Similarly, Eq. (29) directly leads to the phenomenological equations at the microscopic level,

$$\dot{\xi}_\alpha = J_\alpha f_\alpha, \quad J_\alpha = \frac{1}{q}\frac{\partial \dot{\xi}_\alpha}{\partial f_\alpha} = \frac{1}{q}\frac{\partial^2 Q}{\partial f_\alpha^2}, \quad (\alpha = 1,2,\ldots,n) \tag{38}$$

or in matrix form

$$\dot{\xi} = \tilde{J} f, \quad \dot{\xi} = \{\dot{\xi}_1, \dot{\xi}_2, \ldots, \dot{\xi}_n\}^T, \quad f = \{f_1, f_2, \ldots, f_n\}^T \tag{39}$$

where the nonlinear phenomenological coefficient matrix $\tilde{J}$ is a $n \times n$ square matrix and associated with the Hessian matrix of Q or Φ by

$$\tilde{J} = \frac{1}{q} H(Q,f) = \frac{1}{q(q+1)} H(\Phi,f) \tag{40}$$

Note that all the three square matrices are diagonal matrices since the off-diagonal elements of $H(Q,f)$ are

$$\frac{\partial^2}{\partial f_\alpha \partial f_\beta} = \frac{\partial \dot{\xi}_\beta}{\partial f_\alpha} = 0, \quad (\alpha \neq \beta). \tag{41}$$

Obviously, the αth diagonal element of $\tilde{J}$ is just J_α.

3.2 Convexity Of Dissipation. Let us discuss the restriction of the entropy production inequality on the nonlinear phenomenological coefficient matrices, see Eq. (8). In view of Eqs. (35) and (39), it is required

$$\Phi = g^T J g = f^T \tilde{J} f \geqslant 0 \tag{42}$$

for any g or f. Thus, J and $\tilde{J}$ should be positive semidefinite, and then $H(Q,f)$, $H(Q,g)$, $H(\Phi,f)$, and $H(\Phi,g)$ should be also posi-

tive semidefinite. Obviously, If any one of the six matrices is positive semidefinite, the other ones are all positive semidefinite. Note that, if the Hessian matrix of a scalar function, say $Q(f)$, is positive semidefinite, the function is convex see, e.g., Maugin [16]. Therefore, the convexity of the flow potential Q or dissipation function Φ is required by the entropy production inequality. Note that $\tilde{J}$ is a diagonal matrix, and it is positive semidefinite if

$$J_\alpha = \frac{1}{q}\frac{\partial^2 Q}{\partial f_\alpha^2} = \frac{1}{q}\frac{\partial \dot{\xi}_\alpha}{\partial f_\alpha} \geqslant 0, \quad (\alpha = 1,2,\ldots,n) \tag{43}$$

which require that $\dot{\xi}_\alpha$ is a monotonic increasing function of the conjugate force f_α. The requirement can be recast, due to Eq. (29),

$$J_\alpha = \frac{1}{q}\frac{\partial \dot{\xi}_\alpha}{\partial f_\alpha} = \frac{\dot{\xi}_\alpha}{f_\alpha} \geqslant 0, \quad (\alpha = 1,2,\ldots,n) \tag{44}$$

which is equivalent to, in the sense of non-negativeness,

$$f_\alpha \dot{\xi}_\alpha \geqslant 0, \quad (\alpha = 1,2,\ldots,n) \tag{45}$$

As compared with Eq. (8), it is evident that the homogeneous conditions Eq. (29) require that the intrinsic dissipation inequality hold for each internal variable or *locally*.

Due to Eqs. (25) and (31), one obtains,

$$g_\mu \dot{\zeta}_\mu = \Phi(\mathbf{g}, \vartheta, H) = (q+1)Q(\mathbf{g}, \vartheta, H) \tag{46}$$

Differentiating Eq. (46) by g_κ and using Eq. (19), the global homogeneous conditions emerges,

$$\frac{\partial \dot{\zeta}_\alpha}{\partial g_\kappa} g_\alpha = q\dot{\zeta}_\kappa \quad \text{or} \quad \frac{\partial \dot{\boldsymbol{\zeta}}}{\partial \mathbf{g}} \cdot \mathbf{g} = q\dot{\boldsymbol{\zeta}} \tag{47}$$

which shows that the homogeneous property transfers exactly from local internal variables $\boldsymbol{\xi}$ to global internal variables $\boldsymbol{\zeta}$, as compared with Eq. (47) and Eq. (29). It should be emphasized that all deduction in this section is fully independent of the specific relation between $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$, so all results hold for both direct and incremental relations between them.

4 Some Discussions

As mentioned before, with the homogeneous kinetic rate laws, the convexity of Q or Φ with respect to the conjugate forces is required by the entropy production inequality. Here the convexity of Q or Φ with respect to stress $\boldsymbol{\sigma}$ is discussed.

Unlike conventional plasticity theory, the endochronic theory of plasticity proposed by Valanis [17,18] is directly based on irreversible thermodynamics. Here it is shown that the endochronic theory is closely related to the normality structures with homogeneous kinetic rate laws.

4.1 Convexity With Respect To Stress. In general, the convexity of Q or Φ with respect to $\mathbf{f}$ or $\mathbf{g}$ cannot be converted to that with respect to $\boldsymbol{\sigma}$. Although the following parallel normality structures hold,

$$\dot{\xi}_\alpha = \frac{\partial Q}{\partial f_\alpha}, \quad \dot{\zeta}_\mu = \frac{\partial Q}{\partial g_\mu}, \quad \frac{\mathrm{d}^p \boldsymbol{\varepsilon}}{\mathrm{d}t} = \frac{\partial Q}{\partial \boldsymbol{\sigma}}, \tag{48}$$

it does not imply that $\boldsymbol{\sigma}$: $\mathrm{d}^p\boldsymbol{\varepsilon}$ is the dissipated energy increment unlike $(1/V)f_\alpha d\xi_\alpha$ or $g_\mu d\zeta_\mu$. Here, we consider a special case that Φ is a homogeneous function of degree p in $\boldsymbol{\sigma}$, i.e.,

$$\frac{\partial \Phi}{\partial \boldsymbol{\sigma}} : \boldsymbol{\sigma} = p\Phi. \tag{49}$$

In this case, $(1/p)\boldsymbol{\sigma}$ can be understood as the thermodynamic force conjugate to $d^p\boldsymbol{\varepsilon}$. Differentiating Eq. (49) by $\boldsymbol{\sigma}$ leads to

$$\frac{\partial^2 \Phi}{\partial \boldsymbol{\sigma}^2} : \boldsymbol{\sigma} + \frac{\partial \Phi}{\partial \boldsymbol{\sigma}} = p\frac{\partial \Phi}{\partial \boldsymbol{\sigma}} \Rightarrow \frac{\partial^2 \Phi}{\partial \boldsymbol{\sigma}^2} : \boldsymbol{\sigma} = (p-1)\frac{\partial \Phi}{\partial \boldsymbol{\sigma}}. \tag{50}$$

Therefore,

$$\boldsymbol{\sigma} : \frac{\partial^2 \Phi}{\partial \boldsymbol{\sigma}^2} : \boldsymbol{\sigma} = (p-1)\boldsymbol{\sigma} : \frac{\partial \Phi}{\partial \boldsymbol{\sigma}} = (p-1)p\Phi \geqslant 0 \text{ for } p \geqslant 1. \tag{51}$$

Due to the arbitrariness of $\boldsymbol{\sigma}$, the fourth-order tensor $\partial^2\Phi/\partial\boldsymbol{\sigma}^2$ must be positive semidefinite, so Φ is convex with respect to $\boldsymbol{\sigma}$. Thus, Q is also convex with respect to $\boldsymbol{\sigma}$ due to $\Phi = (q+1)Q$.

4.2 Endochronic Constitutive Framework. In the formulation by use of Helmholtz free energy $\phi = \phi(\varepsilon, \vartheta, \mathbf{q})$, a set of phenomenological internal variables $\mathbf{q}$ are used to specify the current state of material internal structure see, e.g., Valanis [18]. The internal variable set contains m second-order tensors,

$$\mathbf{q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m\}, \quad \mathbf{q}_\beta = q_{ij}^\beta, \quad (p = 1,2,\ldots,m) \tag{52}$$

The evolution equations for the internal variables are

$$\frac{\partial \phi}{\partial \mathbf{q}_\beta} + \mathbf{B}_\beta : \frac{\mathrm{d}\mathbf{q}_\beta}{\mathrm{d}z} = \mathbf{0}, \quad (\beta = 1,2,\ldots,m) \tag{53}$$

where $\mathbf{B}_\beta$ is the fourth-order dissipation tensor for the βth internal variable. The evolution of the variables is with respect to a time-like parameter z which is often refereed to as the intrinsic time or the endochronic time. The intrinsic time is monotonically increasing and is defined in terms of the plastic strain. The evolution equations can be recast as

$$\dot{\mathbf{q}}_\beta = \mathbf{B}_\beta^{-1} : \mathbf{f}_\beta, \quad (\beta = 1,2,\ldots,m) \tag{54}$$

where $\dot{\mathbf{q}}_\beta = \mathrm{d}\mathbf{q}_\beta/\mathrm{d}z$ is the rate of the internal variable $\mathbf{q}_\beta$ with respect to the intrinsic time; $\mathbf{f}_\beta = -\partial\phi/\partial\mathbf{q}_\beta$ is the thermodynamic generalized force conjugate to $\mathbf{q}_\beta$. In the endochronic theory, the entropy production inequality is enforced for each internal variable, i.e.,

$$\mathbf{f}_\beta : \dot{\mathbf{q}}_\beta \geqslant 0 \ (\beta = 1,2,\ldots,m) \tag{55}$$

It is easy to show that the endochronic framework is exactly consistent with the normality structure with homogeneous kinetic rate laws from the following two viewpoints:

- Taking $\mathbf{q}$ at $\boldsymbol{\xi}$ level. It is shown in Eq. (54) that $\dot{\mathbf{q}}_\beta$ is linear with $\mathbf{f}_\beta$, or each rate of the internal variable is the homogeneous function of degree one in its conjugate force. Eq. (55) is exactly consistent with Eq. (45).
- Taking $\mathbf{q}$ at $\boldsymbol{\zeta}$ level. Let's divide $\boldsymbol{\xi}$ into m groups, and the βth group of local internal variables are represented only by one averaging variable, $\mathbf{q}_\beta$. Therefore, $\mathbf{q}_\beta$ is fully independent of other elements of $\mathbf{q}$, which leads to Eq. (55). Furthermore, Eq. (54) is just the direct result of Eq. (47) if each rate $\dot{\xi}_\alpha$ is a homogeneous function of degree one in its conjugate force f_α.

  Therefore, it is concluded that the endochronic constitutive framework is just a special case of the normality structures with homogeneous kinetic rate laws.

5 Application to Microcracked Solids

The essential properties of the normality structures with homogeneous kinetic rate laws have been revealed in the preceding sections. In this section, some further discussions are made from different viewpoints. One of the interesting results is that the refined normality structure directly leads to the restriction on quasi-static extension or healing of Griffith cracks by Rice [4]. Based on the discussions, it may be concluded that the homogeneous kinetic rate laws can really be considered as an intrinsic property of certain materials, especially for microcracked solids.

Rice [2] has applied the normality structure to a material sample containing some distribution of Griffith cracks. Let the locus of all crack fronts be denoted by L and let da be a function of position along L describing the amount of local advance of the cracks, and hence constituting the structural rearrangements. It is assumed that the surfaces of cracking have continuously turning tangent planes, without abrupt forking or branching. Therefore, Eq. (3) becomes

$$d^p\psi = -d^p\phi = \frac{1}{V} f_\alpha \, d\xi_\alpha \rightarrow \frac{1}{V}\int_L [F \, da] dL \tag{56}$$

where F denotes the thermodynamic crack extension force per unit length along L. Here the discrete expression of Eq. (3) is replaced by the continuous expression. Similarly, the flow potential defined in Eq. (10) is rewritten as

$$Q = \frac{1}{V}\int_L \int_0^F \dot{a} \, dF \, dL \tag{57}$$

The requirement by the entropy production inequality is, in view of Eq. (8),

$$\frac{1}{V}\int_L [F \, da] dL \geqslant 0. \tag{58}$$

As pointed by Rice [2], at any local crack front,

$$F = G - 2\gamma \tag{59}$$

where G is the Irwin energy release rate and γ is the surface free energy. Rather than the more usually cited condition that $G=2\gamma$ for the onset of crack extension, Rice [4] proposed the restriction on quasi-static extension or healing of Griffith cracks,

$$(G - 2\gamma)\dot{a} \geqslant 0 \tag{60}$$

at any local crack front. Evidently, the inequality (60) is only a sufficient condition for the requirement of the entropy production inequality, Eq. (58), but not a necessary condition for the requirement. In other words, the Rice [4] restriction is not a thermodynamic requirement which can only take the form, Eq. (58). However, this inequality can be considered as the result of the homogeneous kinetic rate laws. The homogeneous crack kinetic rate laws in the sense of Eq. (29) can be written as

$$\frac{\partial \dot{a}}{\partial F} F = q\dot{a} \quad \text{or} \quad \dot{a} = \frac{F}{q}\frac{\partial \dot{a}}{\partial F} \tag{61}$$

at each local crack front. The homogeneous kinetic rate laws lead to the local intrinsic dissipation inequality (45) which, in this case, can be rewritten as, at any local crack front,

$$F\dot{a} \geqslant 0 \quad \text{or} \quad (G - 2\gamma)\dot{a} \geqslant 0 \tag{62}$$

which is just the Rice [4] restriction on quasi-static growth of Griffith cracks.

It should be noted that the homogeneous condition Eq. (61) generally holds for cracking due to the widely used power laws. The subcritical crack growth Δa at a local crack front can often be covered by the following power-law,

$$\dot{a} \propto K^n \tag{63}$$

where K is the stress intensity factor at the crack front. For example, $n=13$ for the nickel-based superalloy Nimonic 80A at a temperature of 650°C [19]. The fatigue crack growth can also be described by similar power laws if taking cycle number N as the generalized time, e.g., the simple Paris equation $da/dN \propto (\Delta K)^n$ where the exponent n can take values as high as 15 to 50 in ceramics [20]. For time-independent cracking, such a power-law can be understood as R-curve. Due to $G \propto K^2$, the power law can be written as

$$\dot{a} = G^q, \quad \text{or} \quad \dot{a} = hG^q \tag{64}$$

where h and $q(=n/2)$ are material parameters. Except for an "ideal" brittle cracking, the surface free energy γ is generally much smaller than the required energy release rate G, i.e., $\gamma \ll G \Rightarrow G \approx F$. Thus, the following crack kinetic rate law possesses a solid physical basis,

$$\dot{a} = hF^q \tag{65}$$

which is consistent with the homogeneous condition (61). Inserting Eq. (65) into Eq. (57), then yields

$$Q = \frac{h}{(q+1)V}\int_L F^{q+1} \, dL \tag{66}$$

The time-independent stable crack growth can be described by R-curve. The stable crack growth conditions are

$$\frac{\partial G}{\partial a} < \frac{\partial R}{\partial a}, \quad G = R. \tag{67}$$

The R-curve, i.e., $R=R(\Delta a)$, can also be covered by the power-law

$$\Delta a = a - a_0 \propto R^q. \tag{68}$$

Broek [21] shows that $q=3.4-5.9$ for an Al–Zn–Mg alloy of different sheet thicknesses. The power-law can be recast as, due to Eq. (67),

$$a - a_0 = hR^q = hG^q \Rightarrow \dot{a} = hqG^{q-1}\dot{G} \tag{69}$$

where h is a scaling constant. With $G \approx F$, such rate laws are obtained

$$\Delta a = hF^q \Rightarrow \dot{a} = hqF^{q-1}\dot{F} \quad \text{or} \quad \frac{da}{dF} = hqF^{q-1} \tag{70}$$

which can be considered as the homogeneous rate laws for time-independent process, as compared with Eq. (65). Incidently, the reasoning chain of this section has also been briefed by Yang et al. [22,23].

6 Maximum Dissipation Rate and Onsager Fluxes

In this section, it will be revealed that the homogeneous kinetic rate laws can be considered as the requirement by the principle of the maximum dissipation rate of Ziegler [6]. Actually, the introduction of the homogeneous kinetic rate laws is inspired by the comparison between the normality structures and the Ziegler [6] orthogonality condition which can be deduced from the principle of the maximum dissipation rate.

Based on the kinetic rate laws defined in Eq. (9), the total dissipation function can be decomposed as

$$\Phi = \sum_{\alpha=1}^{n} \Phi^{(\alpha)}, \quad \Phi^{(\alpha)} = \frac{1}{V} f_\alpha \dot{\xi}_\alpha, \quad (\alpha = 1,2, \ldots, n). \tag{71}$$

Ziegler [6] refers to thermodynamic systems or processes involving only one coherent rate as *elementary*, and refers to thermodynamic systems or processes involving more than one coherent rates as *complex*. A complex system or process is referred to *compound* if it can be uncoupled into elementary subsystems with well-defined dissipation functions like Eq. (71). Thus, the introduction of the Rice [1,2] kinetic rate laws in reality is to define the concerned system as a compound system. It is a very strong assumption.

For each elementary subsystem, the orthogonality condition should hold,

$$\dot{\xi}_\alpha = V\lambda^{(\alpha)}\frac{\partial\Phi^{(\alpha)}}{\partial f_\alpha}, \quad (\alpha = 1,2,\ldots,n) \tag{72}$$

where $\lambda^{(\alpha)}$ is a Lagrangian multiplier, which is required by the principle of maximum dissipation rate, see Ziegler [6]. In general, the orthogonality condition is not consistent with the normality structure with respect to a flow potential Q, Eq. (10). In other words, the Rice [1,2] normality structures generally do not satisfy the principle of maximum dissipation rate.

Let us examine the consistency conditions between the two normality relations. Differentiating Eq. (71) by f_α, it follows that

$$V\frac{\partial\Phi^{(\alpha)}}{\partial f_\alpha} = \dot{\xi}_\alpha + f_\alpha\frac{\partial\dot{\xi}_\alpha}{\partial f_\alpha}, \quad (\alpha = 1,2,\ldots,n). \tag{73}$$

Substituting Eq. (72) into Eq. (73), the consistency condition is obtained,

$$f_\alpha\frac{\partial\dot{\xi}_\alpha}{\partial f_\alpha} = \left(\frac{1}{\lambda^{(\alpha)}} - 1\right)\dot{\xi}_\alpha, \quad (\alpha = 1,2,\ldots,n) \tag{74}$$

which implies that $\dot{\xi}_\alpha$ should be a homogeneous function in f_α. As pointed out by Ziegler [6], even if the local orthogonality condition Eq. (72) holds for each elementary subsystem, it does not imply that the orthogonality condition also holds for the total system, i.e.,

$$\dot{\xi}_\alpha = V\lambda\frac{\partial\Phi}{\partial f_\alpha}. \tag{75}$$

Evidently, the total orthogonality condition can be achieved by requiring that all $\dot{\xi}_\alpha$ are of the same homogeneous degree q, i.e.,

$$\frac{\partial\dot{\xi}_\alpha(f_\alpha,\vartheta,H)}{\partial f_\alpha}f_\alpha = q\dot{\xi}_\alpha(f_\alpha,\vartheta,H) \quad (\alpha = 1,2,\ldots,n), \tag{76}$$

which require that $\lambda^{(\alpha)}$ be a constant, due to Eq. (74),

$$\lambda^{(\alpha)} = \frac{1}{q+1} = \lambda \quad (\alpha = 1,2,\ldots,n), \tag{77}$$

so the total normality condition holds due to Eqs. (71) and (72). In view of Eqs. (10), (19), and (31), the following relations are evident,

$$\dot{\xi}_\alpha = \lambda\frac{\partial\Phi(\mathbf{f},\vartheta,H)}{\partial f_\alpha}, \quad \dot{\zeta}_\mu = \lambda\frac{\partial\Phi(\mathbf{g},\vartheta,H)}{\partial g_\mu} \tag{78}$$

which are just Ziegler's orthogonality conditions [6] and can also be considered as the requirement of the principle of maximum dissipation rate. Therefore, it is concluded that the homogeneous condition on the kinetic rate law, Eq. (76) or Eq. (29), is equivalent to the requirement of the principle of maximum dissipation rate.

It should be pointed out that the structural rearrangements at the level considered by Rice [1,2] are constrained only by the second law of thermodynamics. The constraint by the principle of maximum dissipation rate is not essential at this level and cannot be taken as a general thermodynamic principle. It is instead nothing more than reasonable classification of behavior for certain materials.

Rice's kinetic rate laws of local internal variables, with each rate being stress dependent only via its conjugate thermodynamic force, are corner stones of the normality structure. If the kinetic rate laws are violated, the development of the flow potential along with the normality structures are invalid. However, even in this case such normality structures can be obtained based on the assumption of maximum rate of dissipation. The reason for the inconsistency is obvious. As pointed by Ziegler [6], the principle of maximum dissipation rate is much more general than the orthogonality condition. Indeed, Ziegler [6] even suggests that the second law of thermodynamics can be covered by the principle of maximum rate of dissipation.

In general, the kinetic rate laws of local internal variables take the form,

$$\dot{\xi}_\alpha = \dot{\xi}_\alpha(\mathbf{f},\vartheta,H), \tag{79}$$

and Rice's kinetic rate laws defined by Eq. (9) is just a special case of Eq. (79). For linear dependence of $\dot{\boldsymbol{\xi}}$ on $\mathbf{f}$, the phenomenological equations and Onsager reciprocal relations read,

$$\dot{\xi}_\alpha = L_{\alpha\beta}f_\beta, \quad L_{\alpha\beta} = L_{\beta\alpha} \tag{80}$$

where $\alpha, \beta = 1,2,\ldots,n$. For general nonlinear dependence of $\dot{\boldsymbol{\xi}}$ on $\mathbf{f}$ given by Eq. (79), the nonlinear Onsager reciprocal relations of Edelen [7] read,

$$\frac{\partial\dot{\xi}_\alpha}{\partial f_\beta} = \frac{\partial\dot{\xi}_\beta}{\partial f_\alpha} \tag{81}$$

where $\alpha, \beta = 1,2,\ldots,n$. Evidently, if inserting the linear relation Eq. (80) into Eq. (81), the Onsager reciprocal relations Eq. (80) emerge. The thermodynamic fluxes $\dot{\xi}_\alpha$ satisfying Eq. (81) is termed *Onsager fluxes* by [8]. The nonlinear Onsager relations, Eq. (81), have been shown by [7,8] to result from the requirement that the entropy production rate σ defined by Eq. (8) be a nonnegative, convex function of $\mathbf{f}$ with a minimum at the equilibrium point. Evidently, Rice's kinetic rate laws defined by Eq. (9) satisfy the nonlinear Onsager reciprocal relations (81) automatically. In other words, Rice's restriction on the kinetic rate laws is in reality the requirement of the nonlinear Onsager reciprocal relations.

With the general kinetic rate laws (79), the nonlinear Onsager reciprocal relations, Eq. (81), are just the necessary and sufficient condition that the differential $\dot{\xi}_\alpha \mathrm{d}f_\alpha$ is an exact differential, i.e., $\mathrm{d}Q = (1/V)\dot{\xi}_\alpha \mathrm{d}f_\alpha$, so that

$$\dot{\xi}_\alpha = V\frac{\partial Q}{\partial f_\alpha}, \quad Q = Q(\mathbf{f},\vartheta,H) = \frac{1}{V}\int_0^{\mathbf{f}}\dot{\xi}_\alpha(\mathbf{f},\vartheta,H)\mathrm{d}f_\alpha, \tag{82}$$

which is exactly the same as Eq. (10) but the kinetic rate laws take the form of Eq. (79). Evidently, the normality structure given by Eq. (12) still holds based on Eq. (82).

7 Damage Evolution Laws

It is usually assumed that there exists a scalar damage dissipation potential Q in phenomenological damage models, and then the damage evolution laws are derived from it by normality condition,

$$\dot{\boldsymbol{\Omega}} = \frac{\partial Q}{\partial \mathbf{Y}} \tag{83}$$

where $\boldsymbol{\Omega}$ denotes a damage variable and is considered as a second-order tensor here without a loss of generality, and $\mathbf{Y}$ is the generalized thermodynamic force conjugate to $\boldsymbol{\Omega}$. If further assuming that Q is a quadratic function in the conjugate force $\mathbf{Y}$, e.g., $Q = \frac{1}{2}\mathbf{Y}:\mathbf{J}:\mathbf{Y}$, the phenomenological equation or linear irreversible thermodynamics appears from Eq. (83),

$$\dot{\boldsymbol{\Omega}} = \mathbf{J}:\mathbf{Y} \tag{84}$$

where $\mathbf{J}$ is termed damage characteristic tensor of rank four. Chow and Lu [24] have shown that many classical damage evolution laws can be covered by Eq. (84). The latest damage model of Soh et al. [25] also follows this line.

In phenomenological damage models, it is usually assumed that the current microstructure of the material sample is uniquely characterized by the current damage variable $\boldsymbol{\Omega}$. In this sense, the damage variable $\boldsymbol{\Omega}$ is equivalent to H, the parameter denoting the

Table 1 Damage evolution laws for time dependent and independent processes

	Time-Dependent processes	Time-independent processes
Local kinetics	$\dot{a}=hF^q$	$\Delta a=hF^q$
Flow potentials	$Q=h/(q+1)V\int_L F^{q+1}\mathrm{d}L$	$Q=h/V\int_L F^q\dot{F}\mathrm{d}L$
Damage evolution laws	$\dot{\mathbf{\Omega}}=\mathbf{J}:\dot{\mathbf{Y}}$	$\dot{\Omega}=q\mathbf{J}:\dot{\mathbf{Y}}$

current pattern of microstructural rearrangement of constituent elements of the materials. Therefore, the conjugate force $\mathbf{Y}$ is determined by

$$\mathbf{Y}=\frac{\partial\psi}{\partial\mathbf{\Omega}}=-\frac{\partial\phi}{\partial\mathbf{\Omega}},\quad \phi=\phi(\varepsilon,\vartheta,\mathbf{\Omega}),\quad \psi=\psi(\boldsymbol{\sigma},\vartheta,\mathbf{\Omega}). \tag{85}$$

Evidently, the homogeneous local rate laws, Eq. (29), lead to the following properties, in view of Eqs. (34), (47), and (32),

$$\dot{\mathbf{\Omega}}=\frac{\partial Q}{\partial\mathbf{Y}}=\mathbf{J}:\mathbf{Y},\quad \mathbf{J}=\frac{1}{q}\frac{\partial^2 Q}{\partial\mathbf{Y}^2},\quad \frac{\partial Q}{\partial\mathbf{Y}}:\mathbf{Y}=(q+1)Q. \tag{86}$$

It should be emphasized that these properties are irrelevant to the specific physical meanings and tonsorial characters of the damage tensor, and the quadratic assumption, $Q=\frac{1}{2}\mathbf{Y}:\mathbf{J}:\mathbf{Y}$, is unnecessary.

Swoboda and Yang [26] and Yang et al. [27] try hard to solve the crux under what conditions the phenomenological equation (84) holds, but their answers are plausible and only confined to second-order fabric tensors. Evidently, the essential condition is that the local rate laws are homogeneous function, as shown in Eq. (29). Thus, this crux is concluded in this paper.

7.1 Damage Characteristic Tensor of Microcracked Solids. In this part, microcracks and and their propagation are considered as the dominant microdefects and energy dissipation mechanism in a solid. Indeed, microcracks attracted, and still attract, most interest due to its relevance to the structural reliability and failure, as remarked by Krajcinovic [14]. With the microcracks described in Sec. 5, the damage tensor and microcracks are related by, similar to Eq. (21),

$$\mathrm{d}\mathbf{\Omega}=\int_L \mathbf{R}\,\mathrm{d}a\,\mathrm{d}L \tag{87}$$

where $\mathbf{R}$ denotes the contribution to the damage tensor due to a unit local crack advance per unit length along L at a certain local crack front. Obviously, $\mathbf{\Omega}$ and $\mathbf{R}$ possess the same tensorial characters. The specific form of $\mathbf{R}$ depends on the specific definition of $\mathbf{\Omega}$. Here it is only assumed that there exists a definite distribution of $\mathbf{R}$ along crack fronts for a given microstructure or rearrangement, i.e., $\mathbf{R}=\mathbf{R}(\mathbf{\Omega})$. Then Eq. (16) can be recast as, due to Eq. (56)

$$\mathbf{Y}:\delta\mathbf{\Omega}=\frac{1}{V}\int_L [F\,\delta a]\mathrm{d}L \tag{88}$$

which along with Eq. (87) leads to

$$F=V\mathbf{R}:\mathbf{Y}. \tag{89}$$

The damage characteristic tensor $\mathbf{J}$ is then obtained by Eq. (86) along with Eqs. (66) and (89)

$$\mathbf{J}=\frac{1}{q}\frac{\partial^2 Q}{\partial\mathbf{Y}^2}=hV^q\int_L(\mathbf{R}:\mathbf{Y})^{q-1}\mathbf{R}\mathbf{R}\,\mathrm{d}L \tag{90}$$

It is easy to verify

$$Q=\frac{1}{q+1}\mathbf{Y}:\mathbf{J}:\mathbf{Y} \tag{91}$$

To pursue analytic damage characteristic tensors has been the key object of Swoboda and Yang [26] and Yang et al. [27]. Evidently, their deduced results are just some special cases of Eq. (90).

7.2 Time-Independent Damaging Processes. For time-independent damaging processes, the local kinetics or rate laws can be described by Eq. (70). In this part, let us confine the formulation only to the fully loading processes in which all crack fronts are propagating. In this case, a quasi flow potential can still be calculated by Eq. (57) along with the rate laws Eq. (70),

$$Q=\frac{1}{V}\int_L\int_0^F \dot{a}\,\mathrm{d}F\,\mathrm{d}L=\frac{h}{V}\int_L F^q\dot{F}\,\mathrm{d}L \tag{92}$$

In view of Eq. (89),

$$F=V\mathbf{R}:\mathbf{Y}\Rightarrow\dot{F}=V\mathbf{R}:\dot{\mathbf{Y}} \tag{93}$$

Then the damage evolution law is obtained by the normality condition along with Eqs. (92) and (93)

$$\dot{\mathbf{\Omega}}=\frac{\partial Q}{\partial\mathbf{Y}}=qhV^q\left(\int_L(\mathbf{R}:\mathbf{Y})^{q-1}\mathbf{R}\mathbf{R}\,\mathrm{d}L\right):\dot{\mathbf{Y}} \tag{94}$$

or

$$\dot{\mathbf{\Omega}}=q\mathbf{J}:\dot{\mathbf{Y}},\quad \mathbf{J}=hV^q\int_L(\mathbf{R}:\mathbf{Y})^{q-1}\mathbf{R}\mathbf{R}\,\mathrm{d}L \tag{95}$$

where the $\mathbf{J}$ is the same as the one defined in Eq. (90). Evidently, the damage evolution laws for time-independent processes have the similar structures of the ones for time-dependent processes, as listed in Table 1. It should be noted that the formulation in this part is only valid for fully loading processes.

8 Conclusion

Rice's kinetic rate laws of local internal variables, with each rate being stress dependent only via its conjugate thermodynamic force, are corner stones of the normality structures and certain specific Onsager fluxes of Edelen [7,8]. It is revealed in this paper that nonlinear phenomenological equations and Onsager reciprocal relations emerge naturally from the normality structures if each rate is a homogeneous function of degree q in its conjugate force. Furthermore, the nonlinear phenomenological coefficient matrix is identical to the Hessian matrix of the flow potential function in conjugate forces only scaled by q, and the homogeneous property transfers exactly from local internal variables to global average internal variables.

Within the framework of the normality structures with the homogeneous rate laws, the second law of thermodynamics requires the convexity of the flow potentials and dissipation functions, and the second law also lead to the refined version of Griffith criterion proposed by Rice [4], $(G-2\gamma)\dot{a}\geqslant 0$, for microcracked solids. Furthermore, the revealed remarkable properties help us gain a deep insight into the structures of anisotropic damage evolution laws. The unsolved crux on the conditions of the widely used phenomenological equations in continuum damage mechanics, raised by Swoboda and Yang [26] and Yang et al. [27], is also easily concluded in this framework.

Thermodynamic systems formulated within the framework of normality structures by Rice [1,2] are certain compound systems of Ziegler [6]. The homogeneous property of the rate laws is equivalent to the constraint by the principle of maximum dissipa-

tion rate [6]. Such a property or constraint is not essential as compared with the second law of thermodynamics, but it really represents a broad class of inelastic behaviors, such as microcracking-dominated damaging.

Acknowledgments

The work reported here was supported by the NSF with Grant No. 50279016 and EYTP of China and subsidized by the special funds for major state basic research projects with Grant No. 2002cb412708.

References

[1] Rice, J. R., 1971, "Inelastic Constitutive Relations for Solids: An Integral Variable Theory and Its Application to Metal Plasticity," J. Mech. Phys. Solids, **19**, pp. 433–455.
[2] Rice, J. R., 1975, "Continuum Mechanics and Thermodynamics of Plasticity in Relation to Microscale Deformation Mechanisms," *Constitutive Equations in Plasticity*, A. S. Argon, ed., MIT Press, Cambridge, MA, pp. 23–79.
[3] Onsager, L., 1931, "Reciprocal Relations in Irreversible Processes, I, II," Phys. Rev., **37**, pp. 405–406; 1931, "Reciprocal Relations in Irreversible Processes, I, II," Phys. Rev., **38**, pp. 2265–2279.
[4] Rice, J. R., 1978, "Thermodynamics of Quasi-Static Growth of Griffith Cracks," J. Mech. Phys. Solids, **26**, pp. 61–78.
[5] De Groot, S. R., Mazur, P., 1962, *Non-equilibrium Thermodynamics*, North-Holland, Amsterdam.
[6] Ziegler, H., 1977, *An Introduction to Thermomechanics*, North-Holland, Amsterdam.
[7] Edelen, D. G. B., 1972, "A Nonlinear Onsager Theory of Irreversibility," Int. J. Eng. Sci., **10**, pp. 481–490.
[8] Edelen, D. G. B., 1973, "Asymptotic Stability, Onsager Fluxes and Reaction Kinetics," Int. J. Eng., **11**, pp. 819–839.
[9] Ziegler, H., Wehrli, C., 1987, "On a Principle of Maximal Rate of Entropy Production," J. Non-Equil. Thermodyn., **12**, pp. 229–243.
[10] Rajagopal, K. R., Srinivasa, A. R., 1998, "Mechanics of the Inelastic Behavior of Materials. II. Inelastic Response," Int. J. Plast., **14**, pp. 969–995.
[11] Rajagopal, K. R., Srinivasa, A. R., 2000, "A Thermodynamic Framework for Rate-Type Fluid Models," J. Non-Newtonian Fluid Mech., **88**, pp. 207–227.
[12] Rajagopal, K. R., and Srinivasa, A. R., 2004, "On Thermomechanical Restrictions of Continua," Proc. R. Soc. London, Ser. A, **460**, pp. 631–651.
[13] Puzrin, A. M., Houlsby, G. T., 2003, "Rate-Dependent Hyperplasticity With Internal Functions," J. Eng. Mech., **129**, pp. 252–263.
[14] Krajcinovic, D., 2000, "Damage Mechanics: Accomplishments, Trends and Needs," Int. J. Solids Struct., **37**, pp. 267–277.
[15] Lemaitre, J., Desmorat, R., Sauzay, M., 2000, "Anisotropic Damage Law of Evolution," Eur. J. Mech. A/Solids, **19**, pp. 187–208.
[16] Maugin, G. A., 1999, *The Thermodynamics of Nonlinear Irreversible Behaviors*, World Scientific, Singapore.
[17] Valanis, K. C., 1971, "A Theory of Viscoplasticity Without a Yield Surface, Part I. General Theory," Arch. Mech., **23**, pp. 517–533.
[18] Valanis, K. C., 1975, "On the Foundations of the Endochronic Theory of Viscoplasticity," Arch. Mech., **27**, pp. 857–868.
[19] Delph, T. J., 1999, "A Simple Model for Crack Growth in Creep Resistant Alloys," Int. J. Fract., **98**, pp. 77–86.
[20] Ritchie, R. O., Gilbert, C. J., McNaney, J. M., 2000, "Mechanics and Mechanisms of Fatigue Damage and Crack Growth in Advanced Materials," Int. J. Solids Struct., **37**, pp. 311–329.
[21] Broek, D., 1987, *Elementary Engineering Fracture Mechanics*, Martinus Nijhoff, Dordrecht, The Netherlands.
[22] Yang, Q., Tham, L. G., Swoboda, G., 2002, "Micromechanical Basis of Nonlinear Phenomenological Equations as Damage Evolution Laws," Mech. Res. Commun., **29**, pp. 131–136.
[23] Yang, Q., Tham, L. G., Swoboda, G., 2004, "Relationship Between Refined Griffith Criterion and Power Laws for Cracking," Mech. Res. Commun., **31**, pp. 429–434.
[24] Chow, C. L., Lu, T. J., 1989, "On Evolution Laws of Anisotropic Damage," Eng. Fract. Mech., **34**, pp. 679–701.
[25] Soh, C. K., Liu, Y., Yang, Y., Dong, Y., 2003, "A Displacement Equivalence-Based Damage Model for Brittle Materials—Part I: Theory," ASME J. Appl. Mech., **70**, pp. 681–687.
[26] Swoboda, G., Yang, Q., 1999, "An Energy-Based Damage Model of Geomaterials-II. Deduction of Damage Evolution Laws," Int. J. Solids Struct., **36**, pp. 1735–1755.
[27] Yang, Q., Zhou, W. Y., Swoboda, G., 1999, "Micromechanical Identification of Anisotropic Damage Evolution Laws," Int. J. Fract., **68**, pp. 740–750.

An Efficient and Accurate Numerical Method of Stress Intensity Factors Calculation of a Branched Crack

Xiangqiao Yan
Research Laboratory on Composite Materials,
Harbin Institute of Technology,
Harbin 150001, P.R. China

Based on the analytical solution of Crouch to the problem of a constant discontinuity in displacement over a finite line segment in an infinite elastic solid, in the present paper, the crack-tip displacement discontinuity elements, which can be classified as the left and the right crack-tip elements, are presented to model the singularity of stress near a crack tip. Furthermore, the crack-tip elements together with the constant displacement discontinuity elements presented by Crouch and Starfied are used to develop a numerical approach for calculating the stress intensity factors (SIFs) of general plane cracks. In the boundary element implementation, the left or the right crack-tip element is placed locally at the corresponding left or right crack tip on top of the constant displacement discontinuity elements that cover the entire crack surface and the other boundaries. The method is called the hybrid displacement discontinuity method (HDDM). Numerical examples are given and compared with the available solutions. It can be found that the numerical approach is simple, yet very accurate for calculating the SIFs of branched cracks. As a new example, cracks emanating from a rhombus hole in an infinite plate under biaxial loads are taken into consideration. The numerical results indicate the efficiency of the present numerical approach and can reveal the effect of the biaxial load on the SIFs. In addition, the hybrid displacement discontinuity method together with the maximum circumferential stress criterion (Erdogan and Sih) becomes a very effective numerical approach for simulating the fatigue crack propagation process in plane elastic bodies under mixed-mode conditions. In the numerical simulation, for each increment of crack extension, remeshing of existing boundaries is not required because of an intrinsic feature of the HDDM. Crack propagation is simulated by adding new boundary elements on the incremental crack extension to the previous crack boundaries. At the same time, the element characters of some related elements are adjusted according to the manner in which the boundary element method is implemented. [DOI: 10.1115/1.1796449]

1 Introduction

Among the different configurations of branched cracks, the singly branched crack, as shown in Fig. 1, has received the most attention in the literature. There have been many attempts [1–12] to solve this problem for arbitrarily values of a/b, the ratio of the half-crack length a of the main crack to the branch crack length b. Of particular physical interest is the limiting case as $a/b \rightarrow \infty$, where the solution [3,4] has been used to predict the initial angle of the branching of a crack in brittle solids under mixed-mode loading.

The majority of the analyses on branched cracks were based on the Muskhelishvili potential formulation and conformal mapping of the branched crack geometry. With the development of numerical computational techniques, numerical methods, in particular, finite element methods and boundary element methods are used extensively in solving the crack problems. It is well known that how to model the crack is the key issue in the analyses. Among several elastic two-dimensional crack modeling strategies by the boundary element methods, there exist the multidomain formulation [13], the stress formulation with regularization [14], and the dual boundary element method [15,16]. For each formulation, in order to model the singularity of stress near a crack tip, options are available such as building in the crack-tip stress singularity [17], using the quarter-point boundary element [13], and strategically refining the near-crack-tip nonsingular element. Further details on elastic crack analysis by the boundary element method are given in Refs. [18], [19].

Even though much achievement has been made in crack-modeling techniques, both simple and very accurate crack-modeling techniques still need to be developed, in particular for branched crack problems and crack propagation problems. The displacement discontinuity boundary element method [20,21] is very well suited for analyzing plane crack problems because, physically, one can imagine a displacement discontinuity as a line crack whose opposing surfaces are displaced relative to one another. Based on the analytical solution [21] to the problem of a constant discontinuity in displacement over a finite line segment in an infinite elastic solid, in the present paper, the crack-tip displacement discontinuity elements, which can be classified as the left and the right crack-tip displacement discontinuity elements, are presented to model the singularity of stress near a crack tip. Furthermore, the crack-tip elements together with the constant displacement discontinuity elements presented by Crouch and Starfied are used to develop a numerical approach for calculating the stress intensity factors (SIFs) of general plane cracks. In the boundary element implementation, the left or the right crack-tip element is placed locally at the corresponding left or right crack

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, September 16, 2002; final revision, May 5, 2004. Associate Editor: H. Gao. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

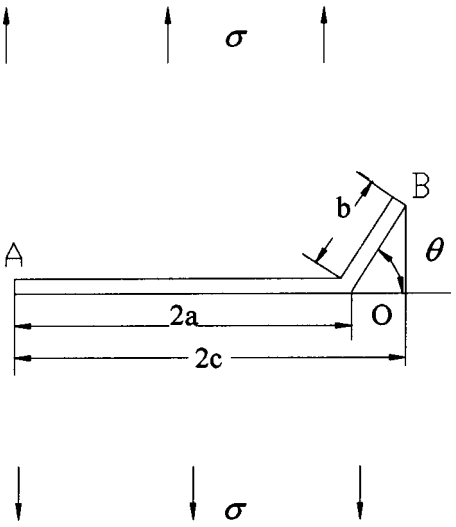


Fig. 1 A singly branched crack

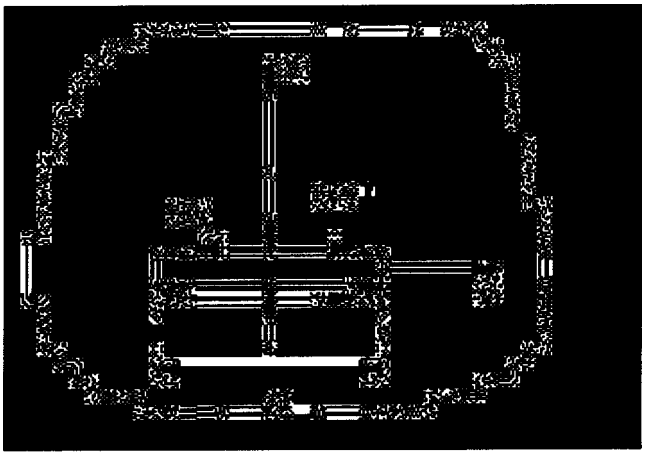


Fig. 2 Schematic of constant displacement discontinuity components D_x and D_y

tip on top of constant displacement discontinuity elements that cover the entire crack surface and the other boundaries. The method is called the hybrid displacement discontinuity method (HDDM). Numerical examples are given and compared with the available solutions. It can be seen that the numerical approach is simple, yet very accurate for calculating the SIFs of branched cracks. As a new example, cracks emanating from a rhombus hole in an infinite plate under biaxial loads are taken into account. The numerical results indicate the efficiency of the present numerical approach and can reveal the effect of the biaxial load on the SIFs.

On the application of boundary element methods to crack propagation analysis, the first attempt to automatically model crack propagation under mixed-mode conditions was given by Ingraffea, Blandford, and Liggett [22] for two-dimensional problems. They used the multiregion method [13] together with the maximum circumferential stress criterion to calculate the direction of crack propagation. Aliabadi [23] pointed out that the difficulty with the multiregion method is that the introduction of artificial boundaries to divide the regions is not unique, and that thus it is not easy to implement it in an automatic procedure. In an incremental crack extension analysis, these artificial boundaries must be repeatedly introduced for each increment of crack extension. In this paper, the hybrid displacement discontinuity method together with the maximum circumferential stress criterion [24] becomes a very effective numerical approach for simulating the fatigue crack propagation process in plane elastic bodies under mixed-mode conditions. In the numerical simulation, for each increment of crack extension, remeshing of existing boundaries is not required because of an intrinsic feature of the HDDM. Crack propagation is simulated by adding new boundary elements on the incremental crack extension to the previous crack boundaries. At the same time, the element characters of some related elements are adjusted according to the manner in which the boundary element method is implemented. As an example, the fatigue propagation process of cracks emanating from a circular hole in a plane elastic plate is simulated using the numerical simulation approach.

By the way, it is pointed out here that finite element simulations [25,26] when used to analyze crack problems have to face large computational problems connected with the discretization of the continuum into finite elements, particularly when some cracks propagate, thus changing the interior boundaries of the solids.

2 The Hybrid Displacement Discontinuity Method

The numerical approach presented in this paper for calculating the SIFs of branched cracks consists of the constant displacement discontinuity elements presented by Crouch and Starfield [20] and the crack-tip displacement discontinuity elements proposed herein.

2.1 Theoretical Foundation of the Constant Displacement Discontinuity Element. The problem of a constant displacement discontinuity over a finite line segment in the (x, y) plane of an infinite elastic solid is specified by the condition that the displacements be continuous everywhere except over the line segment in question. The line segment may be chosen to occupy a certain portion of the x axis, say the portion $|x|<a$, $y=0$. If we consider this segment to be a line crack, we can distinguish its two surfaces by saying that one surface is on the *positive* side of $y=0$, denoted $y=0_+$, and the other is on the *negative* side, denoted $y=0_-$. In crossing from one side of the line segment to the other, the displacements undergo a *constant* specified change in value $D_i=(D_x,D_y)$.

The displacement discontinuities D_i are defined as the difference in displacement between the two sides of the segment:

$$\begin{aligned} D_x &= u_x(x,0_-)-u_x(x,0_+),\\ D_y &= u_y(x,0_-)-u_y(x,0_+). \end{aligned} \tag{1}$$

Because u_x and u_y are positive in the positive x and y coordinate directions, it follows that D_x and D_y are positive as illustrated in Fig. 2. The solution to the subject problem is given by Crouch [21]. The displacements and stresses can be written as

$$\begin{aligned} u_x &= D_x[2(1-\nu)F_3(x,y,a)-yF_5(x,y,a)]\\ &\quad + D_y[-(1-2\nu)F_2(x,y,a)-yF_4(x,y,a)],\\ u_y &= D_x[(1-2\nu)F_2(x,y,a)-yF_4(x,y,a)]\\ &\quad + D_y[2(1-\nu)F_3(x,y,a)-yF_5(x,y,a)], \end{aligned} \tag{2}$$

and

$$\begin{aligned} \sigma_{xx} &= 2GD_x[2F_4(x,y,a)+yF_6(x,y,a)]\\ &\quad + 2GD_y[-F_5(x,y,a)+yF_7(x,y,a)],\\ \sigma_{yy} &= 2GD_x[-yF_6(x,y,a)]\\ &\quad + 2GD_y[-F_5(x,y,a)-yF_7(x,y,a)],\\ \sigma_{xy} &= 2GD_x[-F_5(x,y,a)\\ &\quad + yF_7(x,y,a)]+2GD_y[-yF_6(x,y,a)]. \end{aligned} \tag{3}$$

Functions F_2 through F_7 in these equations are

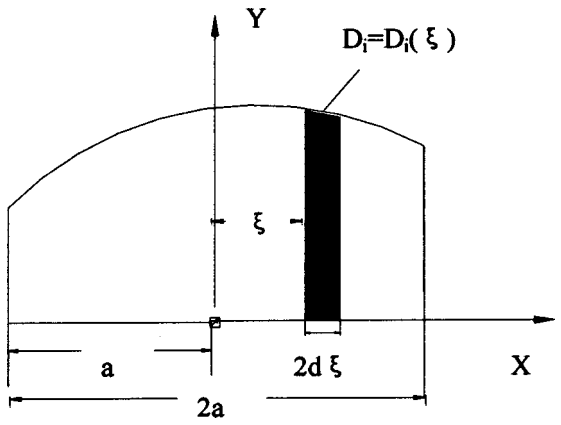


Fig. 3 Schematic of an arbitrary displacement discontinuity function and its differential element

$$
\begin{aligned}
F_2(x,y,a)&=f_{,x}=\frac{1}{4\pi(1-\nu)}\\
&\quad\times[\ln\sqrt{(x-a)^2+y^2}-\ln\sqrt{(x+a)^2+y^2}],\\
F_3(x,y,a)&=f_{,y}=-\frac{1}{4\pi(1-\nu)}\left[\arctan\frac{y}{x-a}-\arctan\frac{y}{x+a}\right],\\
F_4(x,y,a)&=f_{,xy}=\frac{1}{4\pi(1-\nu)}\left[\frac{y}{(x-a)^2+y^2}-\frac{y}{(x+a)^2+y^2}\right],\\
F_5(x,y,a)&=f_{,xx}=-f_{,yy}\\
&=\frac{1}{4\pi(1-\nu)}\left[\frac{x-a}{(x-a)^2+y^2}-\frac{x+a}{(x+a)^2+y^2}\right],\\
F_6(x,y,a)&=f_{,xyy}=-f_{,xxx}\\
&=\frac{1}{4\pi(1-\nu)}\left[\frac{(x-a)^2-y^2}{\{(x-a)^2+y^2\}^2}-\frac{(x+a)^2-y^2}{\{(x+a)^2+y^2\}^2}\right],\\
F_7(x,y,a)&=f_{,yyy}=-f_{,xxy}\\
&=\frac{2y}{4\pi(1-\nu)}\left[\frac{x-a}{\{(x-a)^2+y^2\}^2}-\frac{x+a}{\{(x+a)^2+y^2\}^2}\right],
\end{aligned}
\tag{4}
$$

where

$$
\begin{aligned}
f(x,y)=&\frac{-1}{4\pi(1-\nu)}\left[y\left(\arctan\frac{y}{x-a}-\arctan\frac{y}{x+a}\right)\right.\\
&\left.-(x-a)\ln\sqrt{(x-a)^2+y^2}+x+a\ln\sqrt{(x+a)^2+y^2}\right].
\end{aligned}
\tag{5}
$$

G and ν in these equations are shear modulus and the Poisson's ratio, respectively. Equations (2) and (3) are used by Crouch and Starfield [20] to set up a constant displacement discontinuity boundary element method.

2.2 Basic Formulas Required to Set Up a Higher Displacement Discontinuity Element. Now, consider arbitrary displacement discontinuity distributions along element length $2a$, as shown in Fig. 3:

$$D_i=D_i(\xi)\quad(i=1,2)\tag{6a}$$

or

$$D_x=D_x(\xi),\quad D_y=D_y(\xi).\tag{6b}$$

Based on the solution of the constant discontinuity in displacement given by Crouch [21], the displacements and stresses at a point (x, y) due to a *differential element* with its length $2d\xi$ and its center ξ (see Fig. 3) can be obtained from a differential viewpoint:

$$
\begin{aligned}
du_x&=D_x(\xi)[2(1-\nu)T_3(x,y,\xi,d\xi)-yT_5(x,y,\xi,d\xi)]\\
&\quad+D_y(\xi)[-(1-2\nu)T_2(x,y,\xi,d\xi)-yT_4(x,y,\xi,d\xi)],\\
du_y&=D_x(\xi)[(1-2\nu)T_2(x,y,\xi,d\xi)-yT_4(x,y,\xi,d\xi)]\\
&\quad+D_y(\xi)[2(1-\nu)T_3(x,y,\xi,d\xi)-yT_5(x,y,\xi,d\xi)],
\end{aligned}
\tag{7}
$$

and

$$
\begin{aligned}
d\sigma_{xx}&=2GD_x(\xi)[2T_4(x,y,\xi,d\xi)+yT_6(x,y,\xi,d\xi)]\\
&\quad+2GD_y(\xi)[-T_5(x,y,\xi,d\xi)+yT_7(x,y,\xi,d\xi)],\\
d\sigma_{yy}&=2GD_x(\xi)[-yT_6(x,y,\xi,d\xi)]\\
&\quad+2GD_y(\xi)[-T_5(x,y,\xi,d\xi)-yT_7(x,y,\xi,d\xi)],\\
d\sigma_{xy}&=2GD_x(\xi)[-T_5(x,y,\xi,d\xi)\\
&\quad+yT_7(x,y,\xi,d\xi)]+2GD_y(\xi)[-yT_6(x,y,\xi,d\xi)].
\end{aligned}
\tag{8}
$$

Functions T_2 and T_7 in these equations are given by

$$
\begin{aligned}
T_2(x,y,\xi,d\xi)/d\xi&=V_2(x,y,\xi)=-\frac{1}{4\pi(1-\nu)}\frac{x-\xi}{(x-\xi)^2+y^2},\\
T_3(x,y,\xi,d\xi)/d\xi&=V_3(x,y,\xi)=-\frac{1}{4\pi(1-\nu)}\frac{y}{(x-\xi)^2+y^2},\\
T_4(x,y,\xi,d\xi)/d\xi&=V_4(x,y,\xi)=\frac{2y}{4\pi(1-\nu)}\frac{x-\xi}{\{(x-\xi)^2+y^2\}^2},\\
T_5(x,y,\xi,d\xi)/d\xi&=V_5(x,y,\xi)=\frac{1}{4\pi(1-\nu)}\frac{(x-\xi)^2-y^2}{\{(x-\xi)^2+y^2\}^2},\\
T_6(x,y,\xi,d\xi)/d\xi&=V_6(x,y,\xi)\\
&=\frac{2}{4\pi(1-\nu)}\left\{\frac{(x-\xi)^3}{[(x-\xi)^2+y^2]^3}\right.\\
&\quad\left.-\frac{3(x-\xi)y^2}{[(x-\xi)^2+y^2]^3}\right\},\\
T_7(x,y,\xi,d\xi)/d\xi&=V_7(x,y,\xi)\\
&=\frac{2y}{4\pi(1-\nu)}\left\{\frac{3(x-\xi)^2}{[(x-\xi)^2+y^2]^3}\right.\\
&\quad\left.-\frac{y^2}{[(x-\xi)^2+y^2]^3}\right\}.
\end{aligned}
\tag{9}
$$

Obviously, if the following integrals are obtained,

$$U_{ij}(x,y,a)=\int_{-a}^{a}D_j(\xi)V_i(x,y,\xi)d\xi\quad(i=2,3,\ldots,7;\ \ j=1,2)\tag{10}$$

the displacements and stresses at a point (x, y) due to the whole element can be written as

$$
\begin{aligned}
u_x&=[2(1-\nu)U_{3x}(x,y,a)-yU_{5x}(x,y,a)]\\
&\quad+[-(1-2\nu)U_{2y}(x,y,a)-yU_{4y}(x,y,a)],\\
u_y&=[(1-2\nu)U_{2x}(x,y,a)-yU_{4x}(x,y,a)]\\
&\quad+[2(1-\nu)U_{3y}(x,y,a)-yU_{5y}(x,y,a)],
\end{aligned}
\tag{11}
$$

and

$$
\begin{aligned}
\sigma_{xx}&=2G[2U_{4x}(x,y,a)+yU_{6x}(x,y,a)]\\
&\quad+2G[-U_{5y}(x,y,a)+yU_{7y}(x,y,a)],
\end{aligned}
$$

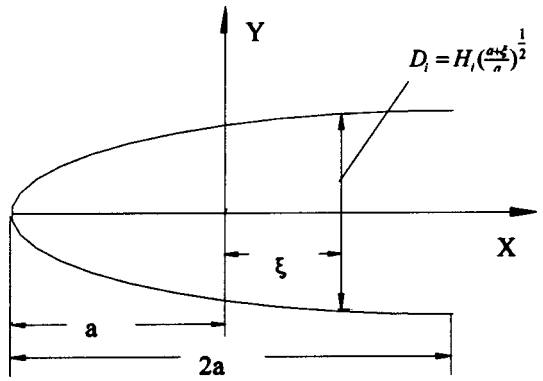


Fig. 4 Schematic of the left crack-tip displacement discontinuity element

$$\begin{aligned}\sigma_{yy}&=2G[-yU_{6x}(x,y,a)]\\&\quad+2G[-U_{5y}(x,y,a)-yU_{7y}(x,y,a)],\\\sigma_{xy}&=2G[-U_{5x}(x,y,a)\\&\quad+yU_{7x}(x,y,a)]+2G[-yU_{6y}(x,y,a)].\end{aligned}\tag{12}$$

The formulas (9)–(12) are the basic formulas required to set up a higher displacement discontinuity element.

2.3 Crack-Tip Displacement Discontinuity Elements. Here, the basic formulas (9)–(12) are used to set up the crack-tip displacement discontinuity elements, which can be classified as the left and the right crack-tip displacement discontinuity elements, to deal with general plane crack problems. The schematic of the left crack-tip displacement discontinuity element is shown in Fig. 4. Its displacement discontinuity functions are chosen as

$$D_x=H_s\left(\frac{a+\xi}{a}\right)^{1/2},\quad D_y=H_n\left(\frac{a+\xi}{a}\right)^{1/2},\tag{13}$$

where H_s and H_n are the tangential and normal displacement discontinuity quantities at the center of the element, respectively. Here, it is noted that the element has the same unknowns as the two-dimensional constant displacement discontinuity element. But it can be seen that the displacement discontinuity functions defined in Eqs. (13) can model the displacement field around the crack tip. The stress field determined by the displacement discontinuity functions (13) possesses $r^{-1/2}$ singularity around the crack tip.

After substituting Eqs. (13) into (10), one has

$$U_{ij}(x,y,a)=H_j\int_{-a}^{a}\left(\frac{a+\xi}{a}\right)^{1/2}V_i(x,y,\xi)d\xi=H_jB_i(x,y,a)\quad(i=2,3,\ldots,7;\ \ j=1,2),\tag{14}$$

where

$$B_i(x,y,a)=\int_{-a}^{a}\left(\frac{a+\xi}{a}\right)^{1/2}V_i(x,y,\xi)d\xi\quad(i=2,3,\ldots,7).\tag{15}$$

After substituting Eq. (14) into Eqs. (11) and (12), one can obtain

$$\begin{aligned}u_x&=H_s[2(1-\nu)B_3(x,y,a)-yB_5(x,y,a)]\\&\quad+H_n[-(1-2\nu)B_2(x,y,a)-yB_4(x,y,a)],\\u_y&=H_s[(1-2\nu)B_2(x,y,a)-yB_4(x,y,a)]\\&\quad+H_n[2(1-\nu)B_3(x,y,a)-yB_5(x,y,a)],\end{aligned}\tag{16}$$

and

$$\begin{aligned}\sigma_{xx}&=2GH_s[2B_4(x,y,a)+yB_6(x,y,a)]\\&\quad+2GH_n[-B_5(x,y,a)+yB_7(x,y,a)],\\\sigma_{yy}&=2GH_s[-yB_6(x,y,a)]\\&\quad+2GH_n[-B_5(x,y,a)-yB_7(x,y,a)],\\\sigma_{xy}&=2GH_s[-B_5(x,y,a)\\&\quad+yB_7(x,y,a)]+2GH_n[-yB_6(x,y,a)].\end{aligned}\tag{17}$$

It can be seen by comparing Eqs. (16) and (17) with Eqs. (2) and (3) that the displacements and stresses due to the crack-tip displacement discontinuity possess the same forms as those due to a constant displacement discontinuity, with $F_l(x,y,a)$ $(i=2,3,\ldots,7)$ in Eqs. (2) and (3) being replaced by $B_i(x,y,a)$ $(i=2,3,\ldots,7)$, and D_x and D_y by H_s and H_n, respectively. This enables the boundary element implementation to be easy.

The computation of B_i $(i=2,3,\ldots,7)$ in Eqs. (16) and (17) will be carried out in the following from four respects.

(1) For an arbitrary point $P(x,y)$ $(y\neq 0)$, generally, the analytical solutions of integrals (15) are obtained difficultly. In this paper, the Gauss numerical integration is used to calculate them. The following transformation is made:

$$\xi=at,\tag{18}$$

and then

$$\begin{aligned}B_i(x,y,a)&=\int_{-a}^{a}\left(\frac{a+\xi}{a}\right)^{1/2}V_i(x,y,\xi)d\xi\\&=a\int_{-1}^{1}V_i(x,y,at)(1+t)^{1/2}dt\quad(i=2,3,\ldots,7).\end{aligned}\tag{19}$$

Therefore, $B_i(x,y,a)$ can be given by

$$B_i(x,y,a)=a\sum_j V_i(x,y,a\zeta_j)(1+\zeta_j)^{1/2}w_j\quad(i=2,3,\ldots,7),\tag{20}$$

where ζ_i and w_i are the Gauss point coordinates and corresponding weighed factors, respectively.

(2) For an arbitrary point $P(x,y)$ $(y=0)$, integrals B_2, B_4, B_5, B_6, and B_7 in Eq. (14) can be solved analytically. For $x>-a$,

$$\begin{aligned}B_2(x,0,a)&=\frac{-1}{4\pi(1-\nu)}\left\{-2\sqrt{2}+\sqrt{\frac{x+a}{a}}\ln\left|\frac{\sqrt{x+a}+\sqrt{2a}}{\sqrt{x+a}-\sqrt{2a}}\right|\right\},\\B_4(x,0,a)&=0,\\B_5(x,0,a)&=\frac{1}{4\pi(1-\nu)}\left\{\frac{\sqrt{2}}{x-a}-\frac{1}{2\sqrt{a(x+a)}}\ln\left|\frac{\sqrt{x+a}+\sqrt{2a}}{\sqrt{x+a}-\sqrt{2a}}\right|\right\},\\B_6(x,0,a)&=\frac{1}{4\pi(1-\nu)}\left\{\frac{\sqrt{2}}{(x-a)^2}-\frac{\sqrt{2}}{2(x^2-a^2)}\right.\\&\quad\left.-\frac{1}{4\sqrt{a}(x+a)^{3/2}}\ln\left|\frac{\sqrt{x+a}+\sqrt{2a}}{\sqrt{x+a}-\sqrt{2a}}\right|\right\},\\B_7(x,0,a)&=0.\end{aligned}\tag{21}$$

While for $x<-a$, let r denote the distance from the crack tip along the crack extension line, i.e.,

$$r=|x|-a.\tag{22}$$

Then

$$B_2(x,0,a)=\frac{-1}{4\pi(1-\nu)}\left\{-2\sqrt{2}+2\sqrt{\frac{r}{a}}\arctan\sqrt{\frac{2a}{r}}\right\},$$

$$B_4(x,0,a)=0,$$

$$B_5(x,0,a)=\frac{1}{4\pi(1-\nu)}\left\{-\frac{\sqrt{2}}{r+2a}+\frac{1}{\sqrt{ar}}\arctan\sqrt{\frac{2a}{r}}\right\},\tag{23}$$

$$B_6(x,0,a)=\frac{1}{4\pi(1-\nu)}\left\{\frac{\sqrt{2}}{(r+2a)^2}-\frac{\sqrt{2}}{2r(r+2a)}-\frac{1}{2\sqrt{a}r^{3/2}}\arctan\sqrt{\frac{2a}{r}}\right\},$$

$$B_7(x,0,a)=0.$$

(3) For an arbitrary point $P(x,y)$ $(y=0)$, the integral B_3 in Eq. (14) is

$$B_3(x,0,a)=\begin{cases}0, & |x|>a\\ +\dfrac{1}{4(1-\nu)}, & y=0_+,\quad |x|<a\\ -\dfrac{1}{4(1-\nu)}, & y=0_-,\quad |x|<a\end{cases}\tag{24}$$

(4) From Eqs. (21) and (24), one can obtain the element self-effects easily:

$$B_2(0,0,a)=\frac{-1}{4\pi(1-\nu)}\left[-2\sqrt{2}+\ln\left|\frac{1+\sqrt{2}}{1-\sqrt{2}}\right|\right],$$

$$B_3(0,0,a)=\begin{cases}+\dfrac{1}{4(1-\nu)}, & y=0_+\\ -\dfrac{1}{4(1-\nu)}, & y=0_-\end{cases}$$

$$B_4(0,0,a)=0,\tag{25}$$

$$B_5(0,0,a)=\frac{1}{4\pi(1-\nu)}\left[-\sqrt{2}-\frac{1}{2}\ln\left|\frac{1+\sqrt{2}}{1-\sqrt{2}}\right|\right]\Big/\ a,$$

$$B_6(0,0,a)=\frac{1}{4\pi(1-\nu)}\left[-\frac{3\sqrt{2}}{2}-\frac{1}{4}\ln\left|\frac{1+\sqrt{2}}{1-\sqrt{2}}\right|\right]\Big/\ a^2,$$

$$B_7(0,0,a)=0.$$

For the right crack-tip displacement discontinuity element, similar formulas can be obtained and do not be given here.

2.4 Implementation of the Present Numerical Approach and Some Illustrations. Crouch and Starfied [20] used Eqs. (2) and (3) to set up the constant displacement discontinuity boundary element method (BEM). Similarly, we can use Eqs. (16) and (17) to set up boundary element equations associated with the crack-tip elements. The constant displacement discontinuity elements presented by Crouch and Starfield [20] together with the crack-tip elements presented in this paper are easily combined to form a very effective numerical approach for calculating the SIFs of general plane cracks. In the boundary element implementation, the left or the right crack-tip element is placed locally at the corresponding left or right crack tip on top of the constant displacement discontinuity elements that cover the entire crack surface and the other boundaries. The method is called as the HDDM.

The hybrid displacement discontinuity method presented in this paper differs from hybrid boundary element codes [27] that, when used to analyze the SIFs of a branched crack, require the plate to be modeled as a finite plate of huge dimensions by fictitious stress elements [20], while the crack could be modeled by displacement discontinuity elements. This brings about a higher computational effort.

Pan [28] pointed out that ''the displacement discontinuity method [20] is quite suitable for cracks in infinite domain where there is no no-crack boundary. However, it alone may not be efficient for finite domain problems, since the kernel functions in DDM involve singularities with order higher than those in the traditional displacement BEM.'' The hybrid displacement discontinuity method is used by the author to calculate the SIFs of complex plane cracks in a finite plate (e.g., a center slant cracked rectangular plate subjected to tension load; cracks emanating from an elliptical hole in a rectangular plate under biaxial loads). These numerical results show that the present numerical approach is also simple, yet very accurate. Because of the limitation to the length of this paper, these results are not reported here.

By the way, it is pointed out here that the displacement discontinuity boundary element program listed in Ref. [20] has one restriction concerning the placement of boundary elements in a problem involving symmetry: a boundary element cannot *lie along* a line of symmetry. Obviously, this restriction means that the symmetric conditions about the *x*-axis and *y*-axis for the crack problems shown in Figs. 7–9 cannot be used and that the symmetric condition about the *x*-axis for the crack problem shown in Fig. 5 cannot be used also. This leads to the result that when the program is used to analyze the crack problems shown in Figs. 5, 7, 8, and 9, it is not much more efficient than the hybrid displacement discontinuity method, which has no such restriction.

Table 1 Variation of SIFs for a center crack in an infinite plate with the number of elements

Number of elements	3	5	7	10	15	25
$K_1/\sigma\sqrt{\pi a}$	0.9621	0.9775	0.9838	0.9885	0.9921	0.995

3 Computational Formulas of Stress Intensity Factors and Simple Test Examples

The objective of many analyses of linear elastic crack problems is to obtain the SIFs K_{I} and K_{II}. Based on the displacement fields around the crack tip, the following formulas exist:

$$K_{\mathrm{I}}=-\frac{G\sqrt{2\pi}}{4(1-\nu)}\lim_{r\to 0}\{D_y(r)/r^{0.5}\},$$

$$K_{\mathrm{II}}=-\frac{G\sqrt{2\pi}}{4(1-\nu)}\lim_{r\to 0}\{D_x(r)/r^{0.5}\},\tag{26}$$

where $D_y(r)$ and $D_x(r)$ are the normal and shear components of displacement discontinuity at a distance r from the crack tip(s). For practical purposes, the limits in Eq. (26) can be approximated by simply evaluating the expression for a fixed value of r that is small in relation to the size of the crack. By means of the crack-tip displacement discontinuity functions defined in Eqs. (13), thus, the approximate formulas of the SIFs K_{I} and K_{II} can be obtained by letting r in Eqs. (26) be a, one-half length of the crack-tip element:

$$K_{\mathrm{I}}=-\frac{\sqrt{2\pi}GH_n}{4(1-\nu)\sqrt{a}},\quad K_{\mathrm{II}}=-\frac{\sqrt{2\pi}GH_s}{4(1-\nu)\sqrt{a}}.\tag{27}$$

To prove the efficiency of the suggested approach, two simple test examples are given here. An infinite plate with a through crack of length $2a$ that is subjected to uniform stress normal to the crack plane at distances sufficiently far away from the crack, is taken to compute the stress intensity factor K_{I}. Owing to its symmetry, only half is taken for the analysis. Table 1 gives the ratio of the numerical solution to the analytical stress intensity factor K_{I} as

Table 2 Variation of SIFs for a center crack in an infinite plate with the ratio of the size of the crack-tip element to that of constant elements

$a_{\text{crack}}/a_{\text{constant}}$	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
$K_1/\sigma\sqrt{\pi a}$	1.2048	1.1690	1.1394	1.1143	1.0928	1.0742	1.0578	1.0433	1.0303
$a_{\text{crack}}/a_{\text{constant}}$	1.05	1.10	1.15	1.20	1.25	1.30	1.35	1.40	1.45
$K_1/\sigma\sqrt{\pi a}$	1.0186	1.0080	0.9984	0.9896	0.9815	0.9741	0.9671	0.9607	0.9547

the number of elements is increased. In this calculation, the crack-tip element and constant elements are taken to be equal in size. Table 2 gives the ratio of the numerical solution to the analytical stress intensity factor K_I as the ratio of the size of the crack-tip element to the one of constant elements is changed. Here, the sizes of the constant elements are taken to be equal and the total number of elements is 11. It can be seen from Table 1 that a good result for the stress intensity factor K_I can be obtained using the crack-tip element. It can be seen from Table 2 that the ratio of the size of the crack-tip element to that of constant elements must be taken to be from 1.0 to 1.3 to obtain a good result with a relative error of less than 3%. This can be regarded as the limitation to the approach presented in the present article.

An inclined crack plate with a through crack of length $2a$ that is subjected to uniform stress at distances sufficiently far away from the crack is taken as another example to compute the SIFs K_I and K_{II}, whose exact solution is available [29]. Here, the SIFs K_I and K_{II} calculated by the present study are normalized by

$$K_I = F_I/(\sigma\sqrt{\pi a}\sin^2\beta), \quad K_{II} = F_{II}/(\sigma\sqrt{\pi a}\sin\beta\cos\beta), \tag{28}$$

where β is the angle between the load and the crack plane. Some numerical results are given in Table 3. In this calculation, the crack-tip elements and constant elements are taken to be the same size and the total number of elements is taken to be 20, i.e., two crack-tip elements and 18 constant elements. It is observed from Table 3 that regardless of the size of the angle β between the load and the crack plane, the present numerical results of the SIFs K_I and K_{II} are in good agreement with the analytical results.

4 Numerical Examples

From the 1970s to today, many researchers have paid attention to branched cracks [1–12], in particular, a singly branched crack. The investigation approaches for these include mostly the Muskhelishvili potential formulation [1,9,30], the conformal mapping method [6–8], the dislocation distribution method [12], and numerical methods, mostly, finite element methods [11,19] and boundary element methods [16,19,31,32]. Here, the present numerical approach is used to calculate the SIFs of branched cracks in an infinite sheet and the present numerical results are compared with the available solutions. Evidently, Bueckner's principle can be used in these analyses.

4.1 A Singly Branched Crack. First, the boundary-element method presented in this article is used to calculate the SIFs of a singly branched crack (see Fig. 1) in an infinite sheet under uniform tension. The SIFs at the branched crack tip B are normalized by

$$F_{IB} = K_{IB}/\sigma\sqrt{\pi c}, \quad F_{IIB} = K_{IIB}/\sigma\sqrt{\pi c} \tag{29}$$

Regarding the discretization of boundary elements, the number of elements discretized on the branched crack is varied with b/a (see Table 4), and the other boundaries are discretized according to the limitation condition that all boundary elements have approximately the same length. For the case of $b/a=0.01$, for example, 2000 elements, which have the same size $2a/2000$, are discretized on the main crack AO, and 10 elements, which have the same size $b/10$, are discretized on the branched crack OB. Here, $2a/2000=b/10$. The element number at the crack tip A is denoted by 1 and the element number at the crack tip B is denoted by 2010. Thus, the element 1 and the element 2010 are the left and the right crack-tip elements, respectively. The elements whose numbers are from 2 to 2009 are all common elements.

Table 5 shows the SIFs at the branched crack tip B obtained in the present article as the branched angle θ and b/a are changed. For comparative purposes, Table 6 lists the analytical results obtained by Kitagawa et al. [6,7] (also see p. 352 in Ref. [29]) by means of the conforming mapping method, whose conjecture has been proven by Lo [9] through the Muskhelishvili potential formulation. It can be seen from Tables 5 and 6 that the present numerical results are in extremely good agreement with those by Kitagawa et al. [6,7].

For a small singly branched crack ($b/a=0.01$), it can be seen by comparing the SIFs given from the present study with those by Kitagawa et al. [6,7] (also see p. 353 in the Ref. [29]) that the agreement is, respectively, within 0.7% and 4% for the stress intensity factors F_{IB} and F_{IIB} (see Table 7).

4.2 A Symmetrically Branched Crack. Second, the present numerical method is used to calculate the SIFs of a symmetrically branched crack (see Fig. 5) in an infinite sheet under uniform tension. The SIFs at the main crack tip A and at the branched crack tip B are normalized by

$$F_I^A = K_{IA}/\sigma\sqrt{\pi c}, \qquad F_I^B = K_{IB}/\sigma\sqrt{\pi c}, \quad F_{II}^B = K_{IIB}/\sigma\sqrt{\pi c}. \tag{30}$$

Regarding the discretization of boundary elements, the number of elements discretized on a branched crack is varied with b/a (see Table 8), and the other boundaries are discretized according to the limitation condition that all boundary elements have approximately the same length. Table 9 shows the present numerical results of the normalized SIFs at the main crack tip A and at the branched crack tip B as the branched angle θ and b/a are changed. For comparative purposes, Table 9 lists also the normalized SIFs given by Kitagawa et al. [6,7] (also see p. 374 in Ref. [29]) by means of the conforming mapping method, whose conjecture has been proven by Lo [9] through the Muskhelishvili potential formulation. It is found from Table 9 that the agreement is, respectively, within 1%, 2.4%, and 3% for the SIFs F_I^A, F_I^B, and F_{II}^B.

4.3 A Skew-Symmetric Branched Crack. Third, the SIFs of a skew-symmetric branched crack in an infinite sheet under

Table 3 Variation of SIFs for an inclined center crack in an infinite plate with the angle β

β	5 deg	10 deg	20 deg	30 deg	40 deg	45 deg	50 deg	60 deg	70 deg	80 deg	85 deg
F_I	0.9895	0.9898	0.9896	0.9898	0.9898	0.9885	0.9897	0.9897	0.9898	0.9897	0.9896
F_{II}	0.9896	0.9897	0.9897	0.9897	0.9897	0.9885	0.9897	0.9897	0.9897	0.9897	0.9896

Table 4 Variation of the number of elements discretized on a branched crack for a singly branched crack with b/a

b/a										
0.01	0.05	0.1	0.2	0.4	0.5	0.6	0.8	1.0	1.5	2.0
10	10	15	30	30	30	30	30	30	45	60

uniform tension (see Fig. 6) are analyzed by means of the present numerical approach. The SIFs at a branched crack tip are normalized by

$$F_{\mathrm{I}}=K_{\mathrm{I}}/\sigma\sqrt{\pi c},\quad F_{\mathrm{II}}=K_{\mathrm{II}}/\sigma\sqrt{\pi c}. \tag{31}$$

Regarding the discretization of boundary elements, the number of elements discretized on a branched crack is varied with b/a (see also Table 8), and the other boundaries are discretized according to the limitation condition that all boundary elements have approximately the same length. Table 10 shows the SIFs at the branched crack tip given by the present study and by Kitagawa et al. [6,7] (also see p. 362 in Ref. [29]) as the branched angle θ and b/a are changed. Evidently, the present numerical results are in extremely good agreement with those obtained by Kitagawa et al. [6,7] by using the conforming mapping method.

4.4 A Doubly Symmetrically Branched Crack. Finally, the SIFs of a doubly symmetrically branched crack in an infinite sheet under uniform tension (see Fig. 7) are analyzed by means of the present numerical approach. The SIFs at a branched crack tip are determined still by using formulas (31). Regarding the discretization of boundary elements, the number of elements discretized on a branched crack is varied with b/a (see also Table 8), and the other boundaries are discretized according to the limitation condition that all boundary elements have approximately the same length. Table 11 shows the SIFs at a branched crack tip as the branched angle θ and b/a are changed. For the comparison purpose, Table 11 lists also the SIFs at the branched crack tip obtained by Vitek [12] (also see p. 386 in Ref. [29] by means of the dislocation distribution method. It can be seen from Table 11 that for the doubly symmetric branched crack the present numerical results are in good agreement with those by Vitek within 2.2%.

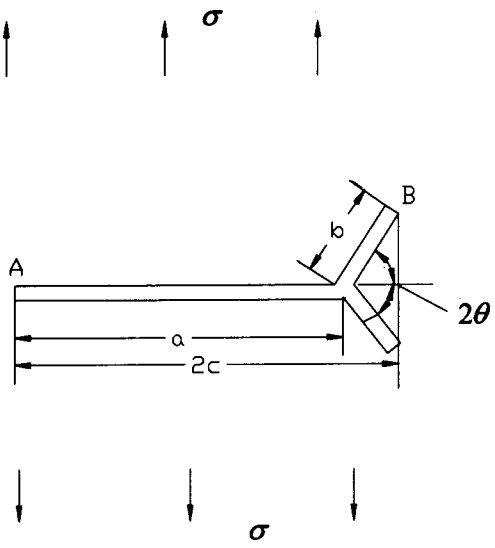


Fig. 5 A symmetrically branched crack

Table 5 Normalized SIFs at the branched crack tip B for a singly branched crack in the present study

	15 deg		30 deg		45 deg		60 deg	
b/a	$F_{\mathrm{I}B}$	$F_{\mathrm{II}B}$	$F_{\mathrm{I}B}$	$F_{\mathrm{II}B}$	$F_{\mathrm{I}B}$	$F_{\mathrm{II}B}$	$F_{\mathrm{I}B}$	$F_{\mathrm{II}B}$
0.01	0.9654	0.1614	0.8709	0.2995	0.7309	0.3964	0.5665	0.4422
0.05	0.9578	0.1938	0.8410	0.3572	0.6693	0.4667	0.4710	0.5105
0.1	0.9530	0.2131	0.8234	0.3918	0.6332	0.5096	0.4140	0.5527
0.2	0.9487	0.2349	0.8065	0.4314	0.5978	0.5595	0.3573	0.6037
0.4	0.9448	0.2554	0.7938	0.4689	0.5725	0.6076	0.3185	0.6539
0.5	0.9439	0.2607	0.7915	0.4786	0.5682	0.6197	0.3127	0.6655
0.6	0.9433	0.2644	0.7902	0.4852	0.5661	0.6277	0.3107	0.6724
0.8	0.9427	0.2688	0.7894	0.4929	0.5655	0.6363	0.3120	0.6775
1.0	0.9425	0.2710	0.7896	0.4966	0.5667	0.6395	0.3158	0.6766
1.5	0.9438	0.2729	0.7926	0.4992	0.5717	0.6388	0.3260	0.6668
2.0	0.9448	0.2726	0.7950	0.4981	0.5761	0.6351	0.3330	0.6563

Table 6 Normalized SIFs at the crack tip B for a singly branched crack by Kitagawa et al. [6,7]

	15 deg		30 deg		45 deg		60 deg	
b/a	$F_{\mathrm{I}B}$	$F_{\mathrm{II}B}$	$F_{\mathrm{I}B}$	$F_{\mathrm{II}B}$	$F_{\mathrm{I}B}$	$F_{\mathrm{II}B}$	$F_{\mathrm{I}B}$	$F_{\mathrm{II}B}$
0.1	0.9540	0.2120	0.8245	0.3895	0.6339	0.5053	0.4106	0.5462
0.2	0.9496	0.2346	0.8076	0.4307	0.5983	0.5578	0.3583	0.5996
0.4	0.9466	0.2556	0.7957	0.4690	0.5741	0.6072	0.3189	0.6514
0.6	0.9457	0.2648	0.7927	0.4858	0.5679	0.6283	0.3112	0.6718
0.8	0.9456	0.2694	0.7922	0.4940	0.5678	0.6375	0.3128	0.6770
1.0	0.9457	0.2718	0.7928	0.4981	0.5694	0.6413	0.3171	0.6775
1.5	0.9463	0.2737	0.7951	0.5008	0.5744	0.6414	0.3273	0.6682
2.0	0.9468	0.2733	0.7971	0.4996	0.5785	0.6377	0.3340	0.6580

Table 7 Normalized SIFs at the small branch crack tip B for a singly branched crack ($b/a=0.01$)

	$F_{\mathrm{I}B}$		$F_{\mathrm{II}B}$	
θ (deg)	Refs. [6], [7]	Present	Refs. [6], [7]	Present
15	0.971	0.9654	0.156	0.1614
30	0.876	0.8709	0.296	0.2995
45	0.732	0.7309	0.389	0.3964
60	0.569	0.5665	0.431	0.4422
75	0.404	0.4000	0.420	0.4368

5 Cracks Emanating From a Rhombus Hole in an Infinite Plate Under Biaxial Loads

In this section, specifically, the boundary element method presented in this article is used to study cracks emanating from a rhombus hole in an infinite plate under biaxial loads. The present numerical results for this crack problem indicate further that the present approach is very effective for calculating the SIFs of complex plane cracks and can reveal the effect of the biaxial load on the SIFs.

Shown in Fig. 8 is the schematic of cracks emanating from a rhombus hole in an infinite plate under biaxial loads. For this problem, symmetric conditions about the x-axis and y-axis can be used. The following cases are considered:

$$\lambda=0,1,-1$$

$$\theta=15\ \text{deg},\ 30\ \text{deg},\ 45\ \text{deg},$$

$$a/b=1.05,\ 1.1,\ 1.2,\ 1.4,\ 1.6,\ 1.8,\ 2.0,\ 2.5,\ 3.0,\ 4.0.$$

Regarding the discretization of boundary elements, the number of elements discretized on a quarter of rhombus hole is 100, and the other boundaries are discretized according to the limitation condition that all boundary elements have approximately the same

Table 8 Variation of the number of elements discretized on a branched crack for a symmetrically branched crack with b/a

b/a										
0.01	0.05	0.1	0.2	0.4	0.5	0.6	0.8	1.0	1.5	2.0
10	10	15	30	30	30	30	30	30	45	60

Table 9 Normalized SIFs for a symmetrically branched crack

		60 deg		45 deg		30 deg	
b/a		Refs. [6], [7]	Present	Refs. [6], [7]	Present	Refs. [6], [7]	Present
0.01	F_{I}^{A}		1.0001		0.9998		0.9994
	F_{I}^{B}		0.5390		0.6499		0.7138
	F_{II}^{B}		0.3158		0.2223		0.0913
0.05	F_{I}^{A}	1.01	1.0054	1.01	1.0043	1.01	1.0027
	F_{I}^{B}	0.45	0.4450	0.59	0.5918	0.70	0.6888
	F_{II}^{B}	0.39	0.3889	0.29	0.2994	0.15	0.1575
0.1	F_{I}^{A}	1.01	1.0097	1.01	1.0080	1.01	1.0057
	F_{I}^{B}	0.39	0.3899	0.56	0.5588	0.68	0.6764
	F_{II}^{B}	0.43	0.4362	0.34	0.3479	0.19	0.1985
0.2	F_{I}^{A}	1.02	1.0159	1.02	1.0135	1.01	1.0105
	F_{I}^{B}	0.34	0.3349	0.54	0.5268	0.66	0.6651
	F_{II}^{B}	0.49	0.4923	0.40	0.4042	0.24	0.2464
0.4	F_{I}^{A}	1.03	1.0248	1.02	1.0199	1.02	1.0155
	F_{I}^{B}	0.29	0.2944	0.51	0.5029	0.65	0.6564
	F_{II}^{B}	0.55	0.5451	0.46	0.4583	0.28	0.2938
0.5	F_{I}^{A}		1.0298		1.0227		1.0175
	F_{I}^{B}		0.2869		0.4982		0.6551
	F_{II}^{B}		0.5579		0.4728		0.3074
0.6	F_{I}^{A}	1.04	1.0355	1.03	1.0257	1.02	1.0192
	F_{I}^{B}	0.28	0.2829	0.50	0.4956	0.65	0.6545
	F_{II}^{B}	0.57	0.5658	0.49	0.4830	0.32	0.3175
0.8	F_{I}^{A}	1.05	1.0487	1.04	1.0320	1.03	1.0225
	F_{I}^{B}	0.28	0.2801	0.50	0.4933	0.65	0.6544
	F_{II}^{B}	0.58	0.5731	0.50	0.4957	0.33	0.3313
1.0	F_{I}^{A}	1.07	1.0632	1.04	1.0387	1.03	1.0255
	F_{I}^{B}	0.28	0.2802	0.50	0.4927	0.66	0.6548
	F_{II}^{B}	0.58	0.5742	0.51	0.5028	0.34	0.3402

length. The present numerical results of the SIFs normalized by $\sigma\sqrt{\pi a}$ are given in Table 12. For purposes of comparison, Table 12 also lists the numerical results in Ref. [29]. From Table 12, it is found that the present numerical results are in excellent agreement with those reported in Ref. [29]. From Table 12, it is found that the effect of the load parameter λ on the SIFs varies with θ and a/b and that the effect of the rhombus angle θ on the SIFs varies with load parameter λ and a/b.

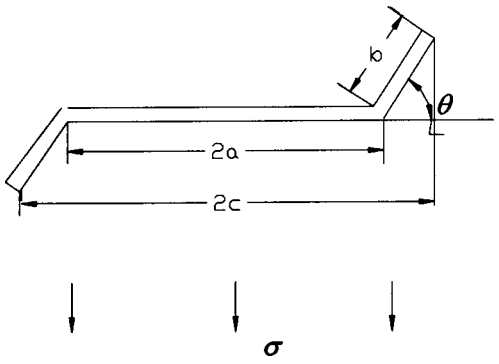


Fig. 6 A skew-symmetric branched crack

6 Numerical Simulation of Fatigue Crack Propagation Process Under Mixed-Mode Conditions

On the application of boundary element methods to crack propagation analysis, the first attempt to automatically model crack propagation under mixed-mode conditions was given by Ingraffea, Blanford, and Liggett [22] for two-dimensional problems. They used the multiregion method [13] together with the maximum circumferential stress criterion to calculate the direction of crack propagation. Aliabadi [23] pointed out that the difficulty with the multiregion method is that the introduction of artificial boundaries to divide the regions is not unique, and thus it is not easy to implement it in an automatic procedure. In an incremental crack extension analysis, these artificial boundaries must be repeatedly introduced for each increment of crack extension. Finite element simulations [25,26] when used to analyze crack problems

Table 10 Normalized SIFs at a branched crack tip for a skew-symmetric branched crack

	60 deg				45 deg				30 deg			
	Refs. [6], [7]		Present		Refs. [6], [7]		Present		Refs. [6], [7]		Present	
b/a	F_{I}	F_{II}	F_{I}	F_{II}	F_{I}	F_{II}	F_{I}	F_{II}	F_{I}	F_{II}	F_{I}	F_{II}
0.01	0.5905	0.4120	0.5936	0.4243	0.7485	0.3686	0.7485	0.3772	0.8809	0.2780	0.8792	0.2834
0.05	0.5232	0.4610	0.5274	0.4731	0.7058	0.4184	0.7069	0.4268	0.8603	0.3182	0.8600	0.3236
0.1	0.4822	0.4920	0.4853	0.5022	0.6805	0.4507	0.6811	0.4560	0.8483	0.3436	0.8479	0.3469
0.2	0.4306	0.5350	0.4384	0.5391	0.6532	0.4888	0.6532	0.4913	0.8356	0.3727	0.8351	0.3741
0.4	0.3934	0.5794	0.3944	0.5830	0.6280	0.5284	0.6273	0.5303	0.8242	0.4021	0.8231	0.4032
0.6	0.3734	0.6031	0.3737	0.6062	0.6161	0.5491	0.6150	0.5508	0.8187	0.4176	0.8172	0.4184
0.8	0.3629	0.6170	0.3626	0.6197	0.6095	0.5617	0.6080	0.5630	0.8156	0.4271	0.8137	0.4277
1.0	0.3570	0.6253	0.3564	0.6278	0.6054	0.5698	0.6036	0.5709	0.8135	0.4335	0.8114	0.4340

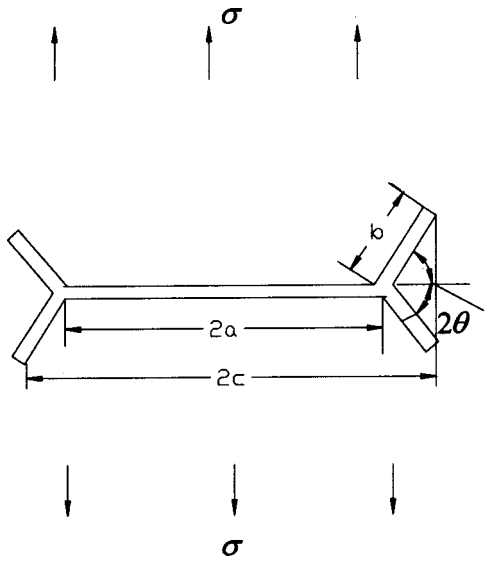


Fig. 7 A doubly symmetrically branched crack

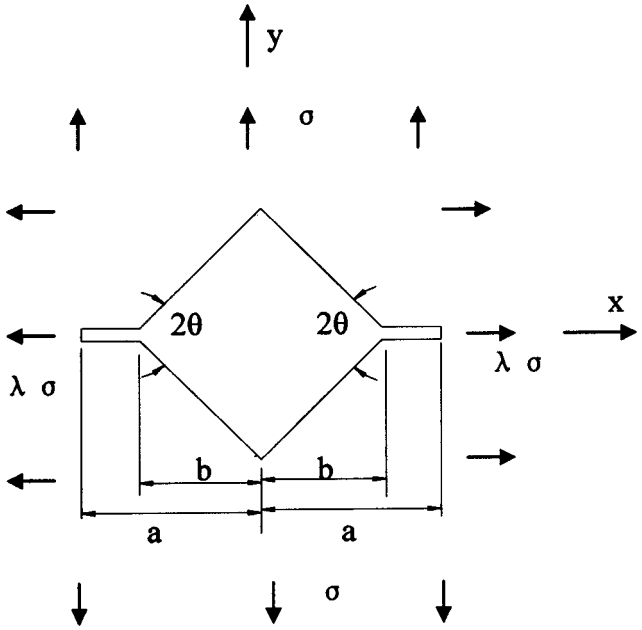


Fig. 8 Cracks emanating from a rhombus hole in an infinite plate under biaxial loads

have to face difficult computational problems connected with the discretization of the continuum into finite elements, particularly when some cracks propagate, thus changing the interior boundaries of the solids.

In this paper, the hybrid displacement discontinuity method and the maximum circumferential stress criterion [24] are combined to form a very effective numerical approach for simulating the fatigue crack propagation process in plane elastic bodies under mixed-mode conditions. In the numerical simulation, for each increment of crack extension, remeshing of existing boundaries is not required because of an intrinsic feature of the HDDM. Crack propagation is simulated by adding new boundary elements on the incremental crack extension to the previous crack boundaries. At the same time, the element characters of some related elements are adjusted according to the manner in which the boundary element method is implemented. As an example, the fatigue propagation process of cracks emanating from a circular hole in a plane elastic plate is simulated using the numerical simulation approach.

It is well known that the fatigue crack propagation, which propagates in a self-similar manner, obeys Paris's equation

$$\Delta a/\Delta N = A(\Delta K_{\mathrm{I}})^m, \tag{32}$$

where $\Delta a/\Delta N$ is the fatigue crack propagation ratio, A and m are material constants, and ΔK_{I} is the range of the stress intensity factor K_{I}.

In general, the fatigue propagation analysis of a crack under mixed-mode conditions involves the determination of the crack propagation direction (e.g., Sih and Barthelemy [33]) except for an equation similar to Eq. (32). Here, the fatigue propagation of a crack under mixed-mode conditions is assumed to satisfy the following conditions:

(1) A fatigue crack will propagate to the direction of the maximum tangential stress surrounding the crack tip.
(2) The fatigue crack propagation rate equation is

$$\frac{\Delta a}{\Delta N} = C(\Delta K_e)^n, \tag{33}$$

where

$$\Delta K_e = \tfrac{1}{2}\cos\frac{\theta_0}{2}[\Delta K_{\mathrm{I}}(1+\cos\theta_0) - 3\Delta K_{\mathrm{II}}\sin\theta_0], \tag{34}$$

where θ_0 is the crack propagation angle predicted according to condition (1), ΔK_{I} and ΔK_{II} are, respectively, the ranges of the stress intensity factors K_{I} and K_{II}, and C and n are material constants that are related to the material constants A and m in Paris's equation (32) by the relations

$$C = A, \quad n = m. \tag{35}$$

As an example, the fatigue propagation process of cracks emanating from a circular hole (see Fig. 9) in an infinite plate sub-

Table 11 Normalized SIFs for a doubly symmetric branched crack

	60 deg				45 deg				30 deg			
	F_{I}		F_{II}		F_{I}		F_{II}		F_{I}		F_{II}	
b/a	Vitek	Present	Vitek	Present	Vitek	Present	Vitek	Present	Vitek	Present	Vitek	Present
0.01	0.557	0.5664	0.292	0.2980	0.662	0.6685	0.200	0.2029	0.724	0.7240	0.073	0.0742
0.05	0.491	0.5014	0.350	0.3565	0.627	0.6348	0.262	0.2646	0.718	0.7188	0.125	0.1258
0.1	0.452	0.4615	0.394	0.3984	0.611	0.6184	0.307	0.3081	0.723	0.7240	0.162	0.1611
0.2	0.410	0.4184	0.454	0.4576	0.600	0.6074	0.368	0.3686	0.740	0.7419	0.211	0.2096
0.4		0.3830		0.5371		0.6127		0.4502		0.7839		0.2745
0.5	0.370	0.3764	0.562	0.5665	0.616	0.6208	0.481	0.4813	0.807	0.8060	0.301	0.2993
0.6		0.3736		0.5916		0.6307		0.5085		0.8282		0.3212
0.8		0.3749		0.6332		0.6532		0.5551		0.8719		0.3589
1.0	0.377	0.3810	0.663	0.6673	0.676	0.6774	0.591	0.5946	0.919	0.9144	0.393	0.3911

Table 12 Normalized SIFs for cracks emanating from a rhombus hole in an infinite plate under biaxial loads

	$\lambda=0$				$\lambda=-1$			$\lambda=+1$		
a/b	$\theta=15$ deg	$\theta=30$ deg	$\theta=45$ deg	$\theta=45$ deg [29]	$\theta=15$ deg	$\theta=30$ deg	$\theta=45$ deg	$\theta=15$ deg	$\theta=30$ deg	$\theta=45$ deg
1.05	1.0146	1.0389	1.0433		1.0313	1.1208	1.2461	0.9979	0.9570	0.8405
1.1	1.0105	1.0372	1.0602	1.07	1.0237	1.1072	1.2487	0.9973	0.9672	0.8717
1.2	1.0066	1.0305	1.0639	1.069	1.0157	1.0834	1.2229	0.9975	0.9776	0.9049
1.4	1.0031	1.0202	1.0535	1.058	1.0081	1.0530	1.1667	0.9981	0.9874	0.9403
1.6	1.0015	1.0137	1.0423	1.046	1.0046	1.0354	1.1244	0.9984	0.9920	0.9602
1.8	1.0005	1.0095	1.0333	1.037	1.0025	1.0244	1.0941	0.9985	0.9946	0.9725
2.0	1.0000	1.0067	1.0264	1.030	1.0013	1.0173	1.0723	0.9987	0.9961	0.9805
2.5	0.9993	1.0028	1.0150		0.9998	1.0078	1.0393	0.9988	0.9978	0.9907
3.0	0.9989	1.0009	1.0087		0.9991	1.0034	1.0225	0.9987	0.9984	0.9949
4.0	0.9986	0.9993	1.0027		0.9986	1.0000	1.0079	0.9986	0.9986	0.9975

jected to uniform cycle stress in the y direction at distances sufficiently far away from the hole is simulated. For this crack problem, the symmetric conditions about the x and y-axes can be used. In this analysis, the shear modulus G, Poisson's ratio ν, the fracture toughness K_{Ic}, the material constants A and m in Paris's equation, the threshold value of the stress intensity factor, ΔK_{th}, cyclic loading parameters, the mean stress σ_m, and the characteristic of cyclic loading, R, are as follows:

$$G=2744 \text{ kg/mm}^2, \quad \nu=0.321, \quad K_{Ic}=116 \text{ kg/mm}^{3/2},$$

$$A=1.039\times 10^{-10},$$

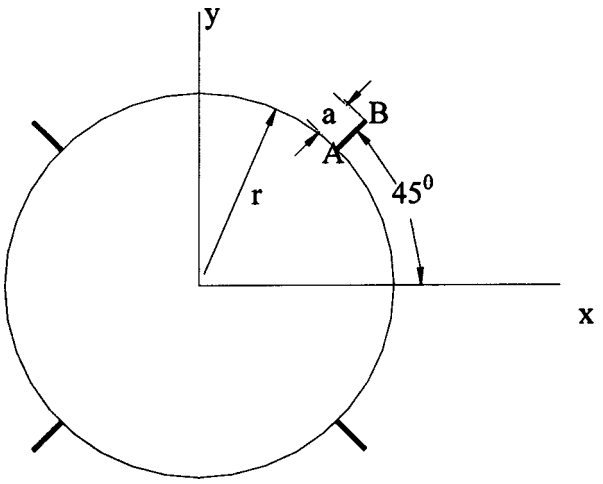


Fig. 9 Cracks emanating from a circular hole in an infinite plate

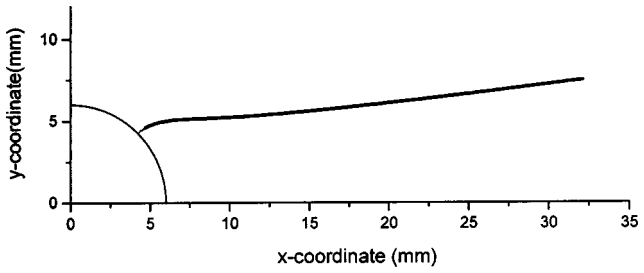


Fig. 10 Fatigue growth path of crack *AB* emanating from the circle hole for the case of $r=6$ mm

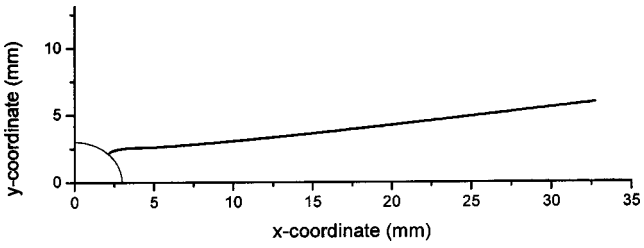


Fig. 11 Fatigue growth path of crack *AB* emanating from the circle hole for the case of $r=3$ mm

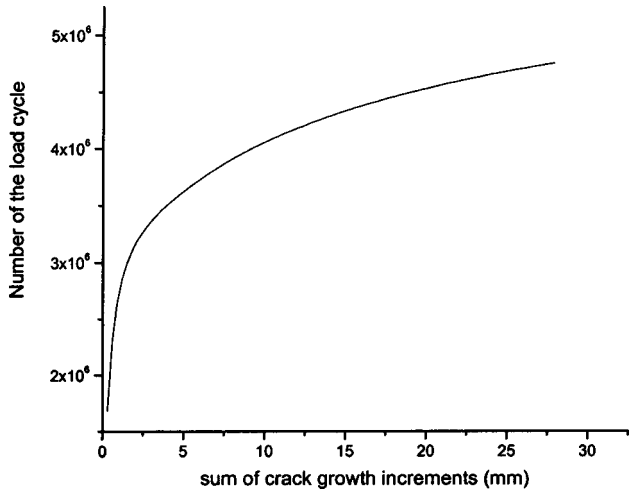


Fig. 12 Variation of the number of load cycle with crack growth for the case of $r=6$ mm

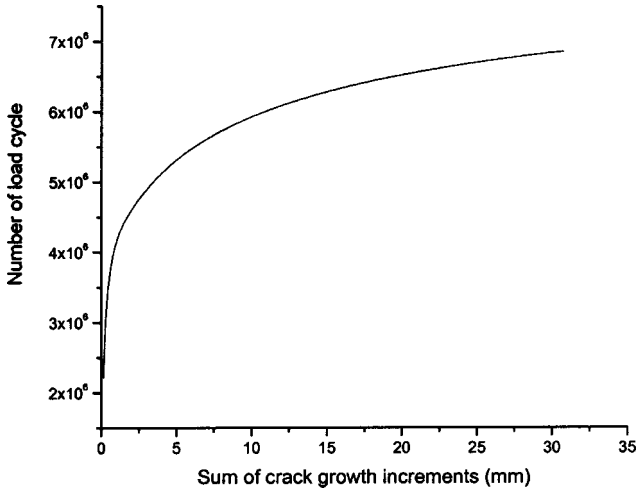


Fig. 13 Variation of the number of load cycle with crack growth for the case of $r=3$ mm

$m = 2.7438, \quad \Delta K_{th} = 0, \quad \sigma_m = 15.333 \ \text{kg/mm}^2, \quad R = 0.048,$

The geometric parameters considered here are

$$R = 3 \ \text{mm}, \quad a/r = 1/20.$$

$$r = 6 \ \text{mm}, \quad a/r = 1/20.$$

Some numerical results are given in Figs. 10–13. Shown in Figs. 10 and 11 are, respectively, the fatigue propagation path of crack *AB* emanating from the circle hole for the cases of $r = 6$ and 3 mm. Shown in Figs. 12 and 13 are, respectively, the variation of the number of the load cycle with the crack propagation for the cases of $r = 6$ and 3 mm.

7 Conclusions

In the present paper, the crack-tip displacement discontinuity elements are presented to model the singularity of stress near a crack tip. Furthermore, the crack-tip elements and the constant displacement discontinuity element presented by Crouch and Starfied are combined to form a numerical approach for calculating the SIFs of general plane cracks. Numerical examples are given and compared with the available solutions. It can be seen that the numerical approach is simple, yet very accurate for calculating the SIFs of branched cracks. As a new example, cracks emanating from a rhombus hole in an infinite plate under biaxial loads are taken into account. The numerical results indicate the efficiency of the present numerical approach and can reveal the effect of the biaxial load on the SIFs. In addition, the hybrid displacement discontinuity method developed in this paper together with the maximum circumferential stress criterion becomes a very effective numerical approach for simulating the fatigue crack propagation process in plane elastic bodies under mixed-mode conditions.

Acknowledgment

Special thanks are due to the National Natural Science Foundation of China (No. 10272037) and the Natural Science Foundation of Heilongjiang, China (No. A-02-05) for supporting the present work.

References

[1] Cotterel, B., and Rice, J. R., 1980, "Slightly Curved or Kinked Cracks," Int. J. Fract., **16**, pp. 155–169.
[2] Anderdon, H., and Erratum, W. G., 1969, "Stress Intensity Factors at the Tips of a Star-Shape Contour in an Infinite Tensile Sheet," J. Mech. Phys. Solids, **17**, pp. 405–417.
[3] Palaniswamy, K., and Knauss, W. G., "On the Problem of Crack Extension in Brittle Solids under General Loading," Report No. SM74-8, Graduate Aeronautical Lab., Caltech.
[4] Billy, B. A., and Cardew, G. E., 1975, "The Crack with a Kinked Tip," Int. J. Fract., **11**, pp. 708–711.
[5] Billy, B. A., Cardew, G. E., and Howard, I. C., 1977, "Stress Intensity Factors at the Tips of Kinked and Forked Cracks," in *Fracture 1977*, Taplin, D. M. R., ed., University of Waterloo Press, Canada, Vol. 3, pp. 197–200.
[6] Kitagawa, H., and Yuuki, R., 1977, "Analysis of Branched Cracks Under Biaxial Stresses," in *Fracture 1977*, Taplin D. M. R., eds., University of Waterloo Press, Canada, Vol. 3, pp. 201–211.
[7] Kitagawa, H., Yuuki, R., and Ohira, T. C., 1975, "Crack-Morphological Aspects in Fracture Mechanics," Eng. Fract. Mech., **7**, pp. 515–529.
[8] Chatterjee, S. N., 1975, "The Stress Field in the Neighborhood of a Branched Crack in an Infinite Sheet," Int. J. Solids Struct., **11**, pp. 521–538.
[9] Lo, K. K., 1978, "Analysis of Branched Cracks," ASME J. Appl. Mech., **45**, pp. 797–802.
[10] Theocaris, P. S., and Loakimidis, N., 1976, "The Symmetrically Branched Cracks in an Infinite Elastic Medium," ZAMP, **27**, pp. 801–814.
[11] Wilson, W. K., and Cherepko, J., 1983, "Analysis of Cracks With Multiple Branches," Int. J. Fract., **22**, pp. 302–315.
[12] Vitek, V., 1977, "Plane Strain Stress Intensity Factors for Branched Cracks," Int. J. Fract., **13-4**, pp. 481–510.
[13] Blandford, G. E., Ingraffea, A. R., and Liggett, J. A., 1981, "Two-Dimensional Stress Intensity Factor Computations Using the Boundary Element Method," Int. J. Numer. Methods Eng., **17**, pp. 387–404.
[14] Balas, J., Sladek, J., and Sladek, V., 1989, *Stress Analysis by Boundary Element Methods*, Elsevier, Amsterdam.
[15] Hong, H., and Chen, J., 1988, "Derivatives of Integral Equations of Elasticity," J. Eng. Mech., **114**, pp. 1028–1044.
[16] Portela, A., and Aliabadi, M. H., 1992, "The Dual Boundary Element Method: Effective Implementation for Crack Problems," Int. J. Numer. Methods Eng., **33**, pp. 1269–1287.
[17] Tanaka, M., and Itoh, H., 1987, "New Crack Elements for Boundary Element Analysis of Elastostatics Considering Arbitrary Stress Singularities," Appl. Math. Model., **11**, pp. 357–363.
[18] Cruse, T. A., 1989, *Boundary Element Analysis in Computational Fracture Mechanics*, Kluwer, Dordrecht.
[19] Aliabadi, M. H., and Rooke, D. P., 1991, *Numerical Fracture Mechanics*, Computational Mechanics Publications, Southampton and Kluwer, Dordrecht.
[20] Crouch, S. L., and Starfield, A. M., 1983, *Boundary Element Method in Solid Mechanics with Application in Rock Mechanics and Geological Mechanics*, Allen and Unwin, London.
[21] Crouch, S. L., 1976, "Solution of Plane Elasticity Problems by Displacement Discontinuity Method," Int. J. Numer. Methods Eng., **10**, pp. 301–343.
[22] Ingraffea, A. R., Blandford, G., and Liggett, J. A., 1987, "Automatic Modeling of Mixed-Mode Fatigue and Quasi-Static Crack Propagation Using the Boundary Element Method," 14th Natl. Symp. on Fracture, ASTM STP 791, pp. 1407–1426.
[23] Aliabadi, M. H., 1997, "Boundary Element Formulation in Fracture Mechanics," Appl. Mech. Rev., **50**, pp. 83–96.
[24] Erdogan, F., and Sih, G. C., 1963, "On the Crack Extension in Plates Under Plane Loading and Transverse Shear," J. Basic Eng., **85**, pp. 519–527.
[25] Charambides, P. G., and McMeeking, R. M., 1987, "Finite Element Method Simulation of a Crack Propagation in a Brittle Microcracked Solid," Mech. Mater., **6**, pp. 71–87.
[26] Huang, X., and Karihaloo, B. L., 1993, "Interaction of Penny Shaped Cracks With a Half Plane Crack," Int. J. Solids Struct., **25**, pp. 591–607.
[27] Scavia, C., 1992, "A Numerical Technique for the Analysis of Cracks Subjected to Normal Compressive Stresses," Int. J. Numer. Methods Eng., **33**, pp. 929–942.
[28] Pan, E., 1997, "A General Boundary Element Analysis of 2-D Linear Elastic Fracture Mechanics," Int. J. Fract., **88**, pp. 41–59.
[29] Murakami, Y., 1987, *Stress Intensity Factors Handbook*, Pergamon, New York.
[30] Chen, Y. Z., 1999, "Stress Intensity Factors for Curved and Kinked Cracks in Plane Extension," Theor. Appl. Fract. Mech., **31**, pp. 223–232.
[31] Liu, N., Altiero, N. J., and Sur, U., 1990, "An Alternative Integral Equation Approach Applied to Kinked Cracks in Finite Plane Bodies," Comput. Methods Appl. Mech. Eng., **84**, pp. 211–226.
[32] Denda, M., and Dong, Y. F., 1999, "Analytical Formulas for a 2-D Crack-Tip Singular Boundary Element for Rectilinear Cracks and Crack Growth Analysis," Eng. Anal. Boundary Elem., **23**, pp. 35–49.
[33] Sih, G. C., and Barthelemy, B. M., 1980, "Mixed Mode Fatigue Crack Growth Prediction," Eng. Fract. Mech., **13**, pp. 439–451.

Analytical Solution for Shear Horizontal Wave Propagation in Piezoelectric Coupled Media by Interdigital Transducer

Q. Wang
Mechanical, Materials, and Aerospace Engineering Department, University of Central Florida, Orlando, FL 32816
Member ASME

S. T. Quek
Department of Civil Engineering, National University of Singapore, Singapore 119260

V. K. Varadan
Department of Engineering Science and Mechanics, The Pennsylvania State University, University Park, PA 16802
Fellow ASME

An analytical solution for the shear horizontal wave propagation excited by interdigital transducer in a piezoelectric coupled semi-infinite medium is developed. This solution is an extension of earlier work on wave propagation in a piezoelectric coupled plate with finitely long interdigital transducer by fully taking account of piezoelectric effects in analysis. In the current analysis, the mathematical model for a semi-infinite metal substrate bonded by a layer of interdigital transducer with infinite length is first derived. The theoretical solutions are obtained in terms of elliptic integration of the first kind and of the standard integral representation for Legendre polynomial. The essential hypothesis for the derivation of the analysis is investigated. Based on the solution for infinitely long interdigital transducer, an analytical solution for the wave propagation in this semi-infinite piezoelectric medium excited by a finitely long interdigital transducer is obtained through Fourier transform. This theoretical research can be applied to health monitoring of structures by interdigital transducer. It could also be used as a framework for the design of interdigital transducer in wave excitation of smart structures.
[DOI: 10.1115/1.1876412]

1 Introduction

An interdigital transducer wafer comprises a thin piezoelectric film on which electrodes in the form of two alternating set of fingers are deposited, as shown in Fig. 1. Such wafer is surface bonded onto a substrate or structure for excitation or reception of waves. In practice, an interdigital transducer is finite in length.

Interdigital transducer was first used in surface acoustic wave devices, which can be found in radar communication equipment as filters and delay lines [1,2], and consumer products such as pagers, mobile phones, and sensors [3–5]. Interdigital transducer has also been used for separating, amplifying, and storing signals as well as signal processing in acousto-electronics [6–8]. Great potentials have been found in using interdigital transducer as sensors for various physical variables, such as force, electric fields, magnetic fields, temperature, and pressure [1]. Interdigital transducer is nowadays used in the area of structural health monitoring due to its controllability of the excited waves and its convenience in operation. Researches and experimental works using interdigital transducer to excite Lamb wave for rapid monitoring of structures have also been attempted [9,10]. The key issue for application of interdigital transducer in structural health monitoring is how to design the size of the interdigital transducer, such as its wavelength and finger width, so that a wave signal with higher magnitude and less dispersive effect can be excited and sensed. Therefore, a complete mechanics analysis for the effect of the interdigital transducer on the wave solution is important and essential for the design of interdigital transducer in its application of structural health monitoring.

The availability of good analytical and design methods for interdigital transducer is essential for all the above applications. Although there has been considerable literature on the analysis of interdigital transducer, improvements can still be made. One major difficulty of the analysis lies in accounting for the full electromechanical coupling in the structure. Tseng [11], Coquin and Tierstan [12], and Joshin and White [13] studied the analysis of the interdigital transducer by solving an electrostatic problem, substituted the distribution of the electric fields into the electromechanical coupled equation, and hence obtained the secondary electric fields and the displacement fields. In the monograph of Parton and Budryavster [8], the analytical solutions for an interdigital transducer which generate Rayleigh surface waves in a hexagonal 6 mm piezoelectric medium was presented based on the same procedure. Another popular analysis was proposed by Balakrirev and Gilinskii [14] to use Green matrix method to solve a two-dimensional (2D) problem for a half-unbounded crystal. In practice, the method is hardly feasible for an arbitrary crystal due to the difficulties in constructing the Green matrix. Some recent progresses are contributed through the finite element method [15–17], boundary element method [18], and 2D's Green function [19]. However, the behavior of surface acoustic wave by interdigital transducer still cannot be accurately modeled analytically and predicted. Kino [20] looked into the theory of excitation of surface acoustic waves on a nonpiezoelectric material by using interdigital transducer. The results were given in terms of the perturbation in acoustic wave velocity. Engan [21] presented the electrostatic field with an infinite number of space harmonics with relative amplitudes given by the corresponding Legendre polynomials. Ogilvy [22] presented an approximate analysis for predicting the generation of elastic waves by interdigital transducer in multilayered piezoelectric materials. However, the width of the interdigital transducer fingers is not explicitly taken into account and neither are the finite dimensions of the interdigital transducer. Similarly, in the analysis of interdigital transducer in health monitoring of structures, the electromechanical effects are neglected in the design of the structure. Hence, the wave characteristics are still solely based on the metal substrate materials. Wang and Varadan [23,24] provided an analytical mathematical solution for

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, February 19, 2003; final revision, September 1, 2004. Associate Editor: K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

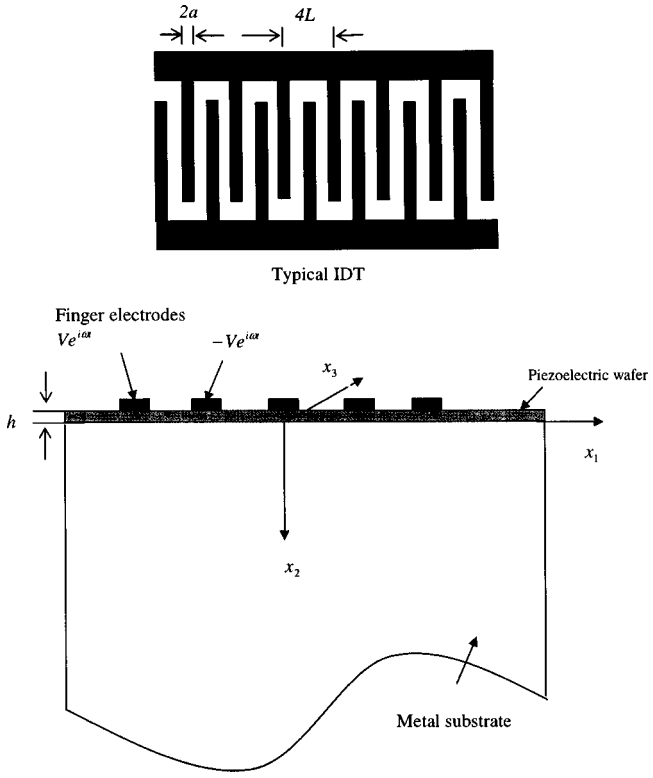


Fig. 1 Piezoelectric coupled medium with surface bonded interdigital transducer

wave propagation in a piezoelectric coupled plate with finitely long interdigital transducer by fully taking account of the piezoelectric effects in their analysis.

As an extension of the research in [23,24], this paper is to derive an analytical solution for shear horizontal wave propagation excited by interdigital transducer in a piezoelectric coupled semi-infinite medium with the coupling electromechanical effects fully modelled. Such a solution is especially essential for the application of interdigital transducer in the health monitoring of structures as discussed. The convergence of the algorithm is also investigated in the current research.

The substrate considered in this paper is a semi-infinite metal, surface bonded with an interdigital transducer abutting the vacuum. The dispersion characteristics of the structure are first obtained. The mathematical solution for the wave propagation in the piezoelectric medium with an infinitely long interdigital transducer is next presented. The solutions in terms of elliptic integration of the first kind and the standard integral representation for Legendre polynomial are derived. The hypothesis for the limitation of the wavelength of interdigital transducer in the model is presented and verified. Thirdly, the analytical solution for the wave propagation in the medium with finitely long interdigital transducer is obtained using Fourier transformation. It is hoped that this paper provides fundamental contributions to understanding wave propagation in piezoelectric structure by interdigital transducer and useful for the design of interdigital transducer in practical applications.

2 Problem Description

Consider a metallic half-space ($x_2>0$) covered by a piezoelectric layer of thickness h ($-h<x_2<0$) with interdigital transducer electrodes deposited on it as shown in Fig. 1. Each "comb-like" electrode has a regular finger spacing of $4L$ and the width of each finger is $2a$. An electric voltage applied across the electrodes will generate an alternating periodic electric field in the x_1-direction and cause wave propagation with a single frequency in the piezoelectric coupled structure. The poling direction of the piezoelectric layer is in the transverse x_3-direction and hence only shear horizontal wave will be studied in this layered structure. Therefore, the only nonvanishing displacement component in both host structure and piezoelectric layer is the one in x_3-direction, which is also independent of x_3. The corresponding shear stresses components and the equations of motion in the host medium are given by

$$\sigma'_{13} = c'_{44}\frac{\partial u'_3}{\partial x_1}, \tag{1a}$$

$$\sigma'_{23} = c'_{44}\frac{\partial u'_3}{\partial x_2} \tag{1b}$$

$$\frac{\partial \sigma'_{13}}{\partial x_1} + \frac{\partial \sigma'_{23}}{\partial x_2} = \rho'\frac{\partial^2 u'_3}{\partial t^2} \tag{1c}$$

from which the propagation of shear horizontal wave excited by interdigital transducer in the host structure is obtained by:

$$c'_{44}\nabla^2 u'_3 = \rho'\frac{\partial^2 u'_3}{\partial t^2} \tag{1d}$$

where c'_{44} is the shear modulus, ρ' the mass density, and u'_3 is the deflection, all of the host medium. The Laplace operator is $\nabla^2 = (\partial/\partial x_1) + (\partial/\partial x_2)$. The shear stress in the host semi-infinite medium in x_2-direction can be written as:

$$\sigma'_{23} = c'_{44}\frac{\partial u'_3}{\partial x_2} \tag{2}$$

In the piezoelectric layer, the equation of the electrostatic field, the constitutive relations, and the equation of motion are give as follows:

$$D_1 = e_{15}\frac{\partial u_3}{\partial x_1} - \Xi_{11}\frac{\partial \phi}{\partial x_1}, \tag{3a}$$

$$D_2 = e_{15}\frac{\partial u_3}{\partial x_2} - \Xi_{11}\frac{\partial \phi}{\partial x_2}, \tag{3b}$$

$$\frac{\partial D_1}{\partial x_1} + \frac{\partial D_2}{\partial x_2} = 0 \tag{3c}$$

$$\sigma_{13} = c_{44}\frac{\partial u_3}{\partial x_1} + e_{15}\frac{\partial \phi}{\partial x_1}, \tag{3d}$$

$$\sigma_{23} = c_{44}\frac{\partial u_3}{\partial x_2} + e_{15}\frac{\partial \phi}{\partial x_2} \tag{3e}$$

$$\frac{\partial \sigma_{13}}{\partial x_1} + \frac{\partial \sigma_{23}}{\partial x_2} = \rho\frac{\partial^2 u_3}{\partial t^2} \tag{3f}$$

from which, the coupling equation for the piezoelectric layer is given by [25]

$$c_{44}\nabla^2 u_3 + e_{15}\nabla^2 \phi = \rho\frac{\partial^2 u_3}{\partial t^2} \tag{4a}$$

$$e_{15}\nabla^2 u_3 - \Xi_{11}\nabla^2 \phi = 0 \tag{4b}$$

where c_{44} is the elastic modulus, e_{15} the piezoelectric coefficient, Ξ_{11} the dielectric constant, ρ the mass density, all of the piezoelectric layer, and u_3 is the deflection, D_1 and D_2 the electric displacement, and ϕ the electric potential, in the piezoelectric layer. The shear stress, electric field and electric displacement in the piezoelectric layer in the x_2-direction are written as,

$$\sigma_{23} = c_{44}\frac{\partial u_3}{\partial x_2} + e_{15}\frac{\partial \phi}{\partial x_2} \tag{5}$$

$$E_2 = -\frac{\partial \phi}{\partial x_2} \tag{6}$$

$$D_2 = e_{15}\frac{\partial u_3}{\partial x_2} + \Xi_{11}E_2 = e_{15}\frac{\partial u_3}{\partial x_2} - \Xi_{11}\frac{\partial \phi}{\partial x_2} \tag{7}$$

The potential $\tilde{\phi}$ in the vacuum (above the piezoelectric layer) can be derived by solving Maxwell's equation as follows:

$$\nabla^2 \tilde{\phi} = 0 \tag{8}$$

The corresponding electric displacement $\tilde{D}$ is given by the relationship

$$\tilde{D} = -\Xi_0 \frac{\partial \tilde{\phi}}{\partial x_2} \tag{9}$$

where Ξ_0 is the dielectric constant of the vacuum.

The boundary conditions for the kinematics, traction and electric fields of the piezoelectric coupled media are expressed as, at $x_2 = 0$:

$$u_3 = u_3' \tag{10}$$

$$\sigma_{23} = \sigma_{23}' \tag{11}$$

$$\phi = 0 \tag{12}$$

at $x_2 = -h$:

$$\sigma_{23} = 0 \tag{13}$$

$$\phi = \tilde{\phi} \tag{14}$$

$$D_2 = \tilde{D} \tag{15}$$

Based on the above equations, the solution for the wave excitation by interdigital transducer in the piezoelectric coupled medium, including the propagation dispersion characteristics, will be obtained and discussed in the following sections.

3 Dispersion Characteristics of Shear Horizontal Wave Propagation for the Close-Circuit Case

Prior to obtaining the wave excitation by interdigital transducer in a coupled structure, the dispersion characteristics of the shear horizontal wave propagation in the piezoelectric coupled medium is first summarized, based on an earlier study [26] where the surfaces of the piezoelectric layer are fully coated with electrode films for closed-circuit case, i.e. the potential on electrode surfaces is equally null.

For the case when the piezoelectric layer is closely connected, the solution of u_3' is given by

$$u_3' = f'(x_2)e^{j(\omega t - kx_1)} \tag{16}$$

where k is the wave number of propagating wave; ω is circular frequency of the motion; $j = \sqrt{-1}$.

Substituting Eq. (16) into Eq. (1*d*) yields,

$$\frac{d^2 f'}{dx_2^2} + \alpha^2 f' = 0 \tag{17}$$

where $\alpha^2 = \omega^2/v'^2 - k^2$, $v'^2 = c_{44}'/\rho'$.

The solution for $v = (\omega/k) < v'$ is found to be

$$u_3' = \overline{A}e^{-\chi' x_2}e^{j(\omega t - kx_1)} \tag{18}$$

where $\chi' = k\sqrt{1-(\omega/kv')^2}$. For $v \geq v'$, the solution represents refracted waves carrying energy away from the layer. Such a wave system is not of significance at any distance because it loses energy quickly and will not discussed further in this paper.

The solution in the piezoelectric layer can be obtained by first substituting

$$\psi = \phi - \frac{e_{15}}{\Xi_{11}}u_3 \tag{19}$$

into Eq. (4*b*) to yield

$$\nabla^2 \psi = 0 \tag{20}$$

where the solution can be written as

$$\psi = (B_1 e^{-kx_2} + B_2 e^{kx_2})e^{j\omega(t-(kx_1/\omega))} \tag{21}$$

In addition, substituting Eq. (4*b*) into (4*a*) gives

$$\overline{c_{44}}\nabla^2 u_3 = \rho\frac{\partial^2 u_3}{\partial t^2} \tag{22}$$

where $\overline{c_{44}} = c_{44} + (e_{15}^2/\Xi_{11})$. The solution of Eq. (22) is given by

$$u_3 = (A_1 e^{-\chi x_2} + A_2 e^{\chi x_2})e^{j\omega(t-(kx_1/\omega))} \quad \text{when } v < v_p, v' \tag{23}$$

$$u_3 = (A_1 \cos \chi x_2 + A_2 \sin \chi x_2)e^{j\omega(t-(kx_1/\omega))} \quad \text{when } v' > v > v_p \tag{24}$$

where $\chi = k\sqrt{|1-(\omega/kv)^2|}$, $v_p^2 = \overline{c_{44}}/\rho$ for the phase velocity of piezoelectric layer.

Substituting Eqs. (21), (23), and (24) into Eqs. (19) and (7), ϕ and D_2 can be expressed as

$$\phi = \left[(B_1 e^{-kx_2} + B_2 e^{kx_2}) + \frac{e_{15}}{\Xi_{11}}(A_1 e^{-\chi x_2} + A_2 e^{\chi x_2})\right]e^{j\omega(t-(k_i x_1/\omega))} \tag{25}$$

$$D_2 = -\Xi_{11}[k(-B_1 e^{-kx_2} + B_2 e^{kx_2})]e^{j\omega(t-(kx_1/\omega))} \tag{26}$$

when $v < v_p$, v', and

$$\phi = \left[(B_1 e^{-kx_2} + B_2 e^{kx_2}) + \frac{e_{15}}{\Xi_{11}}(A_1 \cos \chi x_2 + A_2 \sin \chi x_2)\right]e^{j\omega(t-(kx_1/\omega))} \tag{27}$$

$$D_2 = -\Xi_{11}[k(B_1 e^{-kx_2} - B_2 e^{kx_2})]e^{j\omega(t-(kx_1/\omega))} \tag{28}$$

when $v' > v > v_p$.

Substituting Eqs. (18) and (23)–(26) into Eqs. (2) and (5) gives

$$\sigma_{23}' = c_{44}'(-\chi')\overline{A}e^{-\chi' x_2}e^{j\omega(t-(kx_1/\omega))} \tag{29}$$

$$\sigma_{23} = [(-\chi)\overline{c_{44}}(A_1 e^{-\chi x_2} - A_2 e^{\chi x_2}) + (-k)e_{15}(B_1 e^{-kx_2} - B_2 e^{kx_2})]e^{j\omega(t-(kx_1/\omega))} \tag{30}$$

when $v < v_p$, v', and

$$\sigma_{23} = [(-\chi)\overline{c_{44}}(A_1 \sin \chi x_2 - A_2 \cos \chi x_2) + (-k)e_{15}(B_1 e^{-kx_2} - B_2 e^{kx_2})]e^{j\omega(t-(kx_1/\omega))} \tag{31}$$

when $v' > v > v_p$.

In the paper, the solutions for the case $v < v_p$, v' is of main concern when interdigital transducer is considered in wave excitation as will be explained later. Hence the expressions for the deflection, shears stress and electric variables in the piezoelectric coupled medium given by Eqs. (23), (30), (25), and (26) will be employed.

The boundary conditions pertaining to this closed-circuit case are as follows. At $x_2 = 0$, Eqs. (10)–(12) remain valid, but at $x_2 = -h$, the conditions are given as,

$$\sigma_{23} = 0 \tag{32}$$

$$\phi = 0 \tag{33}$$

Enforcing these five boundary conditions yields:

$$\overline{A} = A_1 + A_2 \tag{34}$$

Table 1 Material properties

	Host structure (Steel)	Piezoelectric layer (PZT4)
Young's module (N/m^2)	$E=210\times10^9$	$c_{44}=8.5\times10^9$
Mass density (kg/m^3)	7.8×10^3	7.5×10^3
e_{15} (C/m^2)	⋯	10.5
e_{31} (C/m^2)	⋯	−4.1
Ξ_0 (F/m)	⋯	8.854×10^{-12}
Ξ_{11}/Ξ_0	⋯	800
Ξ_{33}/Ξ_0	⋯	660

$$(-\chi)\overline{c_{44}}(A_1-A_2)+(-k)e_{15}(B_1-B_2)=(-\chi')c'_{44}\overline{A} \tag{35}$$

$$B_1+B_2+\frac{e_{15}}{\Xi_{11}}(A_1+A_2)=0 \tag{36}$$

$$(-\chi)\overline{c_{44}}(A_1e^{\chi h}-A_2e^{-\chi h})+(-k)e_{15}(B_1e^{kh}-B_2e^{-kh})=0 \tag{37}$$

$$(B_1e^{kh}+B_2e^{-kh})+\frac{e_{15}}{\Xi_{11}}(A_1e^{\chi h}+A_2e^{-\chi h})=0 \tag{38}$$

The existence of nontrivial solutions for the coefficients $\overline{A}$, A_1, A_2, B_1, and B_2 is studied through the usual eigenvalue formulation and briefly summarized. First, A_1 and A_2 may be expressed in terms of B_1 and B_2 from Eqs. (34)–(36) as follows

$$A_1=N_1B_1+N_2B_2 \tag{39}$$

$$A_2=S_1B_1+S_2B_2 \tag{40}$$

Substituting the above two equations into Eqs. (37) and (38) yields the following two equations:

$$Q_1B_1+Q_2B_2=0 \tag{41}$$

$$R_1B_1+R_2B_2=0 \tag{42}$$

The variables N_1, N_2, S_1, S_2 and Q_1, Q_2, R_1, R_2 in Eqs. (39)–(42) are listed in the Appendix. The existence of nontrivial solutions for B_1 and B_2, and hence A_1 and A_2, is thus given by

$$\Delta=R_1Q_2-R_2Q_1=0 \tag{43}$$

The dispersion curves corresponding to different wave modes will be given for the steel-PZT coupled medium. Table 1 lists the material properties used in the numerical simulations. The shear wave velocities for the host steel and piezoelectric materials are $v'_{\text{steel}}=3281$ m/s, $v_p=2351$ m/s. The Bleustein–Gulyaev surface wave velocities in PZT4 can be determined by the equation

$$v_B=v_p\sqrt{1-\frac{k_{15}^4}{(1+k_{15}^2)^2}},$$

where $k_{15}^2=e_{15}^2/c_{44}\Xi_{11}$ [27,28]. Numerically, this surface wave velocity is $v_B=2181$ m/s.

Figure 2 shows the phase velocities for the first four modes, where the nondimensional phase velocity is taken as $\overline{v}=v/v_B$ and the nondimensional wave number is given by $\overline{k}=kh/2\pi$. For the first mode, the phase velocity converges to the Bleustein–Gulyaev wave velocity for a large wave number. This is due to the fact that the surface wave for the piezoelectric layer becomes dominant when the wave number is large compared with the thickness of the layer. The higher modes only exist beyond certain a wave number, for example, the second mode begins at approximately $\overline{k}=0.4$. The wave velocities of the higher modes approach the shear velocity of the piezoelectric layer with increasing wave number. The corresponding $\overline{\omega}-\overline{k}$ is shown in Fig. 3, where the nondimensional

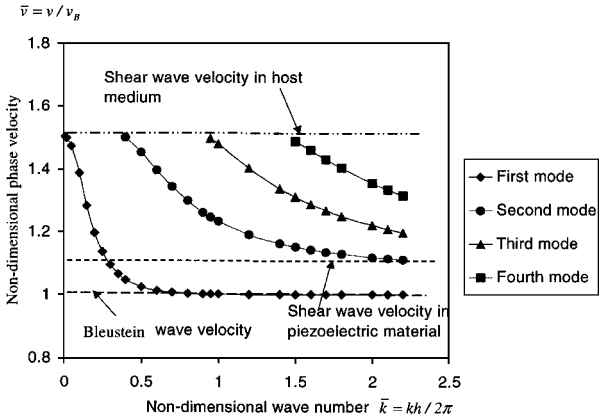


Fig. 2 Dispersive curves for PZT 4 piezoelectric coupled plate

frequency is given by $\overline{\omega}=\overline{v}\overline{k}$. It can be noted that at high wave number the relationship is virtually linear. This figure will be referred to when searching for the wave number at a specific frequency in the process of solving for the shear horizontal wave propagation excited by interdigital transducer in the piezoelectric coupled medium. Thus, the derived results for the shear horizontal wave motion in the piezoelectric coupled medium will provide a foundation for the wave solution excited by the interdigital transducer proposed next.

4 Analytical Solution for Wave Propagation in the Medium With Infinitely Long Interdigital Transducer

To provide a mathematical solution for shear wave propagation excited by an interdigital transducer with finite length, the theoretical solution for the wave motion by an infinitely long interdigital transducer has to be studied preliminarily. For an infinitely long interdigital transducer with periodic finger spacing shown in Fig. 1, the solution of u'_3 in the host medium may be written in periodic form with wavelength of $4L$ corresponding to the finger spacing in each electrode in the x_1-direction. Following Eq. (18), the deflection u'_3 can thus be expressed in Fourier series as

$$u'_3=\sum_{i=i_1}^{\infty}\overline{A_i}e^{-\chi'_ix_2}e^{j(\omega t-k_ix_1)} \tag{44}$$

where

$$i_1=INT\left[\frac{\omega}{v'}\frac{L}{\pi}-\frac{1}{2}\right],$$

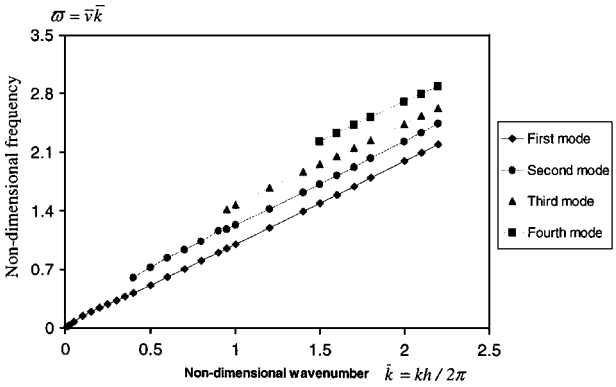


Fig. 3 Relationship between frequency and wave number

$$k_i = \frac{\pi}{L}\left(i + \frac{1}{2}\right),$$

$$\chi_i' = k_i \sqrt{1 - \left(\frac{\omega}{k_i v'}\right)^2}$$

and $INT[.]$ stands for the integer part of the number within the bracket. The corresponding shear stress in the metal core given by Eq. (2) is

$$\sigma_{23}' = \sum_{i=I}^{\infty} c_{44}'(-\chi_i')\overline{A}_i e^{-\chi_i' x_2} e^{j\omega(t-(k_i x_1/\omega))} \tag{45}$$

The variables ψ and u_3 for the piezoelectric layer following Eqs. (21), (23), and (24) can similarly be written as

$$\psi = \sum_{i=0}^{\infty} (B_{1i} e^{-k_i x_2} + B_{2i} e^{k_i x_2}) e^{j\omega(t-(k_i x_1/\omega))} \tag{46}$$

$$u_3 = \sum_{i=I}^{\infty} (A_{1i} e^{-\chi_i x_2} + A_{2i} e^{\chi_i x_2}) e^{j\omega(t-(k_i x_1/\omega))} \text{ when } \frac{\omega}{k_i} < v_p, v' \tag{47}$$

$$u_3 = \sum_{i=i_1}^{i_2} (A_{1i} \cos \chi_i x_2 + A_{2i} \sin \chi_i x_2) e^{j\omega(t-(k_i x_1/\omega))} \text{ when } v' > \frac{\omega}{k_i} > v_p \tag{48}$$

where

$$\chi_i = k_i \sqrt{\left|1 - \left(\frac{\omega}{k_i v}\right)^2\right|},$$

$$i_2 = INT\left[\frac{\omega}{v}\frac{L}{\pi} - \frac{1}{2}\right]$$

and $I = \max(i_1, i_2)$.

The expressions for ϕ and D_2 corresponding to those of Eqs. (25)–(28) are shown below,

$$\phi = \sum_{i=I}^{\infty} \left[(B_{1i} e^{-k_i x_2} + B_{2i} e^{k_i x_2}) + \frac{e_{15}}{\Xi_{11}} (A_{1i} e^{-\chi_i x_2} + A_{2i} e^{\chi_i x_2}) \right] e^{j\omega(t-(k_i x_1/\omega))} \tag{49}$$

$$D_2 = \sum_{i=I}^{\infty} -\Xi_{11}[k_i(-B_{1i} e^{-k_i x_2} + B_{2i} e^{k_i x_2})] e^{j\omega(t-(k_i x_1/\omega))} \tag{50}$$

when $\omega/k_i < v_p, v'$, and

$$\phi = \sum_{i=i_1}^{i_2} \left[(B_{1i} e^{-k_i x_2} + B_{2i} e^{k_i x_2}) + \frac{e_{15}}{\Xi_{11}} (A_{1i} \cos \chi_i x_2 + A_{2i} \sin \chi_i x_2) \right] e^{j\omega(t-(k_i x_1/\omega))} \tag{51}$$

$$D_2 = \sum_{i=i_1}^{i_2} -\Xi_{11}[k_i(-B_{1i} e^{-k_i x_2} - B_{2i} e^{k_i x_2})] e^{j\omega(t-(k_i x_1/\omega))} \tag{52}$$

when $v' > \omega/k_i > v_p$.

The shear stress in the piezoelectric layer based on Eq. (5) is

$$\sigma_{23} = \sum_{i=I}^{\infty} [(-\chi_i)\overline{c}_{44}(A_{1i} e^{-\chi_i x_2} - A_{2i} e^{\chi_i x_2}) + (-k_i) e_{15}(B_{1i} e^{-k_i x_2} - B_{2i} e^{k_i x_2})] e^{j\omega(t-(k_i x_1/\omega))} \tag{53}$$

when $\omega/k_i < v_p, v'$, and

$$\sigma_{23} = \sum_{i=i_1}^{i_2} [(-\chi_i)\overline{c}_{44}(A_{1i} \sin \chi_i x_2 - A_{2i} \cos \chi_i x_2) + (-k_i) e_{15}(B_{1i} e^{-k_i x_2} - B_{2i} e^{k_i x_2})] e^{j\omega(t-(k_i x_1/\omega))} \tag{54}$$

when $v' > \omega/k_i > v_p$.

The potential $\tilde{\phi}$ in the vacuum can be derived below by considering Maxwell equation

$$\nabla^2 \tilde{\phi} = 0 \tag{55}$$

where the solution remains finite as $x_2 \to -\infty$. Hence, $\tilde{\phi}$ and $\tilde{D}$ take the form

$$\tilde{\phi} = \sum_{i=0}^{\infty} C_i e^{k_i x_2} e^{j\omega(t-(k_i x_1/\omega))} \tag{56}$$

$$\tilde{D} = \sum_{i=0}^{\infty} -\Xi_0 k C_i e^{k_i x_2} e^{j\omega(t-(k_i x_1/\omega))} \tag{57}$$

The boundary conditions pertaining to this infinitely long interdigital transducer set-up can be stated as follows. At $x_2 = 0$, Eqs. (10)–(12) hold, whereas at $x_2 = -h$,

$$\sigma_{23} = 0 \tag{58}$$

$$\phi = \tilde{\phi} \tag{59}$$

$$D_2 = \tilde{D} \text{ Outside the electrodes} \tag{60}$$

$$\phi = \tilde{\phi} = V \text{ Inside the electrodes} \tag{61}$$

where V is the magnitude of the alternating voltage applied on the interdigital transducer.

The analytical solution of the wave propagation excited by interdigital transducer in the piezoelectric coupled medium provided below is under the hypothesis of $I = \max(i_1, i_2) = 0$. This hypothesis ensures that all the solutions of the physical variables in the piezoelectric layer follow Eqs. (47), (49), and (50) by proper design of the basic wave number $k_0 = \pi/2L$, i.e., the design of the wavelength of the interdigital transducer. The validity of this hypothesis is discussed hereinafter.

As an illustration, consider a steel-PZT 4 piezoelectric coupled medium where the bulk shear wave velocity of steel and PZT 4 are about $v' = 3281$ m/s, $v_p = 2351$ m/s. If the circular frequency of the excitation voltage is used as 1.4 MHz, the hypothesis of $I = 0$ requires $L < 11.1$ mm, which means the wavelength of interdigital transducer is 44.4 mm. Such requirement is satisfied by most interdigital transducer designs, especially for MEMS designs. Upon the above observation, the hypothesis of using Eqs. (47), (49), and (50) for the solutions of wave propagation in the piezoelectric layer is thus reasonable and valid for most of the designs of the interdigital transducer.

Substituting the solutions into boundary conditions by assuming $I = 0$ results in

$$\sum_{i=0}^{\infty} \overline{A}_i e^{-jk_i x} = \sum_{i=0}^{\infty} (A_{1i} + A_{2i}) e^{-jk_i x_1} \tag{62}$$

$$\sum_{i=0}^{\infty}[(-\chi_i)\overline{c_{44}}(A_{1i}-A_{2i})+(-k_i)e_{15}(B_{1i}-B_{2i})]e^{-jk_ix_1}$$
$$=\sum_{i=0}^{\infty}(-\chi_i')c'_{44}\overline{A_i}e^{-jk_ix_1} \tag{63}$$

$$\sum_{i=0}^{\infty}\left(B_{1i}+B_{2i}+\frac{e_{15}}{\Xi_{11}}+(A_{1i}+A_{2i})\right)e^{-jk_ix_1}=0 \tag{64}$$

$$\sum_{i=0}^{\infty}((-\chi_i)\overline{c_{44}}(A_{1i}e^{\chi_2 h}-A_{2i}e^{-\chi_2 h})+(-k_i)e_{15}(B_{1i}e^{k_ih}$$
$$-B_{2i}e^{-k_ih}))e^{-jk_ix_1}=0 \tag{65}$$

$$\sum_{i=0}^{\infty}\left((B_{1i}e^{k_ih}+B_{2i}e^{-k_ih})+\frac{e_{15}}{\Xi_{11}}+(A_{1i}e^{\chi_2 h}+A_{2i}e^{-\chi_i h})\right)e^{-jk_ix_1}$$
$$=\sum_{i=0}^{\infty}C_ie^{-k_ih}e^{-jk_ix_1} \tag{66}$$

$$\sum_{i=0}^{\infty}(-\Xi_{11}(-k_iB_{1i}e^{k_ih}+k_iB_{2i}e^{-k_ih})+\Xi_0k_iC_ie^{-k_ih})e^{-jk_ix_1}$$
$$=0\quad a<x_1<L \tag{67}$$

$$\sum_{i=0}^{\infty}C_ie^{-k_ih}e^{-jk_ix_1}=V\quad 0<x_1<a \tag{68}$$

The analytical solution for all the six sets of coefficients $\overline{A_i}$, A_{1i}, A_{2i}, B_{1i}, B_{2i}, and C_i $(i=1,2\ldots,\infty)$ will be determined in the same manner as in Sec. 3. The coefficients A_{1i} and A_{2i} may be expressed as follows:

$$A_{1i}=N_{1i}B_{1i}+N_{2i}B_{2i} \tag{69}$$

$$A_{2i}=S_{1i}B_{1i}+S_{2i}B_{2i} \tag{70}$$

where N_{1i}, N_{2i}, S_{1i}, S_{2i}, $i=0,1,2,\ldots3,\infty$ are shown in the Appendix.

Substituting the above two equations into Eqs. (65) and (66) yields

$$\left(N_{1i}e^{\chi_ih}-S_{1i}e^{-\chi_ih}+\frac{k_ie_{15}}{\chi_i\overline{c_{44}}}e^{k_ih}\right)B_{1i}+\left(N_{2i}e^{\chi_ih}-S_{2i}e^{-\chi_ih}\right.$$
$$\left.-\frac{k_ie_{15}}{\chi_i\overline{c_{44}}}e^{k_ih}\right)B_{2i}=0 \tag{71}$$

$$\text{i.e.}\quad Q_{1i}B_{1i}+Q_{2i}B_{2i}=0 \tag{72}$$

$$\left(e^{k_ih}+\frac{e_{15}}{\Xi_{11}}N_{1i}e^{\chi_ih}+\frac{e_{15}}{\Xi_{11}}S_{1i}e^{-\chi_ih}\right)B_{1i}+\left(e^{-k_ih}+\frac{e_{15}}{\Xi_{11}}N_{2i}e^{\chi_ih}\right.$$
$$\left.+\frac{e_{15}}{\Xi_{11}}S_{2i}e^{-\chi_ih}\right)B_{2i}=C_ie^{-k_ih} \tag{73}$$

$$\text{i.e.}\quad R_{1i}B_{1i}+R_{2i}B_{2i}=C_ie^{-k_ih} \tag{74}$$

From Eqs. (72) and (74), B_{1i} and B_{2i} may obtained in terms of C_i as

$$B_{1i}=\frac{C_ie^{-k_ih}}{R_{1i}-\frac{Q_{1i}}{Q_{2i}}R_{2i}} \tag{75}$$

$$B_{2i}=\frac{C_ie^{-k_ih}}{R_{2i}-\frac{Q_{2i}}{Q_{1i}}R_{1i}} \tag{76}$$

Substituting the above two expressions into Eq. (67) gives,

$$\sum_{i=0}^{\infty}k_iC_ie^{-k_ih}\left(\Xi_0+\frac{\Xi_{11}e^{k_ih}}{R_{1i}-\frac{Q_{1i}}{Q_{2i}}R_{2i}}-\frac{\Xi_{11}e^{-k_ih}}{R_{2i}-\frac{Q_{2i}}{Q_{1i}}R_{1i}}\right)e^{-jk_ix_1}$$
$$=0\quad a<x_1<L \tag{77}$$

Denoting

$$\overline{C_i}=C_ie^{-k_ih}\left(\Xi_0+\frac{\Xi_{11}e^{k_ih}}{R_{1i}-\frac{Q_{1i}}{Q_{2i}}R_{2i}}-\frac{\Xi_{11}e^{-k_ih}}{R_{2i}-\frac{Q_{2i}}{Q_{1i}}R_{1i}}\right) \tag{78}$$

Eqs. (67) and (68) may be rearranged as

$$\sum_{i=0}^{\infty}k_i\overline{C_i}e^{-jk_ix_1}=0\quad 0<x_1<a \tag{79}$$

$$\sum_{i=0}^{\infty}\overline{C_i}(1+F_i)e^{-jk_ix_1}=V\quad a<x_1<L \tag{80}$$

where $F_i=\left(\Xi_0+\frac{\Xi_{11}e^{k_ih}}{R_{1i}-\frac{Q_{1i}}{Q_{2i}}R_{2i}}-\frac{\Xi_{11}e^{-k_ih}}{R_{2i}-\frac{Q_{2i}}{Q_{1i}}R_{1i}}\right)^{-1}-1.$

Rewriting Eqs. (79) and (80) in their real function forms gives,

$$\sum_{i=0}^{\infty}\left(1+\frac{1}{2}\right)\overline{C_i}\cos\left(i+\frac{1}{2}\right)\overline{x}=0\quad \overline{a}<\overline{x}<\pi \tag{81}$$

$$\sum_{i=0}^{\infty}\overline{C_i}(1+F_i)\cos\left(i+\frac{1}{2}\right)\overline{x}=V\quad 0<\overline{x}<\overline{a} \tag{82}$$

where $\overline{x}=\pi x_1/L$, $\overline{a}=\pi a/L$ are nondimensional parameters.

The solutions for Eqs. (81) and (82) can be obtained from an infinite system of linear algebraic equations (see Bateman and Erdelyi [27] and Parton and Kudryavtser [8]).

The set of equations is given as,

$$\overline{C_i}=\frac{VP_i(\cos\overline{a})}{\left(i+\frac{1}{2}\right)K\left(\cos\frac{\overline{a}}{2}\right)}-\sum_{n=0}^{\infty}\overline{C_n}F_n\beta_{ni}\quad (i=1,2,\ldots,\infty) \tag{83}$$

where

$$\text{where }\beta_{ni}=\left(i+\frac{1}{2}\right)\int_0^{\overline{\alpha}}P_n(\cos\xi)P_i(\cos\xi)\sin\xi d\xi \tag{84}$$

$$P_i(\cos\xi)=\frac{\sqrt{2}}{\pi}\int_0^{\xi}\frac{\cos\left(i+\frac{1}{2}\right)xdx}{\sqrt{\cos x-\cos\xi}} \tag{85}$$

$P_i(\cos\xi)$ in the above equation is the standard integral representation for the Legendre polynomial, and

$$K\left(\cos\frac{\xi}{2}\right)=\sum_{i=0}^{\infty}\frac{P_i(\cos\xi)}{\left(i+\frac{1}{2}\right)}$$

is the full elliptic integral of the first kind [22].

In practice, finite N terms are used in Eq. (83), where N must be sufficiently large to ensure convergence of the solution. The system of N equations to solve for the coefficients $\overline{C_i}$ $(i=1,2,\ldots,N)$ is

$$[F]\{\overline{C}\} = \frac{V}{K\left(\cos\frac{\overline{a}}{2}\right)}\{P\} \tag{86}$$

where

$$[F] = \begin{bmatrix} 1+F_1\beta_{11} & F_2\beta_{21} & \ldots & F_N\beta_{N1} \\ F_1\beta_{12} & 1+F_2\beta_{22} & \ldots & F_N\beta_{N2} \\ \ldots & \ldots & \ldots & \ldots \\ F_1\beta_{1N} & F_2\beta_{2N} & \ldots & 1+F_N\beta_{NN} \end{bmatrix} \tag{87}$$

$$\{P\} = \begin{Bmatrix} P_1(\cos\overline{a})/(1+1/2) \\ P_2(\cos\overline{a})/(2+1/2) \\ \ldots \\ P_N(\cos\overline{a})/(N+1/2) \end{Bmatrix} \tag{88}$$

and $\{\overline{C}\}=\{\overline{C_1},\overline{C_2},\ldots,\overline{C_N}\}^T$.

From Eq. (86), the solution of $\{\overline{C}\}$ is obtained as,

$$\{\overline{C}\} = \frac{V}{K\left(\cos\frac{\overline{a}}{2}\right)}[F]^{-1}\{P\} \tag{89}$$

Finally, the coefficients $\{C\}=\{C_1,C_2,\ldots,C_N\}^T$ is derived according to Eq. (78)

$$\{C\} = \mathrm{diag}(e^{k_i h}(F_i+1))\{\overline{C}\} = \frac{V}{K\left(\cos\frac{\overline{a}}{2}\right)}\,\mathrm{diag}(e^{k_i h}(F_i+1))[F]^{-1}\{P\} \tag{90}$$

where diag$(\cdot)$ denotes a diagonal matrix.

The coefficients $\{B_1\}=\{B_{11},B_{12},\ldots,B_{1N}\}^T$ and $\{B_2\}=\{B_{21},B_{22},\ldots,B_{2N}\}^T$ can be derived from Eqs. (75) and (76), and the coefficients, $\{A_1\}=\{A_{11},A_{12},\ldots,A_{1N}\}^T$, $\{A_2\}=\{A_{21},A_{22},\ldots,A_{2N}\}^T$, and $\{\overline{A}\}=\{\overline{A_1},\overline{A_2},\ldots,\overline{A_N}\}^T$ can be obtained accordingly as well. All the variables are listed in the Appendix.

The analytical solutions of the deflection in both the metal substrate and the piezoelectric layer, the electric potential and electric displacement in the piezoelectric layer and the electric potential and electric displacement in the vacuum can then be obtained once the results of coefficients $\{\overline{A}\}$, $\{A_1\}$, $\{A_2\}$, $\{B_1\}$, $\{B_2\}$, and $\{C\}$ are known.

To illustrate the convergence of the solution in Eq. (89), the solution for C_i $(i=1,2,\ldots,25)$ is plotted in Fig. 4 when $\overline{a}=0.3$, 0.5, and 0.8, respectively. It can be seen that the convergence of the solution of Eq. (89) is assured numerically.

5 Analytical Solution for Wave Propagation in the Medium With Finitely Long Interdigital Transducer

In engineering applications, the interdigital transducer is usually of finite length. The solution for the wave propagation in the medium with finitely long interdigital transducer will be discussed based on the solution for wave propagation by an infinitely long interdigital transducer provided in the previous section. The derivation of the solution is accomplished through the use of Fourier transform, instead of Fourier series. The length of the interdigital transducer is assumed to be $2L_1$.

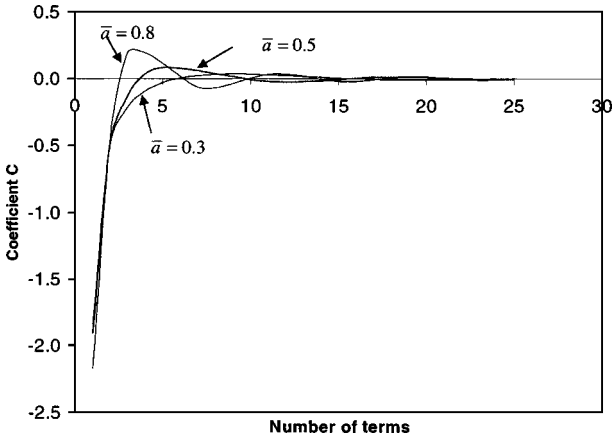


Fig. 4 The converge of the coefficient *C*

First the electric potential in the vacuum $\tilde{\phi}(x_1,x_2,t)$ of the metal for $x_2<-h$ is investigated, where its Fourier transform with respect to x_1 is

$$\tilde{\phi}(x_1,x_2,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\bar{\tilde{\phi}}(\xi,x_2,t)e^{-j\xi x_1}d\xi \tag{91}$$

The image function $\bar{\tilde{\phi}}$ can be written according to Eq. (56) as

$$\bar{\tilde{\phi}}(\xi,x_2,t) = \bar{\tilde{\phi}}(\xi,0,t)e^{\xi x_2} \tag{92}$$

The solution of the image function of $\bar{\tilde{\phi}}(\xi,x_2,t)$ requires the knowledge of the distribution of $\tilde{\phi}$ throughout the boundary. Thus, it is assumed that $\tilde{\phi}(x_1,-h,t)$ is given by Eqs. (56) and (90), which are obtained from the solution for infinitely long interdigital transducer in the electrodes region, and null outside the electrodes region. This assumption should be realistic for sufficiently long transducer gratings. Similar assumption was proposed by Parton [8] when they studied the Lamb wave propagation excited by interdigital transducer without the piezoelectric-mechanical coupling effect considered in the model.

Hence,

$$\bar{\tilde{\phi}}(\xi,-h,t) = \int_{-L_1}^{L_1}\tilde{\phi}(x_1,-h,t)e^{j\xi x_1}dx_1 \tag{93}$$

Substituting Eq. (56) into the above equation yields,

$$\bar{\tilde{\phi}}(\xi,-h,t) = \sum_{i=0}^{N} e^{-k_i h}C_i\left(\frac{\sin(k_i+\xi)L_1}{k_i+\xi}+\frac{\sin(k_i-\xi)L_1}{k_i-\xi}\right)e^{j\omega t}$$
$$= \bar{\tilde{\Phi}}(\xi,-h)e^{j\omega t} \tag{94}$$

where C_i is given by Eq. (90).

The image functions for the Fourier transform of the variables $u_3'(x_1,x_3,t)$ in Eq. (1*d*), $\psi(x_1,x_3,t)$ in Eq. (19), and $u_3(x_1,x_3,t)$ in Eq. (23) with respect to x_1, are $\overline{u_3'}(\xi,x_3,t)$, $\overline{\psi}(\xi,x_3,t)$, and $\overline{u_3}(\xi,x_3,t)$, respectively. Based on similar analyses in Eqs. (16), (23), and (21), the above variables can be written as follows:

$$\overline{u_3'}(\xi,x_2,t) = \overline{U_3'}(\xi)e^{-\chi' x_2}e^{j\omega t} \tag{95}$$

$$\overline{u_3}(\xi,x_2,t) = \left(\overline{U}_{31}(\xi)e^{-\chi x_2}+\overline{U}_{32}(\xi)e^{\chi x_2}\right)e^{j\omega t} \tag{96}$$

$$\psi(\xi,x_2,t) = (\overline{\psi_1}(\xi)e^{-\xi x_2}+\overline{\psi_2}(\xi)e^{\xi x_2})e^{j\omega t} \tag{97}$$

Thus, substituting the above variables into the boundary conditions of Eqs. (10)–(12), (58), and (59) yields,

$$\overline{U}_3' = U_{31} + \overline{U}_{32} \tag{98}$$

$$(-\chi)\overline{c}_{44}(\overline{U}_{31} - \overline{U}_{32}) + (-\xi)e_{15}(\overline{\psi}_1 - \overline{\psi}_2) = (-\chi')c_{44}'\overline{U}_3' \tag{99}$$

$$\overline{\psi}_1 + \overline{\psi}_2 + \frac{e_{15}}{\Xi_{11}}(\overline{U}_{31} + \overline{U}_{32}) = 0 \tag{100}$$

$$(-\chi)\overline{c}_{44}(\overline{U}_{31}e^{\chi h} - \overline{U}_{32}e^{-\chi h}) + (-\xi)e_{15}(\overline{\psi}_1 e^{\xi h} - \overline{\psi}_2 e^{-\xi h}) = 0 \tag{101}$$

$$(\overline{\psi}_1 e^{\xi h} + \overline{\psi}_2 e^{-\xi h}) + \frac{e_{15}}{\Xi_{11}}(\overline{U}_{31}e^{\chi h} + \overline{U}_{32}e^{-\chi h}) = \bar{\bar{\Phi}}(\xi, -h)e^{-\xi h} \tag{102}$$

To solve for $\overline{U}_3'$, $\overline{U}_{31}$, $\overline{U}_{32}$, $\overline{\psi}_1$, and $\overline{\psi}_2$, the same procedure in solving Eqs. (34)–(38) is adopted, that is,

$$\overline{U}_{31} = N_1\overline{\psi}_1 + N_2\overline{\psi}_2 \tag{103}$$

$$\overline{U}_{32} = S_1\overline{\psi}_1 + S_2\overline{\psi}_2 \tag{104}$$

Substituting the above two equations into Eqs. (101) and (102) yields

$$Q_1\overline{\psi}_1 + Q_2\overline{\psi}_2 = 0 \tag{105}$$

$$R_1\overline{\psi}_1 + R_2\overline{\psi}_2 = \bar{\bar{\Phi}}(\xi, -h) \tag{106}$$

from which gives

$$\overline{\psi}_1 = -\frac{Q_2}{\Delta}\bar{\bar{\Phi}}(\xi, -h) \tag{107}$$

$$\overline{\psi}_2 = -\frac{Q_1}{\Delta}\bar{\bar{\Phi}}(\xi, -h) \tag{108}$$

where Δ is given in Eq. (43).

From Eqs. (103), (104), and (98), we have

$$\overline{U}_{31} = \frac{\bar{\bar{\Phi}}(\xi, -h)}{\Delta}(N_2Q_1 - N_1Q_2) = \frac{\Delta_1\bar{\bar{\Phi}}(\xi, -h)}{\Delta} \tag{109}$$

$$\overline{U}_{32} = \frac{\bar{\bar{\Phi}}(\xi, -h)}{\Delta}(S_2Q_1 - S_1Q_2) = \frac{\Delta_2\bar{\bar{\Phi}}(\xi, -h)}{\Delta} \tag{110}$$

$$\overline{U}_3' = \frac{\bar{\bar{\Phi}}(\xi, -h)}{\Delta}(\Delta_1 + \Delta_2) \tag{111}$$

The variables $u_3'(x_1, x_3, t)$, $\psi(x_1, x_3, t)$, and $u_3(x_1, x_3, t)$ can thus be obtained by taking the inverse Fourier transform. As an example, consider $u_3'(x_1, x_3, t)$ where the inverse Fourier transform of $\overline{U}_3'$ is expressed as

$$u_3'(x_1, x_2, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{U}_3'(\xi)e^{-\chi' x_2}e^{j\omega t}e^{-j\xi x_1}d\xi$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\bar{\bar{\Phi}}(\xi, -h)}{\Delta}(\Delta_2 + \Delta_1)e^{j\omega t}e^{-\chi' x_2}e^{-j\xi x_1}d\xi \tag{112}$$

The above improper integral can be solved by the residue theorem. A single pole in the complex integrand of the last equation is, ξ_s, i.e. the root of the following equation,

$$\Delta = 0 \tag{113}$$

The integration of the complex function is along a close path which comprises a curve with radius R in the upper half plane and the real axis. Note that the exponential characteristic of Δ is the order of $e^{2\xi h}$, which ensures the integration is null along the curve in the upper half-plane as $R \to \infty$. The solution for the improper integration of Eq. (112) is thus obtained as,

$$u_3'(x_1, x_2, t) = \frac{j}{2}\frac{\bar{\bar{\Phi}}(\xi_s, -h)}{\Delta'|_{\xi=\xi_s}}(\Delta_2 + \Delta_1)|_{\xi=\xi_s}e^{-\chi' x_2|_{\xi=\xi_s}}e^{j(\omega t - \xi_s x_1)} \tag{114}$$

It is noted that Eq. (113) is the dispersion characteristic equation of the piezoelectric coupled media with a pair of electrodes on the upper and lower surface of the piezoelectric layer shortly connected, which was presented in Sec. 2. The value of ξ_s can be easily obtained from the dispersion curve in Fig. 3 for a given fixed frequency ω.

The deflection in the piezoelectric layer can be similarly obtained using Eqs. (96) and (97)

$$u_3(x_1, x_3, t) = \frac{j}{2}\frac{\bar{\bar{\Phi}}(\xi_s, -h)}{\Delta'|_{\xi=\xi_s}}(\Delta_1 e^{-\chi x_2} + \Delta_2 e^{\chi x_2})|_{\xi=\xi_s}e^{j(\omega t - \xi_s x_1)} \tag{115}$$

The solution for $\psi(x_1, x_3, t)$ is obtained through Eq. (98) as

$$\psi(x_1, x_3, t) = \frac{j}{2}\frac{\bar{\bar{\Phi}}(\xi_s, -h)e^{-\xi_s h}}{\Delta'|_{\xi=\xi_s}}(-Q_1 e^{-\xi x_2} + Q_2 e^{\xi x_2})|_{\xi=\xi_s}e^{j(\omega t - \xi_s x_1)} \tag{116}$$

The electric potential in the piezoelectric layer can be expressed from Eq. (19).

Since $\bar{\bar{\Phi}}(\xi_s, -h)$ appears in the expressions of all physical variables discussed above and is the key term demonstrating the effect of the geometry of interdigital transducer on the wave solutions, the numerical simulations on this term will be conducted to investigate the design of the size of the interdigital transducer on the excited wave motion in the medium. In the following simulations, the length of interdigital transducer is assumed to be half wavelength of it which means only two fingers are used in the calculations.

Figure 5 shows the distribution of $\bar{\bar{\Phi}}(\xi_s, -h)$ with respect to N against the finger width $\bar{a} = \pi a/L$ for the case where $L = 6$ mm and $h = 0.1L$. It is noted first that the convergence is obtained for $N > 15$. Another observation is that the maximum amplitude of $\bar{\bar{\Phi}}(\xi_s, -h)$ occurs sharply around $\bar{a} = 0.9$ and decays rapidly from this value. This conclusion is important in the design of the interdigital transducer since a wave signal with higher amplitude may occur at $\bar{a} = 0.9$ in this case. Thus the design of the interdigital

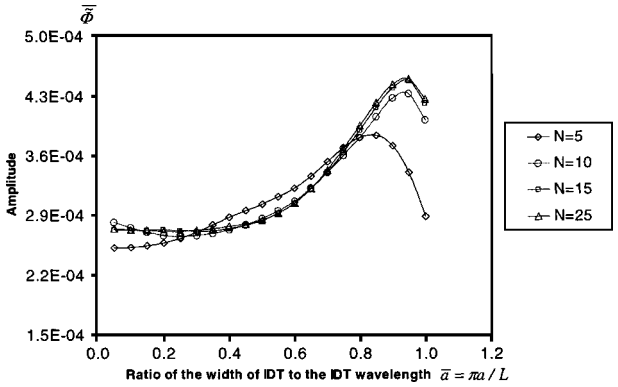


Fig. 5 Convergence of the amplitude

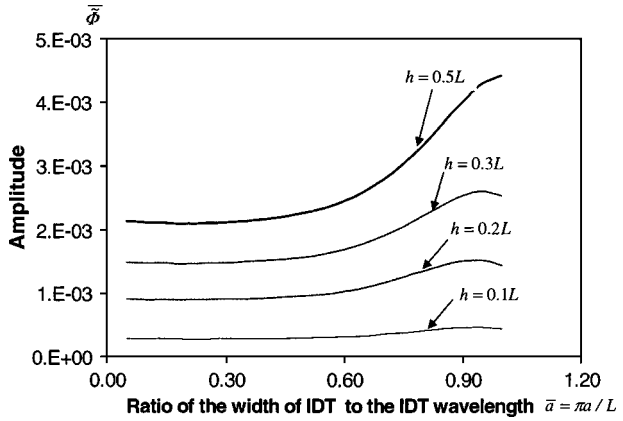


Fig. 6 The amplitude of the excited motion at different heights of the piezoelectric layer

transducer with $\bar{a}=0.9$ is useful in the application of structural health monitoring as a clearer nondispersive signal is essential in damage detection of structures.

The effect of the thickness h of the piezoelectric layer on $\bar{\bar{\Phi}}(\xi_s,-h)$ is plotted in Fig. 6 for $L=6$ mm. The results show that $\bar{\bar{\Phi}}(\xi_s,-h)$ increases with h which seems reasonable. This conclusion implies that thicker piezoelectric layer is useful for producing a clearer wave signal, i.e. wave motion with higher magnitude.

Interesting observations on the L-dependent curves can be found in Fig. 7 for $h=0.6$ mm. For $\bar{a}<0.5$, the amplitude increases with $\bar{a}$, whereas the reverse is observed for $\bar{a}>0.5$. The maximum amplitude occurs at higher $\bar{a}$. The maximum amplitude decreases as L increases until at $L=1$ cm, no distinct maximum amplitude of $\bar{\bar{\Phi}}(\xi_s,-h)$ is observed. This could possibly be due to the condition for assuming $I=0$ in deriving the analytical solution for the wave propagation, where for steel-PZT medium $L<1.1$ cm which is close to $L=1$ cm in the graph. In engineering applications, the length of the interdigital transducer cannot be too short as enough electrical input energy is also important in transmitting the wave motion. Therefore, an optimal design of the length and the finger width of the interdigital transducer has to be investigated according to the proposed simulation conclusions.

After the characteristics of the variable $\bar{\bar{\Phi}}(\xi_s,-h)$ have been derived, the physical phenomenon of other variables can therefore be obtained easily. For example, the variation of the displacement $u_3(x_1,x_3,t)$ can be studied from Eq. (115). The displacement follows a wave propagation motion in x_1 direction. The distribution of the variation in x_2 direction follows the function $\Delta_1 e^{-\chi x_2}+\Delta_2 e^{\chi x_2}$. The effect of the finger width, a, the wavelength of interdigital transducer, L, and the thickness of the piezoelectric layer finger, h, on the displacement all follow the conclusion derived above from Figs. 5–7.

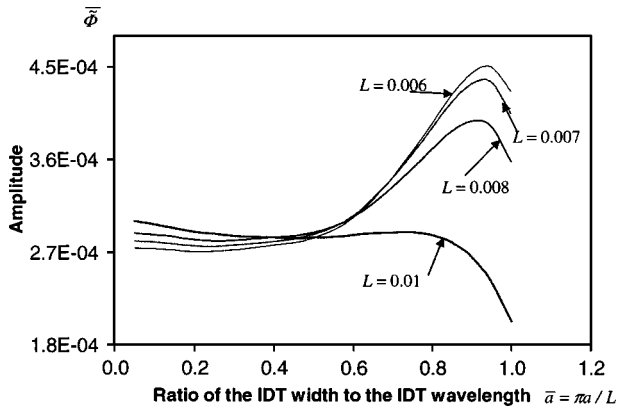


Fig. 7 The variation of the amplitude at different wavelength of IDT

6 Concluding Remarks

An analytical solution for the shear horizontal wave propagation excited by an interdigital transducer in a piezoelectric coupled medium is developed. The dispersion characteristics of the shear horizontal wave in this medium are first presented for the close-circuit case as reference for deriving the subsequent analytical solution. The solution is then shown for the case when an infinitely long interdigital transducer is used on the structure. The mathematical solution is valid for the case where the wavelength is assumed to be designed appropriately, true for most interdigital transducer used in practice with respect to steel-PZT media. The solution of wave propagation for finitely long interdigital transducer is obtained by assuming the distribution of the electric potential in the vacuum is assumed to be the same as that obtained for the case of infinitely long interdigital transducer. The solution reveals that the wave propagation in this piezoelectric coupled medium bonded by interdigital transducer follows the characteristic equation of the dispersion curve for the same structure but with a pair of electrodes shortly connected. Convergence study for the algorithm used in the analysis is investigated where at least 15 terms are needed. Parametric studies showed that the amplitude of the displacement, electric displacement and potential increases as the thickness of the piezoelectric layer. As for the effect of the finger width, the maximum amplitude occurs around $\bar{a}=0.9$. The effect of the length of interdigital transducer is complicated. For smaller $\bar{a}$, longer transducer is required for wave motion with higher magnitude, but reverse effect is observed for bigger $\bar{a}$. Therefore, optimal design for the geometry of interdigital transducer is thus necessitated in engineering applications. Further experimental work will be conducted in the near future. These conclusions are criteria for the design of interdigital transducer in its engineering applications, especially in the application of structural health monitoring in which a nondispersive wave signal with higher magnitude is preferred and essential. It is hoped that the current work could be used as a framework for the design of interdigital transducer in wave excitation of smart structures [28].

Appendix

$$N_1=\frac{1}{2}\left(-\frac{ke_{15}}{\chi\bar{c}_{44}}-\frac{\chi' c'_{44}}{\chi\bar{c}_{44}}\frac{\Xi_{11}}{e_{15}}-\frac{\Xi_{11}}{e_{15}}\right),$$

$$N_2=\frac{1}{2}\left(\frac{ke_{15}}{\chi\bar{c}_{44}}-\frac{\chi' c'_{44}}{\chi\bar{c}_{44}}\frac{\Xi_{11}}{e_{15}}-\frac{\Xi_{11}}{e_{15}}\right),$$

$$S_1=-\frac{1}{2}\left(-\frac{ke_{15}}{\chi\bar{c}_{44}}-\frac{\chi' c'_{44}}{\chi\bar{c}_{44}}\frac{\Xi_{11}}{e_{15}}+\frac{\Xi_{11}}{e_{15}}\right),$$

$$S_2=-\frac{1}{2}\left(\frac{ke_{15}}{\chi\bar{c}_{44}}-\frac{\chi' c'_{44}}{\chi\bar{c}_{44}}\frac{\Xi_{11}}{e_{15}}+\frac{\Xi_{11}}{e_{15}}\right),$$

$$Q_1=N_1e^{\chi h}-S_1e^{-\chi h}+\frac{ke_{15}}{\chi\bar{c}_{44}}e^{kh},\quad Q_2=N_2e^{\chi h}-S_2e^{-\chi h}-\frac{ke_{15}}{\chi\bar{c}_{44}}e^{kh},$$

$$R_1=e^{kh}+\frac{e_{15}}{\Xi_{11}}N_1e^{\chi h}+\frac{e_{15}}{\Xi_{11}}S_1e^{-\chi h},$$

$$R_2=e^{-kh}+\frac{e_{15}}{\Xi_{11}}N_2e^{\chi h}+\frac{e_{15}}{\Xi_{11}}S_2e^{-\chi h},$$

$$N_{1i} = \frac{1}{2}\left(- \frac{k_i e_{15}}{\chi_i \overline{c_{44}}} - \frac{\chi_i' c_{44}'}{\chi_i \overline{c_{44}}} \frac{\Xi_{11}}{e_{15}} - \frac{\Xi_{11}}{e_{15}} \right),$$

$$N_{2i} = \frac{1}{2}\left(\frac{k_i e_{15}}{\chi_i \overline{c_{44}}} - \frac{\chi_i' c_{44}'}{\chi_i \overline{c_{44}}} \frac{\Xi_{11}}{e_{15}} - \frac{\Xi_{11}}{e_{15}} \right),$$

$$S_{1i} = -\frac{1}{2}\left(- \frac{k_i e_{15}}{\chi_i \overline{c_{44}}} - \frac{\chi_i' c_{44}'}{\chi_i \overline{c_{44}}} \frac{\Xi_{11}}{e_{15}} + \frac{\Xi_{11}}{e_{15}} \right),$$

$$S_{2i} = -\frac{1}{2}\left(\frac{k_i e_{15}}{\chi_i \overline{c_{44}}} - \frac{\chi_i' c_{44}'}{\chi_i \overline{c_{44}}} \frac{\Xi_{11}}{e_{15}} + \frac{\Xi_{11}}{e_{15}} \right),$$

$$\{B_1\} = \mathrm{diag}\left(\frac{e^{-k_i h}}{R_{1i} - \dfrac{Q_{1i}}{Q_{2i}} R_{2i}} \right) \{C\} = \frac{V}{K\left(\cos \dfrac{\overline{a}}{2} \right)} \mathrm{diag}\left(\frac{(F_i + 1)}{R_{1i} - \dfrac{Q_{1i}}{Q_{2i}} R_{2i}} \right) \times [F]^{-1}\{P\},$$

$$\{B_2\} = \mathrm{diag}\left(\frac{e^{-k_i h}}{R_{2i} - \dfrac{Q_{2i}}{Q_{1i}} R_{1i}} \right) \{C\} = \frac{V}{K\left(\cos \dfrac{\overline{a}}{2} \right)} \mathrm{diag}\left(\frac{(F_i + 1)}{R_{2i} - \dfrac{Q_{2i}}{Q_{1i}} R_{1i}} \right) \times [F]^{-1}\{P\},$$

$$\{A_1\} = \frac{V}{K\left(\cos \dfrac{\overline{a}}{2} \right)} \mathrm{diag}\left(\frac{N_{1i}(F_i + 1)}{R_{1i} - \dfrac{Q_{1i}}{Q_{2i}} R_{2i}} + \frac{N_{2i}(F_i + 1)}{R_{2i} - \dfrac{Q_{2i}}{Q_{1i}} R_{1i}} \right) [F]^{-1}\{P\},$$

$$\{A_2\} = \frac{V}{K\left(\cos \dfrac{\overline{a}}{2} \right)} \mathrm{diag}\left(\frac{S_{1i}(F_i + 1)}{R_{1i} - \dfrac{Q_{1i}}{Q_{2i}} R_{2i}} + \frac{S_{2i}(F_i + 1)}{R_{2i} - \dfrac{Q_{2i}}{Q_{1i}} R_{1i}} \right) [F]^{-1}\{P\},$$

$$\{\overline{A}\} = \frac{V}{K\left(\cos \dfrac{\overline{a}}{2} \right)} \mathrm{diag}\left(\frac{(N_{1i} + S_{1i})(F_i + 1)}{R_{1i} - \dfrac{Q_{1i}}{Q_{2i}} R_{2i}} + \frac{(N_{2i} + S_{2i})(F_i + 1)}{R_{2i} - \dfrac{Q_{2i}}{Q_{1i}} R_{1i}} \right) \times [F]^{-1}\{P\}.$$

References

[1] Varadan, V. K., and Varadan, V. V., 2000, "Microsensors, Microelectromechanical Systems (MEMS), and Electronics for Smart Structures and Systems," Smart Mater. Struct., **9**, pp. 953–972.

[2] Morgan, D. P., 1985, *Surface-Wave Devices for Signal Processing*, Elsevier, Amsterdam.

[3] Morgan, D. P., 1998, "History of SAW Devices," IEEE Int. Frequency Control Symp., pp. 439–460.

[4] Campbell, C. K., 1998, *Surface Acoustic Wave Devices for Mobile and Wireless Communications*, Academic Press, San Diego, CA.

[5] White, R. M., 1998, "Acoustic Sensors for Physical, Chemical, and Biochemical Application," IEEE Int. Frequency Control Symp., pp. 587–594.

[6] Auld, B. A., 1973, *Acoustic Fields and Waves in Solids*, Vol. I, Wiley, New York.

[7] Auld, B. A., 1973b, *Acoustic Fields and Waves in Solids*, Vol. II, Wiley, New York.

[8] Parton, V. Z., and Kudryavtser, B. A., 1988, *Electromagnetoelasticity*, Gordon & Breach, New York.

[9] Badcock, R. A., and Birt, E. A., 2000, "The Use of 0-3 Piezo-Composite Embedded Lamb Wave Sensors for Detection of Damage in Advanced Fibre Composites," Smart Mater. Struct., **9**, pp. 291–297.

[10] Monkhouse, R. S. C., Wilcox, P. W., Dalton, R. P., and Cawley, P., 2000, "The Rapid Monitoring of Structures Using Interdigital Lamb Wave Transducers," Smart Mater. Struct., **9**, pp. 304–309.

[11] Tseng, G. C., 1968, "Frequency Response of an Interdigital Transducers for Excitation of Surface Elastic Waves," IEEE Trans. Electron Devices, **EQ-15**, pp. 586–594.

[12] Coquin, G. A., and Tierstan, T. E., 1967, "Analysis of the Excitation and Detection of Piezoelectric Surface Waves in Quartz by Means of Surface Electrodes." J. Acoust. Soc. Am., **41**, pp. 921–939.

[13] Joshin, S. G., and White, R. M., 1969, "Excitation and Detection of Surface Elastic Waves in Piezoelectric Crystals." J. Acoust. Soc. Am., **46**, pp. 17–27.

[14] Balakrirev, M. K., and Gilinskii, I. A., 1982, *Waves in Piezocrystal*, Nauka Novosibirsk.

[15] Hasegawa, K., and Koshiba, M., 1990, "Finite-Element Solution of Rayleigh-Wave Scattering From Reflective Gratings on a Piezoelectric Substrate," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, **37**, pp. 99–105.

[16] Yong, Y. K., Garon, R., Kanna, S., and Hashimoto, K. Y., 1998, "Effects of Periodically Missing Fingers and Periodically Shifted Fingers on SAW Propagation in Quartz Resonators," IEEE Int. Frequency Control Symp., pp. 461–469.

[17] Xu, G. S., 2000, "Direct Finite-Element Analysis of the Frequency Response of a Y–Z Lithium Niobate SAW Filter," Smart Mater. Struct., **9**, pp. 973–980.

[18] Hashimoto, K. Y., and Yamaguchi, M., 1991, "Derivation of Coupling-of-Modes Parameters for SAW Device Analysis by Means of Boundary Element Method," IEEE Ultrasonics Symp., pp. 21–26.

[19] Huang, F., and Paige, E. G. S., 1988, "The Scattering of Surface Acoustic Waves by Electric Effects in Two-Dimensional Metal Film Structures," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, **35**, pp. 723–735.

[20] Kino, G. S., and Wagers, R. S., 1973, "Theory of Interdigital Couplers on Nonpiezoelectric Substrates," J. Appl. Phys., **44**, pp. 1480–1488.

[21] Engan, H., 1969, "Excitation of Elastic Surface Wave by Spatial Harmonics of Interdigital Transducers," IEEE Trans. Electron Devices, **ED-16**, pp. 1014–1017.

[22] Ogilvy, J. A., 1996, "An Approximate Analysis of Waves in Layered Piezoelectric Plates From an Interdigital Source Transducer," J. Phys. D, pp. 876–884.

[23] Wang, Q., and Varadan, V. K., 2002, "Wave Propagation in Piezoelectric Coupled Plates by Use of Interdigital Transducer, Part 1: Dispersion Characteristics," Int. J. Solids Struct., **39**, pp. 1119–1130.

[24] Wang, Q., and Varadan, V. K., 2002, "Wave Propagation in Piezoelectric Coupled Plates by Use of Interdigital Transducer, Part 2: Wave Excitation by IDT," Int. J. Solids Struct., **39**, pp. 1131–1144.

[25] Viktorov, I. A., 1981, *Surface Waves in Solids*, Nauka, Moscow (in Russian).

[26] Wang, Q., Quek, S. T., and Varadan, V. K., 2001, "Love Wave Propagation in Piezoelectric Coupled Media," Smart Mater. Struct., **10**, pp. 380–388.

[27] Bateman, H. A., and Erdelyi, 1955, *Higher Transcendental Functions*, McGraw–Hill, New York.

[28] Gulyaev, Y. V., and Plesskii, 1977, "Gap Acoustic Waves in Piezoelectric Material," Akust. Zh., **23**, pp. 716–723.

Consistent Formulations of the Interaction Integral Method for Fracture of Functionally Graded Materials

Jeong-Ho Kim[1]

Glaucio H. Paulino[2]
e-mail: paulino@uiuc.edu

Department of Civil and Environmental Engineering,
Newmark Laboratory,
The University of Illinois at Urbana-Champaign,
205 North Mathews Avenue,
Urbana, IL 61801

The interaction integral method provides a unified framework for evaluating fracture parameters (e.g., stress intensity factors and T stress) in functionally graded materials. The method is based on a conservation integral involving auxiliary fields. In fracture of nonhomogeneous materials, the use of auxiliary fields developed for homogeneous materials results in violation of one of the basic relations of mechanics, i.e., equilibrium, compatibility or constitutive, which naturally leads to three independent formulations: "nonequilibrium," "incompatibility," and "constant-constitutive-tensor." Each formulation leads to a consistent form of the interaction integral in the sense that extra terms are added to compensate for the difference in response between homogeneous and nonhomogeneous materials. The extra terms play a key role in ensuring path independence of the interaction integral. This paper presents a critical comparison of the three consistent formulations and addresses their advantages and drawbacks. Such comparison is made both from a theoretical point of view and also by means of numerical examples. The numerical implementation is based on finite elements which account for the spatial gradation of material properties at the element level (graded elements).
[DOI: 10.1115/1.1876395]

1 Introduction

Solid mechanics problems consist of the following three relations:

- *equilibrium*
- *compatibility*
- *constitutive*

To determine fracture parameters, e.g., stress intensity factors (SIFs) and T stress, by means of the interaction integral (M integral[3]) method, *auxiliary fields* such as displacements ($\boldsymbol{u}^{\text{aux}}$), strains ($\boldsymbol{\varepsilon}^{\text{aux}}$), and stresses ($\boldsymbol{\sigma}^{\text{aux}}$) are needed. In fracture of functionally graded materials (FGMs), the use of the auxiliary fields developed for homogeneous materials results in violation of one of the three relations earlier, which leads to three independent formulations (see Fig. 1): nonequilibrium, incompatibility, and constant-constitutive-tensor formulations. Each formulation leads to a different final form of the resulting M integral, and for *consistency*, extra terms are added to compensate for the difference in response between homogeneous and nonhomogeneous materials. Table 1 illustrates the auxiliary fields corresponding to each formulation. Notice that the nonequilibrium formulation satisfies compatibility ($\boldsymbol{\varepsilon}^{\text{aux}}=(\text{sym}\nabla)\boldsymbol{u}^{\text{aux}}$) and the constitutive relations ($\boldsymbol{\sigma}^{\text{aux}}=\boldsymbol{C}(\boldsymbol{x})\boldsymbol{\varepsilon}^{\text{aux}}$), but violates equilibrium ($\nabla\cdot\boldsymbol{\sigma}^{\text{aux}}\neq 0$ with no body forces). The incompatibility formulation satisfies equilibrium and the constitutive relations, but violates compatibility conditions ($\boldsymbol{\varepsilon}^{\text{aux}}\neq(\text{sym}\nabla)\boldsymbol{u}^{\text{aux}}$). The constant-constitutive-tensor formulation satisfies equilibrium and compatibility conditions, but violates the constitutive relations ($\boldsymbol{\sigma}^{\text{aux}}=\boldsymbol{C}_{\text{tip}}\boldsymbol{\varepsilon}^{\text{aux}}$ with $\boldsymbol{C}_{\text{tip}}\neq\boldsymbol{C}(\boldsymbol{x})$). Conservation integrals based on these three consistent formulations are the focus of this paper.

This paper is organized as follows. Section 2 comments on related work. Section 3 presents auxiliary fields for SIFs and T stress. Section 4 provides three consistent formulations using the interaction integral approach. Sections 5 and 6 establish the relationships between M and SIFs and T stress, respectively. Section 7 provides comparison and critical assessment of the three consistent formulations. Sections 8 presents some numerical aspects relevant to the formulations. Section 9 presents two examples, which test different aspects of the formulations. Finally, Sec. 10 concludes this work.

2 Related Work

The interaction integral method is an accurate and robust scheme for evaluating mixed-mode SIFs and T stress. The method is formulated on the basis of conservation laws, which lead to the establishment of a conservation integral for two admissible states of an elastic solid: *actual* and *auxiliary*. Yau et al. [5] presented the interaction integral method for evaluating SIFs in homogeneous isotropic materials. Wang et al. [6] extended the method to homogeneous orthotropic materials, and Yau [7] used the method for bimaterial interface problems.

Recently, the interaction integral method has been explored in the field of fracture of FGMs. It has been extended for evaluating SIFs [8–11] in isotropic FGMs. Dolbow and Gosz [8] employed the extended finite element method (X-FEM); Rao and Rahman [9] used the element-free Galerkin method; and Kim and Paulino [10,11] used the finite element method (FEM). In addition, the

[1]Present address: Department of Civil and Environmental Engineering, The University of Connecticut, 261 Glenbrook Road U-2037, Storrs, CT 06269.

[2]To whom correspondence should be addressed.

[3]Here, the so-called M integral should not be confused with the M integral of Knowles and Sternberg [1], Budiansky and Rice [2], and Chang and Chien [3]. Also, see the book by Kanninen and Popelar [4] for a review of conservation integrals in fracture mechanics.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, February 24, 2003; final revision, July 27, 2004. Associate Editor: K. Ravi-Chandar. Discussion on the paper should be addressed to the Editor, Professor Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

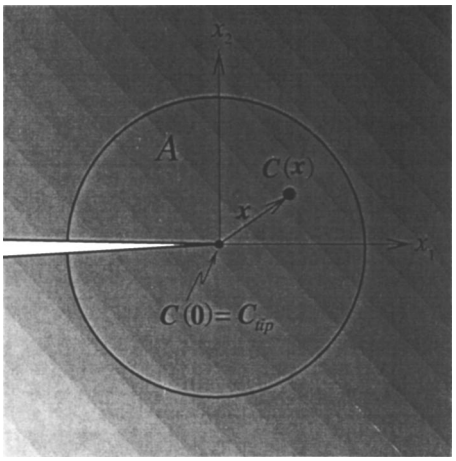


Fig. 1 Motivation for development of alternative consistent formulations. Notice that $C(x) \neq C_{tip}$ for $x \neq 0$. The area A denotes a representative region around the crack tip.

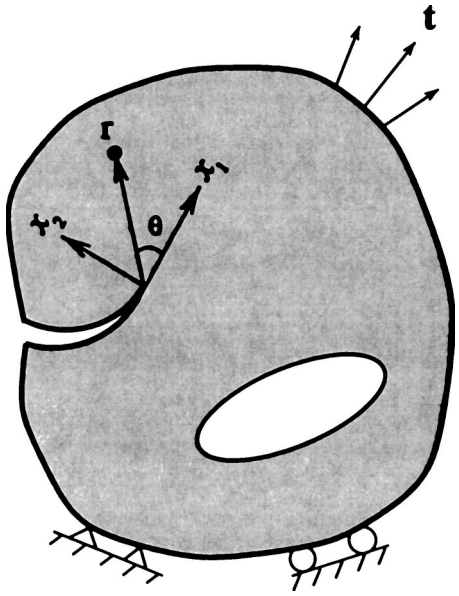


Fig. 2 Cartesian (x_1, x_2) and polar (r, θ) coordinates originating from the crack tip in a nonhomogeneous material subjected to traction (t) and displacement boundary conditions

method has been employed to evaluate T stress in isotropic [11] and orthotropic [12] FGMs. In the aforementioned papers, the interaction integral method has been investigated by means of either an *incompatibility formulation* [8–12] or a *constant-constitutive-tensor formulation* [9]. Thus, for completeness and unification of concepts, this work introduces a *nonequilibrium formulation* for evaluating SIFs and T stress in isotropic and orthotropic FGMs. These three basic formulations (see Sec. 1) will be addressed in this investigation, which includes a critical assessment and comparison of the formulations.

The FEM has been widely used for fracture of FGMs. Eischen [13] evaluated mixed-mode SIFs by means of the path-independent J_k^* integral. Gu et al. [14] evaluated SIFs using the standard J integral. Anlas et al. [15] calculated SIFs by using the path-independent J_1^* integral. Marur and Tippur [16] investigated a crack normal to the material gradient using the FEM in conjunction with experiments. Bao and Cai [17] studied delamination cracking in a graded ceramic/metal substrate under mechanical and thermal loads. Bao and Wang [18] investigated periodic cracking in graded ceramic/metal coatings under mechanical and thermal loads. Kim and Paulino [19] evaluated mixed-mode SIFs by means of the path-independent J_k^* integral, the modified crack closure (MCC), and the displacement correlation technique. Moreover, Kim and Paulino investigated mixed-mode SIFs for cracks arbitrarily oriented in orthotropic FGMs using the MCC method [20] and the path-independent J_k^* integral [21]. The nonsingular stress (T stress) of the Williams's eigenfunction expansion [22] has also been computed by means of the FEM. Becker et al. [23] studied T stress and finite crack kinking in FGMs. They calculated T stress using the difference of the normal stresses along $\theta=0$, i.e., $(\sigma_{xx}-\sigma_{yy})$. Recently, Kim and Paulino [11] proposed a unified approach using the interaction integral method to evaluate T stress and SIFs in FGMs, and also investigated the effect of T stress on crack initiation angles.

Table 1 Comparison of alternative formulations

Nonequilibrium formulation	Incompatibility formulation	Constant-constitutive-tensor formulation
$\boldsymbol{u}^{\text{aux}}$	$\boldsymbol{u}^{\text{aux}}$	$\boldsymbol{u}^{\text{aux}}$
$\boldsymbol{\varepsilon}^{\text{aux}}$	$\boldsymbol{\sigma}^{\text{aux}}$	$\boldsymbol{\varepsilon}^{\text{aux}}$
$\boldsymbol{\sigma}^{\text{aux}}=\boldsymbol{C}(\boldsymbol{x})\boldsymbol{\varepsilon}^{\text{aux}}$	$\boldsymbol{\varepsilon}^{\text{aux}}=\boldsymbol{S}(\boldsymbol{x})\boldsymbol{\sigma}^{\text{aux}}$	$\boldsymbol{\sigma}^{\text{aux}}=\boldsymbol{C}_{\text{tip}}\boldsymbol{\varepsilon}^{\text{aux}}$
$\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}^{\text{aux}}\neq 0$	$\boldsymbol{\varepsilon}^{\text{aux}}\neq(\text{sym}\boldsymbol{\nabla})\boldsymbol{u}^{\text{aux}}$	$\boldsymbol{C}(\boldsymbol{x})\neq\boldsymbol{C}_{\text{tip}}$

Other methods have also been used to investigate fracture of FGMs (see the papers by Erdogan [24], Noda [25], and Paulino et al. [26]). Analytical or semi-analytical approaches have been used by Delale and Erdogan [27], Erdogan [24], Erdogan and Wu [28], and Chan et al. [29]. Delale and Erdogan [30] investigated a crack in a FGM layer between two dissimilar homogeneous half-planes. Gu and Asaro [31] studied a semi-infinite crack in a FGM strip. Shbeeb et al. [32,33] studied multiple cracks interacting in an infinite nonhomogeneous plate. Honein and Herrmann [34] studied conservation laws in nonhomogeneous plane elastostatics and investigated a semi-infinite crack by using the path-independent J_e integral. Gu and Asaro [31] studied orthotropic FGMs considering a four-point bending specimen. Ozturk and Erdogan [35,36] used integral equations to investigate mode I and mixed-mode crack problems in an infinite nonhomogeneous orthotropic medium with a crack aligned with one of the principal material directions. Due to its generality, the FEM is the method of choice in this work.

3 Auxiliary Fields

The interaction integral makes use of auxiliary fields, such as displacements ($\boldsymbol{u}^{\text{aux}}$), strains ($\boldsymbol{\varepsilon}^{\text{aux}}$), and stresses ($\boldsymbol{\sigma}^{\text{aux}}$). These auxiliary fields have to be suitably defined in order to evaluate mixed-mode SIFs and T stress. There are various choices for the auxiliary fields. Here we adopt fields originally developed for homogeneous materials. For each formulation (nonequilibrium, incompatibility, constant-constitutive tensor), the selection of auxiliary fields is done according to Table 1. The auxiliary fields adopted in this paper are described later.

3.1 Fields for SIFs. For evaluating mixed-mode SIFs, we select the auxiliary displacement, strain, and stress fields as the crack-tip asymptotic fields (i.e., $O(r^{1/2})$ for the displacements and $O(r^{-1/2})$ for the strains and stresses) with the material properties sampled at the crack-tip location (e.g., Ref. [13]): Figure 2 shows a crack in a FGM under two-dimensional fields in local Cartesian and polar coordinates originating at the crack tip. The auxiliary displacement, strain, and stress fields are chosen as [22,37]:

$$\boldsymbol{u}^{\text{aux}} = K_I^{\text{aux}}\boldsymbol{f}^{I}(r^{1/2},\theta,\boldsymbol{a}^{\text{tip}}) + K_{II}^{\text{aux}}\boldsymbol{f}^{II}(r^{1/2},\theta,\boldsymbol{a}^{\text{tip}}) \tag{1}$$

$$\boldsymbol{\varepsilon}^{\text{aux}} = (\text{sym}\,\nabla)\boldsymbol{u}^{\text{aux}}, \tag{2}$$

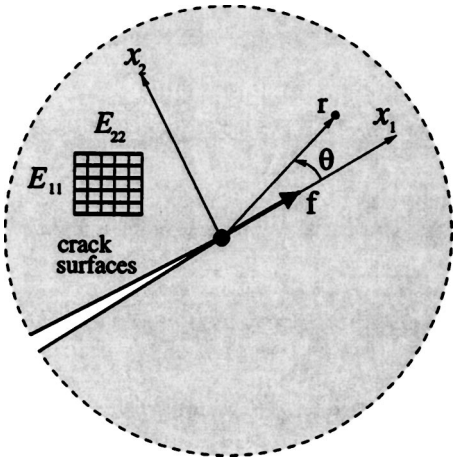


Fig. 3 A point force applied at the crack tip in the direction parallel to the crack surface

$$\boldsymbol{\sigma}^{\text{aux}} = K_I^{\text{aux}}\boldsymbol{g}^I(r^{-1/2},\theta,\boldsymbol{a}^{\text{tip}}) + K_{II}^{\text{aux}}\boldsymbol{g}^{II}(r^{-1/2},\theta,\boldsymbol{a}^{\text{tip}}) \tag{3}$$

where K_I^{aux} and K_{II}^{aux} are the auxiliary mode I and mode II SIFs, respectively, and $\boldsymbol{a}^{\text{tip}}$ denotes contracted notation of the compliance tensor $\boldsymbol{S}$ evaluated at the crack tip, which is explained in Appendix A. The representative functions $\boldsymbol{f}(r^{1/2},\theta,\boldsymbol{a}^{\text{tip}})$ and $\boldsymbol{g}(r^{-1/2},\theta,\boldsymbol{a}^{\text{tip}})$ are given in Appendix B and can also be found in other references, e.g., Refs. [37,38].

3.2 Fields for *T* stress. For evaluating T stress, we choose the auxiliary displacement, strain, and stress fields as those due to a point force in the x_1 direction, applied to the tip of a semi-infinite crack in an infinite homogeneous body as shown in Fig. 3. The auxiliary displacements, strains, and stresses are chosen as [39–41]:

$$\boldsymbol{u}^{\text{aux}} = \boldsymbol{t}^u(\ln r,\theta,f,\boldsymbol{a}^{\text{tip}}) \tag{4}$$

$$\boldsymbol{\varepsilon}^{\text{aux}} = (\text{sym}\,\nabla)\boldsymbol{u}^{\text{aux}} \tag{5}$$

$$\boldsymbol{\sigma}^{\text{aux}} = \boldsymbol{t}^s(r^{-1},\theta,f,\boldsymbol{a}^{\text{tip}}) \tag{6}$$

where f is the point force applied to the crack tip, and $\boldsymbol{a}^{\text{tip}}$ denotes contracted notation of the compliance tensor $\boldsymbol{S}$ evaluated at the crack tip, which is defined in Appendix A. The representative functions $\boldsymbol{t}^u(\ln r,\theta,f,\boldsymbol{a}^{\text{tip}})$ and $\boldsymbol{t}^s(r^{-1},\theta,f,\boldsymbol{a}^{\text{tip}})$ are given in Appendix C and can be found in other references, e.g., Refs. [39,41].

For orthotropic materials, the auxiliary fields may be determined by either the Lekhnitskii or Stroh formalism [12]. There is no difficulty in determining the auxiliary fields in the case of isotropic materials [11].

4 *M*-integral formulations

The standard J integral [42] is given by

$$J = \lim_{\Gamma_s\to 0}\int_{\Gamma_s}(\mathcal{W}\delta_{1j} - \sigma_{ij}u_{i,1})n_j d\Gamma \tag{7}$$

where $\mathcal{W}$ is the strain energy density expressed by

$$\mathcal{W} = \tfrac{1}{2}\sigma_{ij}\varepsilon_{ij} = \tfrac{1}{2}C_{ijkl}\varepsilon_{kl}\varepsilon_{ij} \tag{8}$$

and n_j is the outward normal vector to the contour Γ_s, as shown in Fig. 4. The portion of Γ with applied displacements is denoted Γ_u, and the portion of Γ with applied traction is denoted Γ_τ. Moreover $\Gamma=\Gamma_u+\Gamma_\tau$. Using a plateau-type weight function varying from $q=1$ on Γ_s to $q=0$ on Γ_0 [10] and assuming that the crack faces are traction-free, Eq. (7) becomes

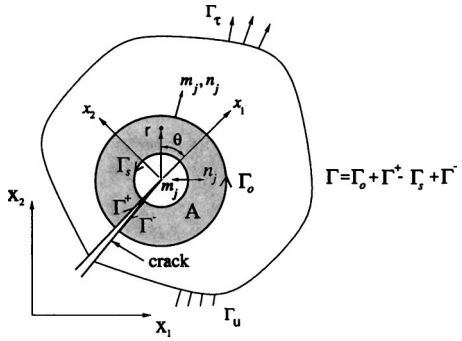


Fig. 4 Conversion of the contour integral into an EDI where $\Gamma=\Gamma_0+\Gamma^+-\Gamma_s+\Gamma^-$, $m_j=n_j$ on Γ_0 and $m_j=-n_j$ on Γ_s

$$J = \lim_{\Gamma_s\to 0}\oint_{\Gamma}(\sigma_{ij}u_{i,1} - \mathcal{W}\delta_{1j})m_j q d\Gamma \tag{9}$$

Applying the divergence theorem to Eq. (9), the equivalent domain integral (EDI) is obtained as

$$J = \int_A(\sigma_{ij}u_{i,1} - \mathcal{W}\delta_{1j})q_{,j}dA + \int_A(\sigma_{ij}u_{i,1} - \mathcal{W}\delta_{1j})_{,j}qdA \tag{10}$$

The J integral of the superimposed fields (actual and auxiliary fields) is obtained as

$$\begin{aligned} J^s = &\int_A\left\{(\sigma_{ij}+\sigma_{ij}^{\text{aux}})(u_{i,1}+u_{i,1}^{\text{aux}}) - \tfrac{1}{2}(\sigma_{ik}+\sigma_{ik}^{\text{aux}})(\varepsilon_{ik}+\varepsilon_{ik}^{\text{aux}})\delta_{1j}\right\}q_{,j}dA \\ &+\int_A\left\{(\sigma_{ij}+\sigma_{ij}^{\text{aux}})(u_{i,1}+u_{i,1}^{\text{aux}}) - \tfrac{1}{2}(\sigma_{ik}+\sigma_{ik}^{\text{aux}})(\varepsilon_{ik}+\varepsilon_{ik}^{\text{aux}})\right. \\ &\left.\times(\delta_{1j})\right\}_{,j}qdA \end{aligned} \tag{11}$$

which is conveniently decomposed into

$$J^s = J + J^{\text{aux}} + M \tag{12}$$

where J^{aux} is given by

$$\begin{aligned} J^{\text{aux}} = &\int_A(\sigma_{ij}^{\text{aux}}u_{i,1}^{\text{aux}} - \mathcal{W}^{\text{aux}}\delta_{1j})q_{,j}dA + \int_A\left\{\sigma_{ij}^{\text{aux}}u_{i,1}^{\text{aux}}\right. \\ &\left.-\tfrac{1}{2}\sigma_{ik}^{\text{aux}}\varepsilon_{ik}^{\text{aux}}\delta_{1j}\right\}_{,j}qdA \end{aligned}$$

and the resulting interaction integral (M) is given by

$$\begin{aligned} M = &\int_A\left\{\sigma_{ij}u_{i,1}^{\text{aux}} + \sigma_{ij}^{\text{aux}}u_{i,1} - \tfrac{1}{2}(\sigma_{ik}\varepsilon_{ik}^{\text{aux}} + \sigma_{ik}^{\text{aux}}\varepsilon_{ik})\delta_{1j}\right\}q_{,j}dA \\ &+\int_A\left\{\sigma_{ij}u_{i,1}^{\text{aux}} + \sigma_{ij}^{\text{aux}}u_{i,1} - \tfrac{1}{2}(\sigma_{ik}\varepsilon_{ik}^{\text{aux}} + \sigma_{ik}^{\text{aux}}\varepsilon_{ik})\delta_{1j}\right\}_{,j}qdA \end{aligned} \tag{13}$$

This general form of M integral becomes a specific form of M integral for each of the three formulations, which is explained in the next section.

4.1 Nonequilibrium Formulation. The name of the formulation is based on the fact that the auxiliary stress field

$$\sigma_{ij}^{\text{aux}} = C_{ijkl}(\boldsymbol{x})\varepsilon_{kl}^{\text{aux}} \tag{14}$$

does not satisfy equilibrium because it differs from

$$\sigma_{ij}^{\text{aux}} = (C_{ijkl})_{\text{tip}} \varepsilon_{kl}^{\text{aux}}, \tag{15}$$

where $C_{ijkl}(\boldsymbol{x})$ is the constitutive tensor of the actual FGM and $(C_{ijkl})_{\text{tip}}$ is the constitutive tensor at the crack tip (see Fig. 1). The derivatives of the auxiliary stress field are

$$\sigma_{ij,j}^{\text{aux}} = C_{ijkl,j}(\boldsymbol{x})\varepsilon_{kl}^{\text{aux}} + C_{ijkl}(\boldsymbol{x})\varepsilon_{kl,j}^{\text{aux}} = \underline{(C_{ijkl})_{\text{tip}}\varepsilon_{kl,j}^{\text{aux}}} + C_{ijkl,j}(\boldsymbol{x})\varepsilon_{kl}^{\text{aux}} + [C_{ijkl}(\boldsymbol{x}) - (C_{ijkl})_{\text{tip}}]\varepsilon_{kl,j}^{\text{aux}}, \tag{16}$$

where the underlined term in Eq. (16) vanishes. Thus this argument confirms that the auxiliary stress field selected in this formulation (Eq. (14)) does not satisfy equilibrium, i.e., $\sigma_{ij,j}^{\text{aux}} \neq 0$ (no body forces or inertia). This choice of the auxiliary fields has been discussed by Dolbow and Gosz [8], but a nonequilibrium formulation was not provided in their paper. The nonequilibrium in the stress field has to be taken into account in the interaction integral formulation, which is discussed in detail later.

Using the following equality:

$$\sigma_{ij}\varepsilon_{ij}^{\text{aux}} = C_{ijkl}(\boldsymbol{x})\varepsilon_{kl}\varepsilon_{ij}^{\text{aux}} = \sigma_{kl}^{\text{aux}}\varepsilon_{kl} = \sigma_{ij}^{\text{aux}}\varepsilon_{ij} \tag{17}$$

one rewrites Eq. (13) as

$$M = \int_A \{\sigma_{ij}u_{i,1}^{\text{aux}} + \sigma_{ij}^{\text{aux}}u_{i,1} - \sigma_{ik}\varepsilon_{ik}^{\text{aux}}\delta_{1j}\}q_{,j}dA + \int_A \{\sigma_{ij}u_{i,1}^{\text{aux}} + \sigma_{ij}^{\text{aux}}u_{i,1} - \sigma_{ik}\varepsilon_{ik}^{\text{aux}}\delta_{1j}\}_{,j}qdA = M_1 + M_2 \tag{18}$$

The last term of the second integral (M_2) in Eq. (18) is expressed as

$$\begin{aligned}(\sigma_{ik}\varepsilon_{ik}^{\text{aux}}\delta_{1j})_{,j} &= (\sigma_{ik}\varepsilon_{ik}^{\text{aux}})_{,1} = (\sigma_{ij}\varepsilon_{ij}^{\text{aux}})_{,1} = (C_{ijkl}\varepsilon_{kl}\varepsilon_{ij}^{\text{aux}})_{,1} \\ &= C_{ijkl,1}\varepsilon_{kl}\varepsilon_{ij}^{\text{aux}} + C_{ijkl}\varepsilon_{kl,1}\varepsilon_{ij}^{\text{aux}} + C_{ijkl}\varepsilon_{kl}\varepsilon_{ij,1}^{\text{aux}} \\ &= C_{ijkl,1}\varepsilon_{kl}\varepsilon_{ij}^{\text{aux}} + \sigma_{ij}^{\text{aux}}\varepsilon_{ij,1} + \sigma_{ij}\varepsilon_{ij,1}^{\text{aux}}\end{aligned} \tag{19}$$

Substitution of Eq. (19) into Eq. (18) leads to

$$M_2 = \int_A (\sigma_{ij,j}u_{i,1}^{\text{aux}} + \sigma_{ij}u_{i,1j}^{\text{aux}} + \sigma_{ij,j}^{\text{aux}}u_{i,1} + \sigma_{ij}^{\text{aux}}u_{i,1j})qdA - \int_A (C_{ijkl,1}\varepsilon_{kl}\varepsilon_{ij}^{\text{aux}} + \sigma_{ij}^{\text{aux}}\varepsilon_{ij,1} + \sigma_{ij}\varepsilon_{ij,1}^{\text{aux}})qdA \tag{20}$$

Using compatibility (actual and auxiliary) and equilibrium (actual) (i.e., $\sigma_{ij,j}=0$ with no body force), one simplifies Eq. (20) as

$$M_2 = \int_A \{\sigma_{ij,j}^{\text{aux}}u_{i,1} - C_{ijkl,1}\varepsilon_{kl}\varepsilon_{ij}^{\text{aux}}\}qdA \tag{21}$$

Therefore the resulting interaction integral (M) becomes

$$M = \int_A \{\sigma_{ij}u_{i,1}^{\text{aux}} + \sigma_{ij}^{\text{aux}}u_{i,1} - \sigma_{ik}\varepsilon_{ik}^{\text{aux}}\delta_{1j}\}q_{,j}dA + \int_A \{\underline{\sigma_{ij,j}^{\text{aux}}u_{i,1}} - C_{ijkl,1}\varepsilon_{kl}\varepsilon_{ij}^{\text{aux}}\}qdA \tag{22}$$

where the underlined term is a nonequilibrium term, which appears due to nonequilibrium of the auxiliary stress fields. The existence of the final form of M integral for FGMs in Eq. (22) has been proved by Kim [43] and Paulino and Kim [44].

4.2 Incompatibilty Formulation. The incompatibility formulation satisfies equilibrium ($\sigma_{ij,j}^{\text{aux}}=0$ with no body forces) and the constitutive relationship ($\varepsilon_{ij}^{\text{aux}}=S_{ijkl}(\boldsymbol{x})\sigma_{kl}^{\text{aux}}$), but violates compatiblity conditions ($\varepsilon_{ij}^{\text{aux}} \neq (u_{i,j}^{\text{aux}}+u_{j,i}^{\text{aux}})/2$). Thus Eq. (20) is also valid for this formulation. Using equilibrium (actual and auxiliary) and compatibility (actual), one simplifies M_2 as

$$M_2 = \int_A \{\sigma_{ij}(u_{i,1j}^{\text{aux}} - \varepsilon_{ij,1}^{\text{aux}}) - C_{ijkl,1}\varepsilon_{kl}\varepsilon_{ij}^{\text{aux}}\}qdA$$

Therefore the resulting interaction integral (M) becomes

$$M = \int_A \{\sigma_{ij}u_{i,1}^{\text{aux}} + \sigma_{ij}^{\text{aux}}u_{i,1} - \sigma_{ik}\varepsilon_{ik}^{\text{aux}}\delta_{1j}\}q_{,j}dA + \int_A \{\underline{\sigma_{ij}(u_{i,1j}^{\text{aux}} - \varepsilon_{ij,1}^{\text{aux}})} - C_{ijkl,1}\varepsilon_{kl}\varepsilon_{ij}^{\text{aux}}\}qdA \tag{23}$$

where the underlined term is an incompatibility term, which appears due to incompatibility of the auxiliary strain fields. The existence of the final form of M integral for FGMs in Eq. (23) has been proved by Kim [43].

4.3 Constant-Constitutive-Tensor Formulation. The constant-constitutive-tensor formulation satisfies equilibrium ($\sigma_{ij,j}^{\text{aux}}=0$ with no body forces) and compatiblity conditions ($\varepsilon_{ij}^{\text{aux}}=(u_{i,j}^{\text{aux}}+u_{j,i}^{\text{aux}})/2$), but violates the constitutive relationship ($\sigma_{ij}^{\text{aux}}=(C_{ijkl})_{\text{tip}}\varepsilon_{kl}^{\text{aux}}$ with $(C_{ijkl})_{\text{tip}} \neq C_{ijkl}(\boldsymbol{x})$). Notice that $\sigma_{ij}\varepsilon_{ij}^{\text{aux}} \neq \sigma_{ij}^{\text{aux}}\varepsilon_{ij}$ due to the violated constitutive relationship. Thus Eq. (13) becomes

$$\begin{aligned}M = &\int_A \left\{\sigma_{ij}u_{i,1}^{\text{aux}} + \sigma_{ij}^{\text{aux}}u_{i,1} - \tfrac{1}{2}(\sigma_{ik}\varepsilon_{ik}^{\text{aux}} + \sigma_{ik}^{\text{aux}}\varepsilon_{ik})\delta_{1j}\right\}q_{,j}dA \\ &+ \int_A \left\{\sigma_{ij,j}u_{i,1}^{\text{aux}} + \sigma_{ij}u_{i,1j}^{\text{aux}} + \sigma_{ij,j}^{\text{aux}}u_{i,1} + \sigma_{ij}^{\text{aux}}u_{i,1j} - \tfrac{1}{2}(\sigma_{ij,1}\varepsilon_{ij}^{\text{aux}} \right. \\ &\left. + \sigma_{ij}\varepsilon_{ij,1}^{\text{aux}} + \sigma_{ij,1}^{\text{aux}}\varepsilon_{ij} + \sigma_{ij}^{\text{aux}}\varepsilon_{ij,1})\right\}qdA\end{aligned} \tag{24}$$

Using equilibrium and compatibility conditions for both actual and auxiliary fields, one obtains M as

$$M = \int_A \left\{\sigma_{ij}u_{i,1}^{\text{aux}} + \sigma_{ij}^{\text{aux}}u_{i,1} - \tfrac{1}{2}(\sigma_{ik}\varepsilon_{ik}^{\text{aux}} + \sigma_{ik}^{\text{aux}}\varepsilon_{ik})\delta_{1j}\right\}q_{,j}dA + \int_A \tfrac{1}{2}\{\sigma_{ij}\varepsilon_{ij,1}^{\text{aux}} - \sigma_{ij,1}\varepsilon_{ij}^{\text{aux}} + \sigma_{ij}^{\text{aux}}\varepsilon_{ij,1} - \sigma_{ij,1}^{\text{aux}}\varepsilon_{ij}\}qdA \tag{25}$$

Notice that the resulting M involves derivatives of the actual strain and stress fields, which arises due to the material mismatch, and may cause loss of accuracy from a numerical point of view. The existence of the final form of M integral for FGMs in Eq. (25) has been proved by Kim [43].

5 Extraction of Stress Intensity Factors

For mixed-mode crack problems on orthotropic materials, the energy release rates $\mathcal{G}_I$ and $\mathcal{G}_{II}$ are related to mixed-mode SIFs as follows [37]:

$$\mathcal{G}_I = -\frac{K_I}{2}a_{22}^{\text{tip}}\,\text{Im}\left[\frac{K_I(\mu_1^{\text{tip}}+\mu_2^{\text{tip}}) + K_{II}}{\mu_1^{\text{tip}}\mu_2^{\text{tip}}}\right] \tag{26}$$

$$\mathcal{G}_{II} = \frac{K_{II}}{2}a_{11}^{\text{tip}}\,\text{Im}[K_{II}(\mu_1^{\text{tip}}+\mu_2^{\text{tip}}) + K_I(\mu_1^{\text{tip}}\mu_2^{\text{tip}})] \tag{27}$$

where Im denotes the imaginary part of the complex function. Thus

$$J_{\text{local}} = \mathcal{G} = \mathcal{G}_I + \mathcal{G}_{II} = c_{11}K_I^2 + c_{12}K_IK_{II} + c_{22}K_{II}^2 \tag{28}$$

where

$$\begin{aligned}c_{11} &= -\frac{a_{22}^{\text{tip}}}{2}\,\text{Im}\left(\frac{\mu_1^{\text{tip}}+\mu_2^{\text{tip}}}{\mu_1^{\text{tip}}\mu_2^{\text{tip}}}\right) \\ c_{12} &= -\frac{a_{22}^{\text{tip}}}{2}\,\text{Im}\left(\frac{1}{\mu_1^{\text{tip}}\mu_2^{\text{tip}}}\right) + \frac{a_{11}^{\text{tip}}}{2}\,\text{Im}(\mu_1^{\text{tip}}\mu_2^{\text{tip}}) \\ c_{22} &= \frac{a_{11}^{\text{tip}}}{2}\,\text{Im}(\mu_1^{\text{tip}}+\mu_2^{\text{tip}})\end{aligned} \tag{29}$$

For two admissible fields, which are the actual ($\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}$) and auxiliary ($\boldsymbol{u}^{\text{aux}}, \boldsymbol{\varepsilon}^{\text{aux}}, \boldsymbol{\sigma}^{\text{aux}}$) fields, one obtains [6]:

$$J^s_{\text{local}} = c_{11}(K_I + K_I^{\text{aux}})^2 + c_{12}(K_I + K_I^{\text{aux}})(K_{II} + K_{II}^{\text{aux}}) + c_{22}(K_{II} + K_{II}^{\text{aux}})^2 = J_{\text{local}} + J^{\text{aux}}_{\text{local}} + M_{\text{local}} \tag{30}$$

where J_{local} is given by Eq. (28), $J^{\text{aux}}_{\text{local}}$ is given by

$$J^{\text{aux}}_{\text{local}} = c_{11}(K_I^{\text{aux}})^2 + c_{12}K_I^{\text{aux}}K_{II}^{\text{aux}} + c_{22}(K_{II}^{\text{aux}})^2 \tag{31}$$

and M_{local} is given by

$$M_{\text{local}} = 2c_{11}K_IK_I^{\text{aux}} + c_{12}(K_IK_{II}^{\text{aux}} + K_I^{\text{aux}}K_{II}) + 2c_{22}K_{II}K_{II}^{\text{aux}} \tag{32}$$

The mode I and mode II SIFs are evaluated by solving the following linear algebraic equations:

$$M^{(1)}_{\text{local}} = 2c_{11}K_I + c_{12}K_{II}, \quad (K_I^{\text{aux}} = 1.0, K_{II}^{\text{aux}} = 0.0) \tag{33}$$

$$M^{(2)}_{\text{local}} = c_{12}K_I + 2c_{22}K_{II}, \quad (K_I^{\text{aux}} = 0.0, K_{II}^{\text{aux}} = 1.0) \tag{34}$$

where the superscript in $M^{(i)}_{\text{local}}$ $(i=1,2)$ is used just to indicate that the values are distinct in each case. For isotropic materials, the off-diagonal terms of c_{ij} drop, and Eqs. (33) and (34) become

$$M^{(1)}_{\text{local}} = \frac{2}{E^*_{\text{tip}}}K_I \quad (K_I^{\text{aux}} = 1.0, K_{II}^{\text{aux}} = 0.0) \tag{35}$$

$$M^{(2)}_{\text{local}} = \frac{2}{E^*_{\text{tip}}}K_{II}, \quad (K_I^{\text{aux}} = 0.0, K_{II}^{\text{aux}} = 1.0) \tag{36}$$

respectively, where $E^*_{\text{tip}} = E_{\text{tip}}$ for plane stress and $E^*_{\text{tip}} = E_{\text{tip}}/(1 - \nu^2_{\text{tip}})$ for plane strain. The relationships of Eqs. (33) and (34), and Eqs. (35) and (36) are the same as those for homogeneous orthotropic [6] and isotropic [5] materials, respectively, except that, for FGMs, the material properties are evaluated at the crack-tip location. Notice that, for the orthotropic case, there is no need for Newton's iteration, which is needed with other approaches such as the path-independent J_k integral [21] and the MCC integral [20]. Here the SIFs for mode I and mode II are naturally decoupled (cf. Eqs. (33) and (34)).

6 Extraction of *T* Stress

T stress can be extracted from the interaction integral by nullifying the contributions of both singular (i.e., $O(r^{-1/2})$) and higher-order (i.e., $O(r^{1/2})$ and higher) terms. The derivation is explained in detail by Kim and Paulino [11,12] and Paulino and Kim [44]. From the earlier derivation of Eq. (13), the *M* integral in the form of line integral is obtained as

$$M_{\text{local}} = \lim_{\Gamma_s \to 0} \int_{\Gamma_s} \{\sigma_{ik}\varepsilon^{\text{aux}}_{ik}\delta_{1j} - \sigma_{ij}u^{\text{aux}}_{i,1} - \sigma^{\text{aux}}_{ij}u_{i,1}\}n_j d\Gamma \tag{37}$$

Here we can consider only the stress parallel to the crack direction, i.e.:

$$\sigma_{ij} = T\delta_{1i}\delta_{1j} \tag{38}$$

Substituting Eq. (38) into Eq. (37), one obtains

$$M_{\text{local}} = -\lim_{\Gamma_s \to 0} \int_{\Gamma_s} \sigma^{\text{aux}}_{ij}n_j u_{i,1} d\Gamma = Ta^{\text{tip}}_{11} \lim_{\Gamma_s \to 0} \int_{\Gamma_s} \sigma^{\text{aux}}_{ij}n_j d\Gamma \tag{39}$$

Because the force f is in equilibrium (see Fig. 3):

$$f = -\lim_{\Gamma_s \to 0} \int_{\Gamma_s} \sigma^{\text{aux}}_{ij}n_j d\Gamma \tag{40}$$

and thus the following relationship is obtained:

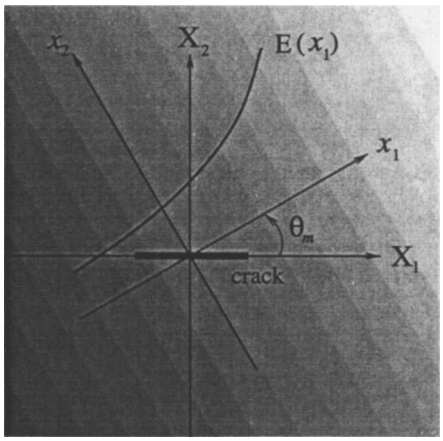


Fig. 5 Crack geometry in a nonhomogeneous material, which is graded along the x_1 direction

$$T = \frac{M_{\text{local}}}{fa^{\text{tip}}_{11}} \tag{41}$$

where a^{tip}_{11} is a material parameter at the crack tip location for plane stress, and is replaced by b^{tip}_{11} for plane strain (cf. Eq. (65)). For isotropic materials, Eq. (41) becomes

$$T = \frac{E^*_{\text{tip}}}{f}M_{\text{local}} \tag{42}$$

where $E^*_{\text{tip}} = E_{\text{tip}}$ for plane stress and $E^*_{\text{tip}} = E_{\text{tip}}/(1-\nu^2_{\text{tip}})$ for plane strain.

7 Comparison and Critical Assessment

The three formulations presented earlier are *consistent* in the sense that extra terms are added to account for the difference in response between homogeneous and nonhomogeneous materials. However, each formulation has an independent final form (see Eqs. (22), (23), and (25)) due to the different characteristics of the auxiliary fields. The final form of the *M* integral for each of these formulations is compared and assessed from a theoretical point of view later.

The nonequilibrium formulation results in the simplest final *M* integral thus requiring the least computation and implementation effort among the three formulations. This is observed by comparing Eqs. (22), (23), and (25). Moreover, the nonequilibrium formulation is equivalent to the incompatibility formulation, because both formulations involve the same constitutive relations and corresponding material derivatives. This equivalence is observed in the numerical examples of Sec. 9. However, the constant-constitutive-tensor formulation [9] requires the derivatives of the actual stress field, which may introduce accuracy problems with standard C^0 elements commonly used in the displacement-based FEM.

In order to further compare the three consistent formulations, let's consider an exponentially graded material in which Poisson's ratio is constant and Young's modulus varies in any direction (see Fig. 5):

$$E(x_1) = E_0 \exp(\delta x_1) = E_0 \exp(\beta_1 X_1 + \beta_2 X_2) \tag{43}$$

$$\nu = \text{constant} \tag{44}$$

where $\boldsymbol{X} = (X_1, X_2)$ refers to a global coordinate system, x_1 is the direction of material gradation (inclined by θ_m with respect to the X_1 coordinate), and the nonhomogeneity parameters δ, β_1, and β_2 are related by

$$\beta_1 = \delta \cos\theta_m, \quad \beta_2 = \delta \sin\theta_m \tag{45}$$

This selection of material property leads to simplification of the resulting M integrals and allows one to better assess and compare the characteristics of the formulations. Moreover, exponentially graded materials have been extensively investigated in the technical literature, e.g., Refs. [8,15,19,21,24,27–36,45–48]. The resulting M integrals corresponding to the three formulations are derived later in the global coordinate system, which is used in the numerical implementation (see Sec. 8 later).

7.1 Nonequilibrium Formulation. The derivatives of interest, with respect to the global coordinate system, are $(m=1,2)$

$$\sigma^{\text{aux}}_{ij,j} = C_{ijkl,j}(\boldsymbol{X})\varepsilon^{\text{aux}}_{kl} + C_{ijkl}(\boldsymbol{X})\varepsilon^{\text{aux}}_{kl,j} = \beta_j C_{ijkl}(\boldsymbol{X})\varepsilon^{\text{aux}}_{kl} + C_{ijkl}(\boldsymbol{X})\varepsilon^{\text{aux}}_{kl,j} = \beta_j C_{ijkl}(\boldsymbol{X})\varepsilon^{\text{aux}}_{kl} + \alpha_p (C_{ijkl})_{\text{tip}}\varepsilon^{\text{aux}}_{kl,j} = \beta_j \sigma^{\text{aux}}_{ij} \tag{46}$$

$$C_{ijkl,m} = \beta_m C_{ijkl}(\boldsymbol{X}) \tag{47}$$

where $\alpha_p = \exp(\beta_1 X_1 + \beta_2 X_2)$ is a factor that arises due to the proportionality of C_{ijkl} for the material gradation considered. The global interaction integral $(M_m)_{\text{global}}$ $(m=1,2)$ is given by

$$(M_m)_{\text{global}} = \int_A \{\sigma_{ij} u^{\text{aux}}_{i,m} + \sigma^{\text{aux}}_{ij} u_{i,m} - \sigma_{ik}\varepsilon^{\text{aux}}_{ik}\delta_{1j}\}\frac{\partial q}{\partial X_j} dA + \int_A \{\sigma^{\text{aux}}_{ij,j} u_{i,m} - C_{ijkl,m}\varepsilon_{kl}\varepsilon^{\text{aux}}_{ij}\} q dA \tag{48}$$

Substitution of Eqs. (46) and (47) into Eq. (48) yields $(m=1,2)$:

$$(M_m)_{\text{global}} = \int_A \{\sigma_{ij} u^{\text{aux}}_{i,m} + \sigma^{\text{aux}}_{ij} u_{i,m} - \sigma_{ik}\varepsilon^{\text{aux}}_{ik}\delta_{mj}\}\frac{\partial q}{\partial X_j} dA + \int_A \{\beta_j \sigma^{\text{aux}}_{ij} u_{i,m} - \beta_m \sigma_{ij}\varepsilon^{\text{aux}}_{ij}\} q dA \tag{49}$$

Notice that, for this particular case, a simpler expression than that for the general case is obtained (cf. Eq. (22)). The derivatives of material properties are represented by the material nonhomogeneity β in Eq. (49). Moreover, the contribution of the nonequilibrium term to the M integral is related to the value of β.

7.2 Incompatibility Formulation. The derivatives of interest, with respect to the global coordinate system, are $(m=1,2)$:

$$\varepsilon^{\text{aux}}_{ij,m} = S_{ijkl,m}(\boldsymbol{X})\sigma^{\text{aux}}_{kl} + S_{ijkl}(\boldsymbol{X})\sigma^{\text{aux}}_{kl,m} = -\beta_m S_{ijkl}(\boldsymbol{X})\sigma^{\text{aux}}_{kl} + S_{ijkl}(\boldsymbol{X})\sigma^{\text{aux}}_{kl,m} = -\beta_m \varepsilon^{\text{aux}}_{ij} + S_{ijkl}(\boldsymbol{X})\sigma^{\text{aux}}_{kl,m} \tag{50}$$

together with Eq. (47). The global interaction integral $(M_m)_{\text{global}}$ $(m=1,2)$ is given by

$$(M_m)_{\text{global}} = \int_A \{\sigma_{ij} u^{\text{aux}}_{i,m} + \sigma^{\text{aux}}_{ij} u_{i,m} - \sigma_{ik}\varepsilon^{\text{aux}}_{ik}\delta_{mj}\}\frac{\partial q}{\partial X_j} dA + \int_A \{\sigma_{ij}(u^{\text{aux}}_{i,mj} - \varepsilon^{\text{aux}}_{ij,m}) - C_{ijkl,m}\varepsilon_{kl}\varepsilon^{\text{aux}}_{ij}\} q dA \tag{51}$$

Substitution of Eqs. (50) and (47) into Eq. (51) yields $(m=1,2)$:

$$(M_m)_{\text{global}} = \int_A \{\sigma_{ij} u^{\text{aux}}_{i,m} + \sigma^{\text{aux}}_{ij} u_{i,m} - \sigma_{ik}\varepsilon^{\text{aux}}_{ik}\delta_{mj}\}\frac{\partial q}{\partial X_j} dA + \int_A \{\sigma_{ij} u^{\text{aux}}_{i,mj} - \sigma^{\text{aux}}_{ij,m}\varepsilon_{ij}\} q dA \tag{52}$$

Notice that, for this particular case, the final M integral does not involve any derivatives of material properties (cf. Eq. (23)). In this formulation, the first integral of Eq. (52) is the same as that for the nonequilibrium formulation, because both formulations use the same constitutive tensor $\boldsymbol{C}(\boldsymbol{X})$.

7.3 Constant-Constitutive-Tensor Formulation. The derivatives of interest, with respect to the global coordinate system, are $(m=1,2)$:

$$\sigma_{ij,m} = C_{ijkl,m}(\boldsymbol{X})\varepsilon_{kl} + C_{ijkl}(\boldsymbol{X})\varepsilon_{kl,m} = \beta_m C_{ijkl}(\boldsymbol{X})\varepsilon_{kl} + C_{ijkl}(\boldsymbol{X})\varepsilon_{kl,m} = \beta_m \sigma_{ij} + C_{ijkl}(\boldsymbol{X})\varepsilon_{kl,m} \tag{53}$$

$$\sigma^{\text{aux}}_{ij,m} = (C_{ijkl})_{\text{tip}}\varepsilon^{\text{aux}}_{kl,m} \tag{54}$$

The global interaction integral $(M_m)_{\text{global}}$ $(m=1,2)$ is given by

$$M = \int_A \left\{\sigma_{ij} u^{\text{aux}}_{i,m} + \sigma^{\text{aux}}_{ij} u_{i,m} - \frac{1}{2}(\sigma_{ik}\varepsilon^{\text{aux}}_{ik} + \sigma^{\text{aux}}_{ik}\varepsilon_{ik})\delta_{mj}\right\}\frac{\partial q}{\partial X_j} dA + \int_A \frac{1}{2}\{\sigma_{ij}\varepsilon^{\text{aux}}_{ij,m} - \sigma_{ij,m}\varepsilon^{\text{aux}}_{ij} + \sigma^{\text{aux}}_{ij}\varepsilon_{ij,m} - \sigma^{\text{aux}}_{ij,m}\varepsilon_{ij}\} q dA \tag{55}$$

Substitution of Eqs. (53) and (54) into Eq. (55) yields $(m=1,2)$:

$$M = \int_A \left\{\sigma_{ij} u^{\text{aux}}_{i,m} + \sigma^{\text{aux}}_{ij} u_{i,m} - \frac{1}{2}(\sigma_{ik}\varepsilon^{\text{aux}}_{ik} + \sigma^{\text{aux}}_{ik}\varepsilon_{ik})\delta_{mj}\right\}\frac{\partial q}{\partial X_j} dA + \int_A \frac{1}{2}\{\sigma_{ij}\varepsilon^{\text{aux}}_{ij,m} - \beta_m \sigma_{ij}\varepsilon^{\text{aux}}_{ij} - C_{ijkl}\varepsilon_{kl,m}\varepsilon^{\text{aux}}_{ij} + \sigma^{\text{aux}}_{ij}\varepsilon_{ij,m} - (C_{ijkl})_{\text{tip}}\varepsilon^{\text{aux}}_{kl,m}\varepsilon_{ij}\} q dA \tag{56}$$

where $C_{ijkl} \equiv C_{ijkl}(\boldsymbol{X})$. Notice that, for this case, the final M integral requires the derivatives of the actual strain field, which may have numerical accuracy problems. The derivatives of material properties are represented by the material nonhomogeneity β in Eq. (56). Moreover, the first integral of Eq. (56) is different from those for the other two formulations.

8 Some Numerical Aspects

For numerical computation by means of the FEM, the M integral is evaluated first in global coordinates ($(M_m)_{\text{global}}$) and then transformed to local coordinates (M_{local}). The M integrals $(M_m)_{\text{global}}$ for the three consistent formulations have derivatives of material properties in common. In this paper, we do not use closed-form expressions for derivatives of material properties because these expressions would be specific to each specific function or micromechanics model. Thus, for the sake of generality, we determine such derivatives by using shape function derivatives of finite elements [19,45].

The derivatives involving material derivatives for each formulation are

- nonequilibrium: $\sigma^{\text{aux}}_{ij,j} = C_{ijkl,j}\varepsilon^{\text{aux}}_{kl} + C_{ijkl}\varepsilon^{\text{aux}}_{kl,j}$ (57)
- incompatibility: $\varepsilon^{\text{aux}}_{ij,m} = S_{ijkl,m}\sigma^{\text{aux}}_{kl} + S_{ijkl}\sigma^{\text{aux}}_{kl,m}$ (58)
- constant-constitutive-tensor: $\sigma_{ij,m} = C_{ijkl,m}\varepsilon_{kl} + C_{ijkl}\varepsilon_{kl,m}$ (59)

A simple and general approach to evaluate such derivatives consists of using shape function derivatives [11]. Thus the derivatives of a generic quantity P (e.g., C_{ijkl}, S_{ijkl}, or ε_{ij}) are obtained as

$$\frac{\partial P}{\partial X_m} = \sum_{i=1}^{n} \frac{\partial N_i}{\partial X_m} P_i, \quad (m=1,2) \tag{60}$$

where n is the number of element nodes and $N_i = N_i(\xi,\eta)$ are the element shape functions which can be found in many references, e.g., Ref. [49]. The derivatives $\partial N_i/\partial X_m$ are obtained as

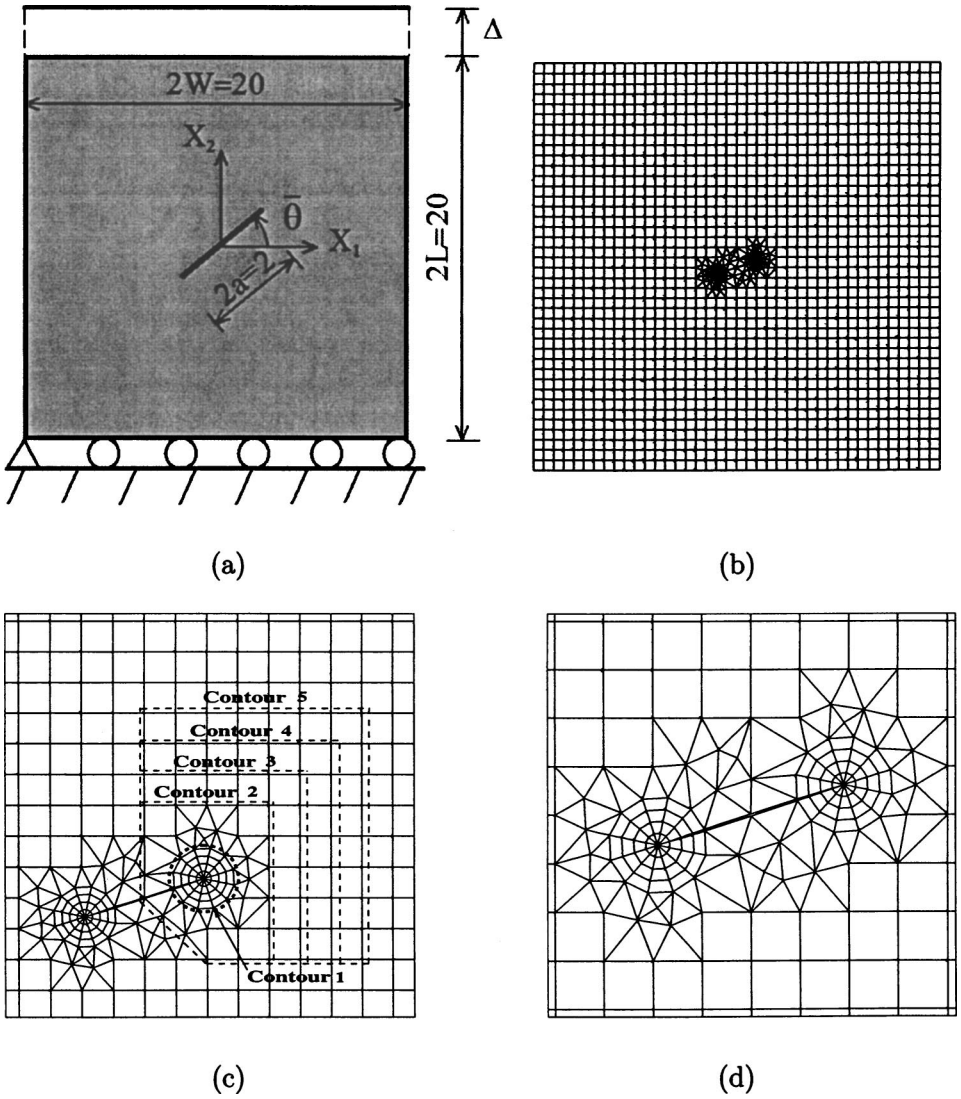


Fig. 6 Example 1: FGM plate with an inclined crack with geometric angle $\bar{\theta}$: (*a*) geometry and boundary conditions (BCs) under fixed-grip loading; (*b*) typical finite element mesh; (*c*) contours for EDI computation of *M* integral; (*d*) mesh detail using 12 sectors (*S*12) and four rings (*R*4) around the crack tips (θ=18° counter-clockwise)

$$\begin{Bmatrix} \partial N_i/\partial X_1 \\ \partial N_i/\partial X_2 \end{Bmatrix} = \boldsymbol{J}^{-1} \begin{Bmatrix} \partial N_i/\partial \xi \\ \partial N_i/\partial \eta \end{Bmatrix} \tag{61}$$

where $\boldsymbol{J}^{-1}$ is the inverse of the standard Jacobian matrix relating (X_1, X_2) with (ξ, η) [49].

9 Numerical Examples

The performance of the interaction integral for evaluating SIFs and *T* stress in isotropic and orthotropic FGMs is examined by means of numerical examples. This paper employs the three formulations, such as nonequilibrium, incompatibility, and constant-constitutive tensor, for numerical investigation. The following examples are presented

(1) Inclined center crack in a plate
(2) Strip with an edge crack

All the examples are analyzed using the FEM code I-FRANC2D[4]. (Illinois; FRacture ANalysis Code 2D), which is based on the code FRANC2D [50,51] developed at Cornell University. The I-FRANC2D element library for FGMs consists of *graded elements* [19,46,45], which incorporate the material gradient at the size scale of the element. The specific graded elements used here are based on the *generalized isoparametric formulation* presented by Kim and Paulino [19], who have also compared the performance of these elements with that of conventional homogeneous elements which produce a step-wise constant approximation to a continuous material property field [45].

All the geometry is discretized with isoparametric graded elements [19]. The specific elements used consist of singular quarter-point six-node triangles (*T6qp*) for crack-tip discretization, eight-node serendipity elements (*Q8*) for a circular region around crack-tip elements, and regular six-node triangles (*T6*) in a transition zone to *Q*8 elements (see, for example, Fig. 6, for a typical crack tip region discretization).

All the examples consist of SIFs and *T* stress results for both isotropic and orthotropic FGMs, and those results are obtained by the interaction integral in conjunction with the FEM. In order to validate SIFs and *T* stress solutions, the FEM results for the first example (an inclined center crack in an exponentially graded plate

[4]The FEM code I-FRANC2D was formerly called FGM-FRANC2D [19].

Table 2 Example 1: comparison of normalized mixed-mode SIFs in *isotropic* FGMs for $\beta a = 0.5$ ($K_0 = \bar{\varepsilon} E^0 \sqrt{\pi a}$) (see Fig. 6). Contour 5 shown in Fig. 6(*c*) is used for the constant-constitutive-tensor formulation. The results for the nonequilibrium and incompatibility formulations are almost identical and thus the results from the latter formulation are not reported here.

Method	$\bar{\theta}$	K_I^+/K_0	K_{II}^+/K_0	K_I^-/K_0	K_{II}^-/K_0
Konda and Erdogan [47]	0°	1.424	0.000	0.674	0.000
	18°	1.285	0.344	0.617	0.213
	36°	0.925	0.548	0.460	0.365
	54°	0.490	0.532	0.247	0.397
	72°	0.146	0.314	0.059	0.269
	90°	0.000	0.000	0.000	0.000
Nonequilibrium	0°	1.4234	0.0000	0.6657	0.0000
	18°	1.2835	0.3454	0.6104	0.2112
	36°	0.9224	0.5502	0.4559	0.3625
	54°	0.4880	0.5338	0.2451	0.3943
	72°	0.1451	0.3147	0.0587	0.2670
	90°	0.0000	0.0000	0.0000	0.0000
Constant-constitutive tensor	0°	1.4262	0.0000	0.6629	0.0000
	18°	1.2807	0.3452	0.6081	0.2101
	36°	0.9224	0.5512	0.4546	0.3607
	54°	0.4862	0.5348	0.2460	0.3931
	72°	0.1439	0.3144	0.0596	0.2670
	90°	0.0000	0.0000	0.0000	0.0000
Dolbow and Gosz [8] (X-FEM)	0°	1.445	0.000	0.681	0.000
	18°	1.303	0.353	0.623	0.213
	36°	0.930	0.560	0.467	0.364
	54°	0.488	0.540	0.251	0.396
	72°	0.142	0.316	0.062	0.268
	90°	0.000	0.000	0.000	0.000

subjected to fixed-grip loading) are compared with available semi-analytical and numerical solutions. The second example involves hyperbolic-tangent functions for material properties and investigates the effect of translation of these properties with respect to the crack-tip location.

9.1 Inclined Center Crack in a Plate. Figure 6(*a*) shows an inclined center crack of length $2a$ located with a geometric angle $\bar{\theta}$ (counter-clockwise) in a plate subjected to fixed-grip loading; Fig. 6(*b*) shows the complete mesh configuration; Fig. 6(*c*) shows five contours used for EDI computation of the *M* integral; and Fig. 6(*d*) shows the mesh detail using 12 sectors (*S12*) and four rings (*R4*) of elements around the crack tips. The displacement boundary condition is prescribed such that $u_2 = 0$ along the lower edge, and $u_1 = 0$ for the node at the lower left-hand side. The mesh discretization consists of 1641 *Q8*, 94 *T6*, and 24 *T6qp* elements, with a total of 1759 elements and 5336 nodes. The fixed-grip loading results in a uniform strain $\varepsilon_{22}(X_1, X_2) = \bar{\varepsilon}$ in a corresponding uncracked structure, which corresponds to $\sigma_{22}(X_1, 10) = \bar{\varepsilon} \bar{E}^0 e^{\beta X_1}$ for isotropic FGMs and $\sigma_{22}(X_1, 10) = \bar{\varepsilon} \bar{E}_{22}^0 e^{\beta X_1}$ for orthotropic FGMs (see Fig. 6(*a*)). Young's moduli and shear modulus are exponential functions of X_1, while Poisson's ratio is constant. The following data were used in the FEM analyses:

plane stress, 2×2 Gauss quadrature

dimensionless nonhomogeneity parameter: $\beta a = 0.5$

$$a/W = 0.1, \quad L/W = 1.0, \quad \bar{\theta} = 0° \text{ to } 90°, \quad \bar{\varepsilon} = 1$$

Isotropic case

$$E(X_1) = E^0 e^{\beta X_1}, \quad \nu(X_1) = \nu$$

$$E^0 = 1.0, \quad \nu = 0.3$$

Orthotropic case

$$E_{11}(X_1) = E_{11}^0 e^{\beta X_1}, \quad E_{22}(X_1) = E_{22}^0 e^{\beta X_1},$$
$$G_{12}(X_1) = G_{12}^0 e^{\beta X_1}, \quad \nu_{12}(X_1) = \nu_{12}^0$$

$$E_{11}^0 = 10^4, \quad E_{22}^0 = 10^3, \quad G_{12}^0 = 1216, \quad \nu_{12}^0 = 0.3$$

Table 2 compares the present FEM results for normalized SIFs obtained by the nonequilibrium and constant-constitutive-tensor formulations of the *M* integral with semi-analytical solutions provided by Konda and Erdogan [47] and the extended FEM results by Dolbow and Gosz [8] for various geometric angles of a crack in isotropic FGMs. The difference in the result for SIFs between nonequilibrium and incompatibility formulations is found to be in the order $O(10^{-4})$ in this example, and thus the results are not provided. The converged results obtained by the nonequilibrium formulation are in good agreement with those by Konda and Erdogan [47] (maximum difference 1.3%, average difference 0.6%), those by Dolbow and Gosz [8], and those obtained by the constant-constitutive-tensor formulation. For the nonequilibrium and incompatibility formulations, a domain including almost half of the square plate is used, and converged solutions are obtained. However, for the constant-constitutive-tensor formulation, contour 5 as shown in Fig. 6(*c*) is used. We observe that the accuracy for the constant-constitutive-tensor formulation are reasonable for small size of contours such as contours 1–5, but as the contour becomes large than contour 5, the solution does not converge, and accuracy deteriorates. As explained in the theoretical discussion, the constant-constitutive-tensor formulation may have numerical problems in the accuracy of derivatives of actual strain or stress fields. To reduce domain dependence, mesh discretization over the plate shown in Fig. 6(*b*) needs to be improved.

Figure 7 shows $J = (K_I^2 + K_{II}^2)/E_{\text{tip}}$ value calculated by the interaction integral for the right crack tip of an inclined crack with $\bar{\theta} = 18$ deg using five contours for EDI computations as shown in Fig. 6(*c*). The nonequilibrium formulation is used both considering and neglecting the nonequilibrium term (see Eq. (22)), and the incompatibility formulation is used both considering and neglect-

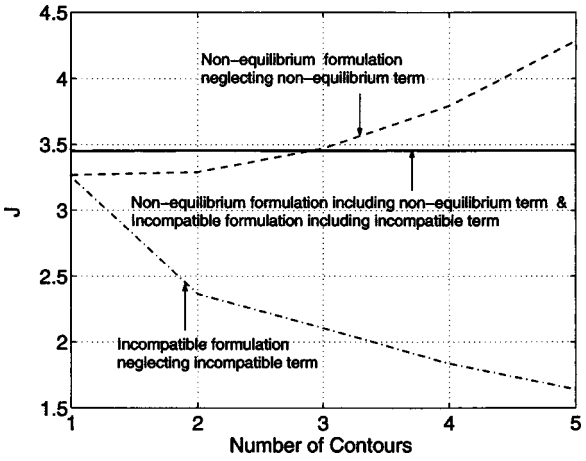


Fig. 7 Example 1: comparison of $J=(K_I^2+K_{II}^2)/E_{tip}$ for the right crack tip of an inclined crack with $\bar{\theta}=18°$ using the *M* integral. The nonequilibrium formulation is used both considering and neglecting the nonequilibrium term (see Eq. (22)). The incompatibility formulation is used both considering and neglecting the incompatible term (see Eq. (23))

ing the incompatible term (see Eq. (23)). The solutions obtained by considering the nonequilibrium term for the nonequilibrium formulation, and the incompatibility term for the incompatibility formulation are not distinguishable in a graphical form. Notice that the converged solution is obtained when including either the nonequilibrium or the incompatibility term, however, such behavior is generally not observed when neglecting either term.

Table 3 compares the present FEM results for normalized SIFs in orthotropic FGMs obtained by the nonequilibrium formulation of the *M* integral with those obtained by the incompatibility formulation for various geometric angles of a crack in orthotropic FGMs. Notice that the two formulations provide similar FEM results for SIFs for each geometric angle. Comparison of Tables 2 and 3 indicates that the material orthotropy shows significant effect on SIFs, and the SIFs K_I^+ (right crack tip) and K_{II}^- (left crack tip) for the orthotropic case are greater than or equal to those for the isotropic case, however, the SIFs K_{II}^+ and K_I^- for the orthotropic case are smaller than or equal to those that for the isotropic case.

Table 4 compares the present FEM results for normalized *T* stress in isotropic FGMs obtained by the nonequilibrium formulation of the *M*-integral with those reported by Paulino and Dong [48] who used the singular integral equation method. Table 5 compares the present FEM results for normalized *T* stress obtained by the nonequilibrium formulation of the *M* integral with those obtained by the incompatibility formulation for orthotropic FGMs. Notice that the two formulations provide similar FEM results for *T* stress for each geometric angle. For the isotropic case, *T* stress at both right and left crack tips changes sign in the range of angle $\bar{\theta}=30$ deg–45 deg (see Table 4), while, for the orthotropic case, it changes sign in the range of angle $\bar{\theta}=15$ deg–30 deg (see Table 5). Comparison of Tables 4 and 5 indicates that the material orthotropy shows significant effect on *T* stress in terms of both sign and magnitude.

Table 4 Example 1: comparison of normalized *T* stress in *isotropic* FGMs for $\beta a=0.5$ ($\sigma_0=\bar{\varepsilon}E^0$) (see Fig. 6)

	Nonequilibrium		Paulino and Dong [48]	
$\bar{\theta}$	$T(+a)/\sigma_0$	$T(-a)/\sigma_0$	$T(+a)/\sigma_0$	$T(-a)/\sigma_0$
0°	−0.896	−0.858	−0.867	−0.876
15°	−0.773	−0.747	−0.748	−0.763
30°	−0.434	−0.436	−0.420	−0.444
45°	0.036	0.011	0.039	0.010
60°	0.513	0.484	0.513	0.490
75°	0.868	0.850	0.870	0.858
90°	0.994	0.994	1.000	1.000

9.2 Strip With an Edge Crack. Figure 8(*a*) shows an edge crack of length "*a*" in a plate, and Fig. 8(*b*) shows the complete mesh discretization using 12 sectors (*S12*) and four rings (*R4*) of elements around the crack tip. Figures 8(*c*)–8(*e*) illustrate the three considered types of hyperbolic-tangent material gradation with respect to the crack tip: reference configuration, translation to the left, and translation to the right, respectively. The fixed-grip displacement loading results in a uniform strain $\varepsilon_{22}(X_1,X_2)=\bar{\varepsilon}$ in a corresponding uncracked structure. The displacement boundary condition is prescribed such that $u_2=0$ along the lower edge and $u_1=0$ for the node at the left-hand side. The mesh discretization consists of 208 *Q8*, 37 *T6*, and 12 *T6qp* elements, with a total of 257 elements and 1001 nodes.

Young's moduli and shear modulus are hyperbolic-tangent functions with respect to the global (X_1,X_2) Cartesian coordinates, while Poisson's ratio is constant (Fig. 9). The following data were used for the FEM analysis:

plane strain, 2×2 Gauss quadrature

$$a/W=0.5,\quad L/W=2.0,\quad \bar{\varepsilon}=0.25,\quad d=(-0.5 \text{ to } 0.5)$$

Isotropic case

$$E(X_1)=(E^-+E^+)/2+\tanh[\beta(X_1+d)](E^--E^+)/2$$

Table 3 Example 1: Comparison of normalized mixed-mode SIFs in *orthotropic* FGMs for $\beta a=0.5$ ($K_0=\bar{\varepsilon}E_{22}^0\sqrt{\pi a}$) (see Fig. 6)

Formulation	$\bar{\theta}$	K_I^+/K_0	K_{II}^+/K_0	K_I^-/K_0	K_{II}^-/K_0
Nonequilibrium	0°	1.4279	0.0000	0.6663	0.0000
	18°	1.3224	0.2176	0.5997	0.2436
	36°	1.0177	0.4097	0.4150	0.4160
	54°	0.6008	0.4477	0.1814	0.4379
	72°	0.2154	0.2906	0.0056	0.2822
	90°	0.0000	0.0000	0.0000	0.0000
Incompatiblity	0°	1.4285	0.0000	0.6663	0.0000
	18°	1.3224	0.2194	0.5997	0.2427
	36°	1.0177	0.4111	0.4149	0.4156
	54°	0.6008	0.4480	0.1809	0.4373
	72°	0.2158	0.2906	0.0052	0.2823
	90°	0.0000	0.0000	0.0000	0.0000

Table 5 Example 1: comparison of normalized *T* stress in *orthotropic* FGMs for $\beta a=0.5$ ($\sigma_0=\bar{\varepsilon}E^0_{22}$) (see Fig. 6)

	Nonequilibrium		Incompatibility	
$\bar{\theta}$	$T(+a)/\sigma_0$	$T(-a)/\sigma_0$	$T(+a)/\sigma_0$	$T(-a)/\sigma_0$
0°	−2.822	−2.725	−2.832	−2.712
15°	−1.407	−1.402	−1.384	−1.407
30°	0.156	0.079	0.168	0.074
45°	0.785	0.700	0.785	0.702
60°	0.971	0.909	0.970	0.910
75°	1.003	0.973	1.002	0.973
90°	0.996	0.996	0.997	0.997

$$\beta a = 15.0, \quad \nu = 0.3$$

$$(E^-, E^+) = (1.00, 3.00)$$

Orthotropic case

$$E_{11}(X_1) = (E^-_{11} + E^+_{11})/2 + \tanh[\alpha(X_1 + d)](E^-_{11} - E^+_{11})/2$$

$$E_{22}(X_1) = (E^-_{22} + E^+_{22})/2 + \tanh[\beta(X_1 + d)](E^-_{22} - E^+_{22})/2$$

$$G_{12}(X_1) = (G^-_{12} + G^+_{12})/2 + \tanh[\gamma(X_1 + d)](G^-_{12} - G^+_{12})/2$$

$$\alpha a = \beta a = \gamma a = 15.0, \quad \nu_{12} = 0.3$$

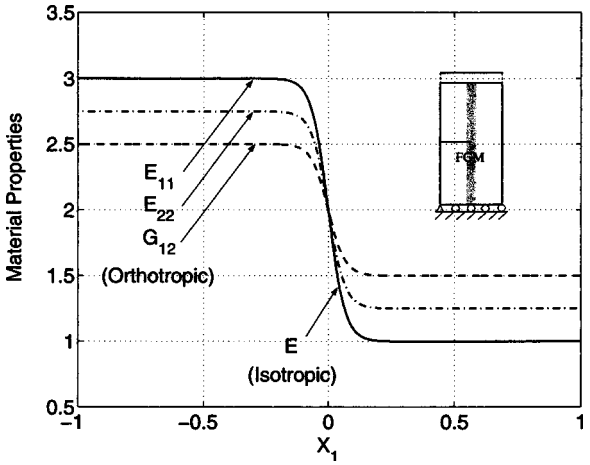


Fig. 9 Example 2: variation of material properties: E_{11}, E_{22}, and G_{12} for the orthotropic case, and E for the isotropic case

$$(E^-_{11}, E^+_{11}) = (1.00, 3.00), \quad (E^-_{22}, E^+_{22}) = (1.25, 2.75),$$
$$(G^-_{12}, G^+_{12}) = (1.50, 2.50)$$

Table 6 compares the present FEM results for mode I SIF (K_I) obtained by the nonequilibrium formulation with those obtained by the incompatibility formulation for various translation factors "d" of hyperbolic-tangent material variation considering both iso-

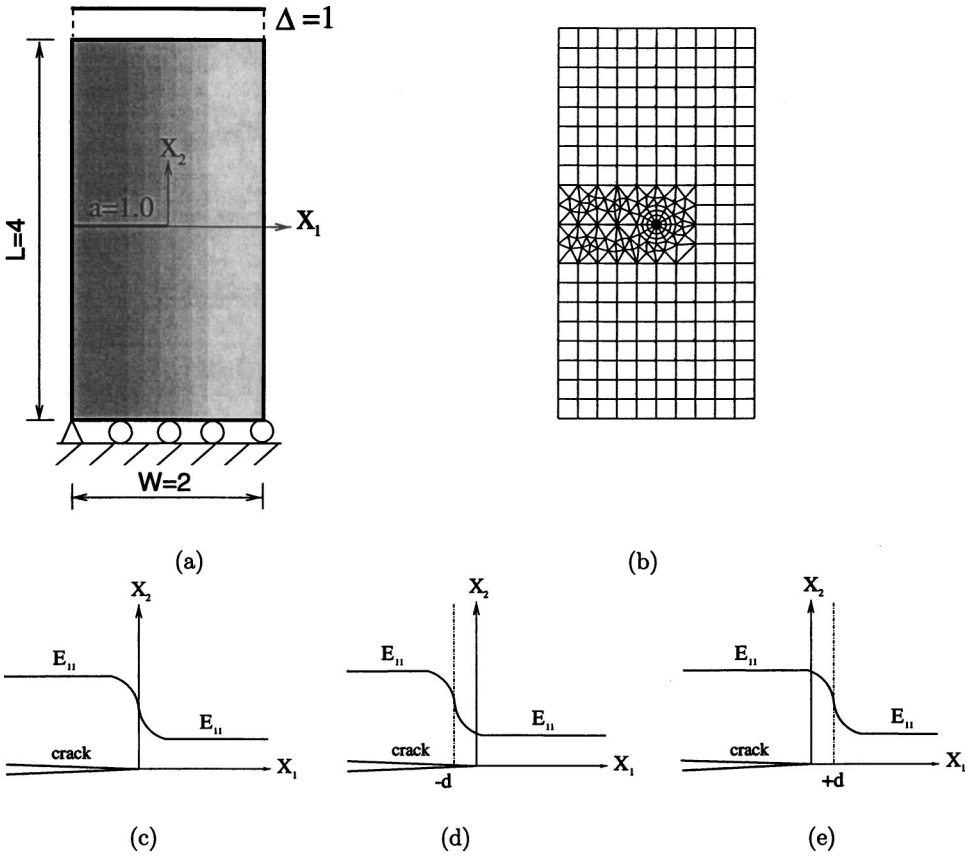


Fig. 8 Example 2: strip with an edge crack in hyperbolic-tangent materials: (*a*) geometry and BCs; (*b*) complete finite element mesh with 12 sectors (*S*12) and four rings (*R*4) around the crack tip; (*c*) reference configuration ($d=0.0$); (*d*) translation of material gradation to the left ($d=+0.5$); (*e*) translation of material gradation to the right ($d=-0.5$)

Table 6 Example 2: comparison of mode I SIF (K_I) for an edge crack considering translation (d) of hyperbolic-tangent material variation (see Fig. 8)

	Nonequilibrium		Incompatibility	
d	Iso	Ortho	Iso	Ortho [12]
−0.5	1.212	1.164	1.186	1.158
−0.4	1.211	1.167	1.201	1.163
−0.3	1.211	1.175	1.190	1.173
−0.2	1.218	1.189	1.209	1.189
−0.1	1.231	1.212	1.212	1.217
0	1.030	1.047	1.026	1.049
0.1	0.595	0.701	0.588	0.697
0.2	0.486	0.615	0.487	0.614
0.3	0.451	0.585	0.451	0.585
0.4	0.430	0.567	0.430	0.567
0.5	0.419	0.554	0.419	0.554

tropic and orthotropic FGMs. For the orthotropic case, the FEM results obtained by the nonequilibrium formulation are compared with those obtained by the incompatibility formulation reported by Kim and Paulino [12]. Notice that the two equivalent formulations provide similar FEM results for mode I SIF for each translation factor d. For the isotropic FGMs, the mode I SIF decreases with the translation factor d for the range between −0.1 and 0.5. For the orthotropic FGMs, the mode I SIF increases with the translation factor d for the range between −0.5 and −0.1, however, it decreases as d increases further. Table 6 also indicates that mode I SIFs for the orthotropic case are smaller than those for the isotropic case for each translation factor d from −0.5 to −0.1, however, the SIFs for the orthotropic case are greater than those for the isotropic case for $d=0$ to 0.5.

Table 7 compares the present FEM results for T stress obtained by the nonequilibrium formulation with those obtained by the incompatibility formulation for various translation factors d of hyperbolic-tangent material variation considering both isotropic and orthotropic FGMs. Notice that the two formulations provide similar FEM results, and the T stresses are negative for all the translation factors d considered. For both isotropic and orthotropic FGMs, the T stress decreases with the translation factor d for the range between −0.5 and 0.0, however, it increases as d increases further. Table 7 also indicates that T stress for the orthotropic case is greater than or equal to that for the isotropic case for each translation factor.

Table 7 Example 2: comparison of T stress for an edge crack considering translation (d) of hyperbolic-tangent material variation (see Fig. 8)

	Nonequilibrium		Incompatibility	
d	Iso	Ortho	Iso	Ortho
−0.5	−0.463	−0.393	−0.452	−0.394
−0.4	−0.478	−0.407	−0.470	−0.406
−0.3	−0.507	−0.434	−0.493	−0.439
−0.2	−0.580	−0.499	−0.571	−0.501
−0.1	−0.797	−0.686	−0.797	−0.702
0	−1.123	−0.923	−1.181	−0.962
0.1	−0.444	−0.364	−0.431	−0.362
0.2	−0.218	−0.205	−0.217	−0.205
0.3	−0.175	−0.171	−0.175	−0.171
0.4	−0.157	−0.157	−0.157	−0.157
0.5	−0.152	−0.151	−0.152	−0.152

10 Conclusions

This paper provides a critical assessment and comparison of three consistent formulations: nonequilibrium, incompatibility, and constant-constitutive-tensor formulations. Each formulation leads to a consistent form of the interaction integral in the sense that extra terms are added to compensate for the difference in response between homogeneous and nonhomogeneous materials. These extra terms play a key role in ensuring path independence of the interaction integral for FGMs. In terms of numerical computations, the nonequilibrium formulation leads to the simplest final form of the M integral among the three formulations. In terms of numerical accuracy, the nonequilibrium formulation is equivalent to the incompatibility formulation, which is observed in numerical examples involving various types of material gradation. The constant-constitutive-tensor formulation requires the derivatives of the actual stress and strain field, and may have numerical accuracy problems with standard C^0 elements commonly used in the displacement-based FEM, as observed in example 1.

From numerical investigations, we observe that both material gradation and orthotropy have a significant influence on SIFs and T stress (i.e., both sign and magnitude), and the crack tip location also shows a significant influence on the fracture parameters in hyperbolic-tangent materials. We also observe that the extra terms (e.g., nonequilibrium or incompatible terms) ensure convergence to target solutions (SIFs or T stress).

Acknowledgments

The authors gratefully acknowledge the support from NASA-Ames, Engineering for Complex Systems Program, and the NASA-Ames Chief Engineer (Dr. Tina Panontin) through Grant No. NAG 2-1424. They also acknowledge additional support from the National Science Foundation (NSF) under Grant No. CMS-0115954 (Mechanics and Materials Program).

Nomenclature

a = half crack length

$\boldsymbol{a}$ or a_{ij} = contracted notation of the compliance tensor ($\boldsymbol{S}$ or S_{ijkl}) for plane stress; $i=1,2,6$; $j=1,2,6$

$\boldsymbol{a}^{\text{tip}}$ or a_{ij}^{tip} = $\boldsymbol{a}$ or a_{ij} evaluated at the crack tip location; $i,j=1,2,6$

$\boldsymbol{A}$ = a 2×2 complex matrix

b_{ij} = contracted notation of the compliance tensor for plane strain; $i=1,2,6$; $j=1,2,6$

b_{ij}^{tip} = b_{ij} evaluated at the crack tip location; $i,j=1,2,6$

$\boldsymbol{B}$ = a 2×2 complex matrix

c_{11}, c_{22}, c_{12} = coefficients in the relationship between J and stress intensity factors (K_I and K_{II})

$\boldsymbol{C}(\boldsymbol{\theta})$ = a 2×2 diagonal matrix

C_{ijkl} or $\boldsymbol{C}$ = constitutive tensor; $i,j,k,l=1,2,3$

d = translation factor in hyperbolic-tangent function

d_0 = x_1 coordinate of a fixed point

e = natural logarithm base, $e=2.71828182\ldots$

E = Young's modulus for isotropic materials

E^0 = Young's modulus E evaluated at the origin

E_{tip} = Young's modulus E evaluated at the crack tip

E_{11}, E_{22} = Young's moduli with respect to the principal axes of orthotropy

E_{11}^0, E_{22}^0 = Young's moduli E_{11},E_{22} evaluated at the origin

f = a point force

$\boldsymbol{f}$ = a 2×1 force vector

$\boldsymbol{f}^I$, $\boldsymbol{f}^{II}$ = representative functions for auxiliary displacements for SIFs

G_{12} = shear modulus for orthotropic materials

G_{12}^0 = shear modulus G_{12} evaluated at the origin
$\mathcal{G}$ = energe release rates
$\boldsymbol{g}^I$, $\boldsymbol{g}^{II}$ = representative functions for auxiliary stresses for SIFs
$\mathcal{G}_I$ = mode I energe release rate
$\mathcal{G}_{II}$ = mode II energe release rate
$\mathcal{H}$ = contour integral
$\boldsymbol{h}$ = a 2×1 real matrix
Im = imaginary part of the complex function
J = path-independent J integral for the actual field
J^{aux} = J integral for the auxiliary field
J^s = J integral for the superimposed fields (actual plus auxiliary)
K_I = mode I stress intensity factor
K_{II} = mode II stress intensity factor
K_0 = normalizing factor for stress intensity factors, $K_0=\overline{\varepsilon}\overline{E}^0\sqrt{\pi a}$ for the isotropic case and $K_0=\overline{\varepsilon}\overline{E}_{22}^0\sqrt{\pi a}$ for the orthotropic case
L = length of a plate
$\boldsymbol{L}$ = a 2×2 real matrix
M = interaction integral (M integral)
N_i = shape functions for node i of an element
$\boldsymbol{N}_3(\boldsymbol{\theta})$ = a 2×2 real matrix
m_i, n_i = unit normal vectors on the contour of the domain integral
P = a generic property (C_{ijkl}, S_{ijkl}, or ε_{ij})
$\boldsymbol{P}(\boldsymbol{\theta})$ = a 2×2 diagonal matrix
p_k = coefficients of the asymptotic displacements for orthotropic materials; $k=1,2$
q_k = coefficients of the asymptotic displacements for orthotropic materials; $k=1,2$
q = weight function in the domain integral
r = radial direction in polar coordinates
Re = real part of the complex function
S_{ijkl} or $\boldsymbol{S}$ = compliance tensor; $i,j,k,l=1,2,3$
$\boldsymbol{S}(\boldsymbol{\theta})$ = a 2×2 real matrix
T = elastic T stress
$\boldsymbol{t}^u$ = representative functions for auxiliary displacements for T stress
$\boldsymbol{t}^s$ = representative functions for auxiliary stresses for T stress
u_i = displacements for the actual field; $i=1,2$
$\boldsymbol{u}^{\text{aux}}$ or u_i^{aux} = a vector for auxiliary displacements; $i=1,2$
W = width of a plate
$\mathcal{W}$ = strain energy density
$\mathcal{W}^{\text{aux}}$ = strain energy density for the auxiliary field
x_i = local Cartesian coordinates; $i=1,2$
X_i = global Cartesian coordinates; $i=1,2$
z_k = complex variable, $z_k=x_k+iy_k$; $k=1,2$
α = material nonhomogeneity parameter for gradation of E_{11}
α_k = the real part of μ_k; $k=1,2$
β = material nonhomogeneity parameter for gradation of E_{22} or E
β_k = the imaginary part of μ_k; $k=1,2$
γ = material nonhomogeneity parameter for gradation of G_{12}
Γ = contour for J and M integrals
Γ_0 = outer contour
Γ_s = inner contour
Γ^+ = contour along the upper crack face
Γ^- = contour along the lower crack face
δ_{ij} = Kronecker delta; $i,j=1,2$
ε_{ij} = strains for the actual field; $i=1,2,3$; $j=1,2,3$
ε_k = contracted notation of ε_{ij}; $k=1,\ldots,6$
$\boldsymbol{\varepsilon}^{\text{aux}}$ or $\varepsilon_{ij}^{\text{aux}}$ = a vector for auxiliary strains; $i,j=1,2,3$
θ = angular direction in polar coordinates with respect to the local Cartesian coordinates
$\overline{\theta}$ = the angle of the local Cartesian coordinates with respect to the global Cartesian coordinates
θ_m = indication of direction of material gradation with respect to the crack
κ = material parameter for isotropic materials; $(3-\nu)/(1+\nu)$ for plane stress and $3-4\nu$ for plane strain
κ^{tip} = material parameter κ evaluated at the crack tip
μ_k = roots of the characteristic equation; $k=1,2$
μ_k^{tip} = μ_k evaluated at the crack tip location; $k=1,2$
$\overline{\mu_k}$ = complex conjugate of μ_k; $k=1,2$
ν = Poisson's ratio for isotropic materials
ν_{12}, ν_{21} = Poisson's ratios for orthotropic materials
σ_k = contracted notation of σ_{ij}; $k=1,\ldots,6$
σ_0 = normalizing factor; $\sigma_0=\overline{\varepsilon}\overline{E}^0$ for the isotropic case $\sigma_0=\overline{\varepsilon}\overline{E}_{22}^0$ for the orthotropic case
σ_{ij} = stresses for the actual fields; $i=1,2,3$; $j=1,2,3$
$\boldsymbol{\sigma}^{\text{aux}}$ or σ_{ij}^{aux} = a vector for auxiliary stresses; $i,j=1,2,3$

Appendix A: Anisotropic Elasticity

The generalized Hooke's law for stress-strain relationship is given by [40]:

$$\varepsilon_i = a_{ij}\sigma_j, \quad a_{ij} = a_{ji}(i,j=1,2,\ldots,6) \tag{A1}$$

where the compliance coefficients, a_{ij}, are contracted notations of the compliance tensor S_{ijkl} and

$$\varepsilon_1=\varepsilon_{11}, \quad \varepsilon_2=\varepsilon_{22}, \quad \varepsilon_3=\varepsilon_{33}, \quad \varepsilon_4=2\varepsilon_{23}, \quad \varepsilon_5=2\varepsilon_{13},$$
$$\varepsilon_6=2\varepsilon_{12}$$

$$\sigma_1=\sigma_{11}, \quad \sigma_2=\sigma_{22}, \quad \sigma_3=\sigma_{33}, \quad \sigma_4=\sigma_{23}, \quad \sigma_5=\sigma_{13}, \quad \sigma_6=\sigma_{12} \tag{A2}$$

For plane stress, the a_{ij} components of interest are

$$a_{ij}(i,j=1,2,6) \tag{A3}$$

and for plane strain, the a_{ij} components are exchanged with b_{ij} as follows:

$$b_{ij} = a_{ij} - \frac{a_{i3}a_{j3}}{a_{33}}(i,j=1,2,6) \tag{A4}$$

Two-dimensional anisotropic elasticity problems can be formulated in terms of the analytic functions, $\phi_k(z_k)$, of the complex variable, $z_k=x_k+iy_k$ $(k=1,2)$, $i=\sqrt{-1}$, where

$$x_k = x+\alpha_k y, \quad y_k=\beta_k y(k=1,2) \tag{A5}$$

The parameters α_k and β_k are the real and imaginary parts of $\mu_k=\alpha_k+i\beta_k$, which can be determined from the following characteristic equation [40]:

$$a_{11}\mu^4 - 2a_{16}\mu^3 + (2a_{12}+a_{66})\mu^2 - 2a_{26}\mu + a_{22} = 0 \tag{A6}$$

where the roots μ_k are always complex or purely imaginary in conjugate pairs as $\mu_1,\overline{\mu_1};\mu_2,\overline{\mu_2}$.

Appendix B: Representative Functions for SIFs

For orthotropic FGMs, the representative functions $\boldsymbol{f}(r^{1/2},\theta,\boldsymbol{a}^{\text{tip}})$ in Eq. (1) are given by [37]:

$$f_1^I = \sqrt{2r/\pi}\,\mathrm{Re}\left[\frac{1}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\{\mu_1^{\mathrm{tip}}p_2\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta} - \mu_2^{\mathrm{tip}}p_1\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}\}\right]$$

$$f_1^{II} = \sqrt{2r/\pi}\,\mathrm{Re}\left[\frac{1}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\{p_2\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta} - p_1\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}\}\right]$$

$$f_2^I = \sqrt{2r/\pi}\,\mathrm{Re}\left[\frac{1}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\{\mu_1^{\mathrm{tip}}q_2\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta} - \mu_2^{\mathrm{tip}}q_1\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}\}\right]$$

$$f_2^{II} = \sqrt{2r/\pi}\,\mathrm{Re}\left[\frac{1}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\{q_2\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta} - q_1\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}\}\right]$$

where Re denotes the real part of the complex function, μ_1^{tip} and μ_2^{tip} denote crack-tip material parameters, which are obtained from Eq. (A6) and taken for $\beta_k > 0$ ($k=1,2$), and p_k and q_k are given by

$$p_k = a_{11}^{\mathrm{tip}}(\mu_k^{\mathrm{tip}})^2 + a_{12}^{\mathrm{tip}} - a_{16}^{\mathrm{tip}}\mu_k^{\mathrm{tip}}$$

$$q_k = a_{12}^{\mathrm{tip}}\mu_k^{\mathrm{tip}} + \frac{a_{22}^{\mathrm{tip}}}{\mu_k^{\mathrm{tip}}} - a_{26}^{\mathrm{tip}} \tag{B1}$$

respectively. The functions $\boldsymbol{g}(r^{-1/2},\theta,\boldsymbol{a}^{\mathrm{tip}})$ in Eq. (3) are given by [37]:

$$g_{11}^I = \frac{1}{\sqrt{2\pi r}}\,\mathrm{Re}\left[\frac{\mu_1^{\mathrm{tip}}\mu_2^{\mathrm{tip}}}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\left\{\frac{\mu_2^{\mathrm{tip}}}{\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta}} - \frac{\mu_1^{\mathrm{tip}}}{\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}}\right\}\right]$$

$$g_{11}^{II} = \frac{1}{\sqrt{2\pi r}}\,\mathrm{Re}\left[\frac{1}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\left\{\frac{(\mu_2^{\mathrm{tip}})^2}{\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta}} - \frac{(\mu_1^{\mathrm{tip}})^2}{\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}}\right\}\right]$$

$$g_{22}^I = \frac{1}{\sqrt{2\pi r}}\,\mathrm{Re}\left[\frac{1}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\left\{\frac{\mu_1^{\mathrm{tip}}}{\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta}} - \frac{\mu_2^{\mathrm{tip}}}{\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}}\right\}\right]$$

$$g_{22}^{II} = \frac{1}{\sqrt{2\pi r}}\,\mathrm{Re}\left[\frac{1}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\left\{\frac{1}{\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta}} - \frac{1}{\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}}\right\}\right]$$

$$g_{12}^I = \frac{1}{\sqrt{2\pi r}}\,\mathrm{Re}\left[\frac{\mu_1^{\mathrm{tip}}\mu_2^{\mathrm{tip}}}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\left\{\frac{1}{\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}} - \frac{1}{\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta}}\right\}\right]$$

$$g_{12}^{II} = \frac{1}{\sqrt{2\pi r}}\,\mathrm{Re}\left[\frac{1}{\mu_1^{\mathrm{tip}}-\mu_2^{\mathrm{tip}}}\left\{\frac{\mu_1^{\mathrm{tip}}}{\sqrt{\cos\theta+\mu_1^{\mathrm{tip}}\sin\theta}} - \frac{\mu_2^{\mathrm{tip}}}{\sqrt{\cos\theta+\mu_2^{\mathrm{tip}}\sin\theta}}\right\}\right] \tag{B2}$$

Notice that, in the earlier expressions, the graded material parameters are sampled at the crack tip.

For isotropic FGMs, the representative functions $\boldsymbol{f}(r^{1/2},\theta,\boldsymbol{a}^{\mathrm{tip}})$ for displacements in Eq. (1), and $\boldsymbol{g}(r^{-1/2},\theta,\boldsymbol{a}^{\mathrm{tip}})$ for stresses in Eq. (3) are given in many references (e.g., Ref. [38]). The graded material parameters are sampled at the crack tip.

Appendix C: Representative Functions for *T* Stress

The presentation follows the Stroh formalism [39]. For othotropic FGMs, the representative functions $\boldsymbol{t}^u(\ln r,\theta,f,\boldsymbol{a}^{\mathrm{tip}})$ in Eq. (4) are given by [39]:

$$t_1^u = -\frac{h_1}{2\pi}\ln r - \frac{1}{2}(S_{11}h_1 + S_{12}h_2)$$

$$t_2^u = -\frac{h_2}{2\pi}\ln r - \frac{1}{2}(S_{21}h_1 + S_{22}h_2) \tag{C1}$$

The parameters S_{ij} and h_i in Eq. (C1) are the components in the 2×2 matrix $\boldsymbol{S}(\theta)$, and the 2×1 vector $\boldsymbol{h}$ as follows:

$$\boldsymbol{S}(\theta) = \frac{2}{\pi}\,\mathrm{Re}[\boldsymbol{A}\boldsymbol{C}(\theta)\boldsymbol{B}^T] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22}\end{bmatrix}$$

$$\boldsymbol{h} = \boldsymbol{L}^{-1}\boldsymbol{f} = \begin{Bmatrix} h_1 \\ h_2\end{Bmatrix} \tag{C2}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \lambda_1^{\mathrm{tip}}p_1^{\mathrm{tip}} & \lambda_2^{\mathrm{tip}}p_2^{\mathrm{tip}} \\ \lambda_1^{\mathrm{tip}}q_1^{\mathrm{tip}} & \lambda_2^{\mathrm{tip}}q_2^{\mathrm{tip}}\end{bmatrix},\quad \boldsymbol{B} = \begin{bmatrix} -\lambda_1^{\mathrm{tip}}\mu_1^{\mathrm{tip}} & -\lambda_2^{\mathrm{tip}}\mu_2^{\mathrm{tip}} \\ \lambda_1^{\mathrm{tip}} & \lambda_2^{\mathrm{tip}}\end{bmatrix}$$

$$\boldsymbol{C}(\theta) = \begin{bmatrix} \ln s_1(\theta) & 0 \\ 0 & \ln s_2(\theta)\end{bmatrix},\quad s_k(\theta) = \cos\theta + \mu_2^{\mathrm{tip}}\sin\theta$$

$$\boldsymbol{L}^{-1} = \mathrm{Re}[i\boldsymbol{A}\boldsymbol{B}^{-1}],\quad \boldsymbol{f} = [f,0]^T \tag{C3}$$

in which p_k^{tip} and q_k^{tip} ($k=1,2$) are given by Eq. (B1), and λ_k^{tip} ($k=1,2$) is the normalization factor given by the expression

$$2(\lambda_k^{\mathrm{tip}})^2(q_k^{\mathrm{tip}}/\mu_k^{\mathrm{tip}} - \mu_k^{\mathrm{tip}}p_k^{\mathrm{tip}}) = 1. \tag{C4}$$

The representative functions $\boldsymbol{t}^s(r^{-1},\theta,f,\boldsymbol{a}^{\mathrm{tip}})$ in Eq. (6) are given by [39]:

$$t_{11}^s = \sigma_{rr}^{\mathrm{aux}}\cos^2\theta,\quad t_{22}^s = \sigma_{rr}^{\mathrm{aux}}\sin^2\theta,\quad t_{12}^s = \sigma_{rr}^{\mathrm{aux}}\sin\theta\cos\theta \tag{C5}$$

where the auxiliary stresses are given by [39]:

$$\sigma_{rr}^{\mathrm{aux}} = \frac{1}{2\pi r}\boldsymbol{n}^T(\theta)\boldsymbol{N}_3(\theta)\boldsymbol{h},\quad \sigma_{\theta\theta}^{\mathrm{aux}} = \sigma_{r\theta}^{\mathrm{aux}} = 0 \tag{C6}$$

in which

$$\boldsymbol{n} = [\cos\theta,\sin\theta]^T,\quad \boldsymbol{N}_3(\theta) = 2\,\mathrm{Re}[\boldsymbol{B}\boldsymbol{P}(\theta)\boldsymbol{B}^T]$$

$$\boldsymbol{P}(\theta) = \begin{bmatrix} \mu_1(\theta) & 0 \\ 0 & \mu_2(\theta)\end{bmatrix},\quad \mu_k(\theta) = \frac{\mu_k^{\mathrm{tip}}\cos\theta - \sin\theta}{\mu_k^{\mathrm{tip}}\sin\theta + \cos\theta},\quad (k=1,2) \tag{C7}$$

For isotropic FGMs, the representative functions $\boldsymbol{t}^u(\ln r,\theta,f,\boldsymbol{a}^{\mathrm{tip}})$ in Eq. (4) for displacements, and $\boldsymbol{t}^s(r^{-1},\theta,f,\boldsymbol{a}^{\mathrm{tip}})$ for stresses in Eq. (6) are given in many references (e.g., Ref. [41]). The graded material parameters are sampled at the crack tip.

References

[1] Knowles, J. K., and Sternberg, E., 1972, "On a Class of Conservation Laws in Linearized and Finite Elastostatics," Arch. Ration. Mech. Anal., **44**(2), pp. 187–211.

[2] Budiansky, B., and Rice, J. R., 1973, "Conservation Laws and Energy-Release Rates," ASME J. Appl. Mech., **40**(1), pp. 201–203.

[3] Chang, J. H., and Chien, A. J., 2002, "Evaluation of M-Integral for Anisotropic Elastic Media With Multiple Defects," Int. J. Fract., **114**(3), pp. 267–289.

[4] Kanninen, M. F., and Popelar, C. H., 1985, *Advanced Fracture Mechanics*, Oxford University Press, New York.

[5] Yau, J. F., Wang, S. S., and Corten, H. T., 1980, "A Mixed-Mode Crack Analysis of Isotropic Solids Using Conservation Laws of Elasticity," ASME J. Appl. Mech., **47**(2), pp. 335–341.

[6] Wang, S. S., Corten, H. T., and Yau, J. F., 1980, "Mixed-Mode Crack Analysis of Rectilinear Anisotropic Solids Using Conservation Laws of Elasticity," Int. J. Fract., **16**(3), pp. 247–259.

[7] Yau, J. F., 1979, "Mixed-Mode Fracture Analysis Using a Conservation Integral," PhD thesis, Department of Theoretical and Applied Mechanics, University of Illinois at Urbana-Champaign.

[8] Dolbow, J., and Gosz, M., 2002, "On the Computation of Mixed-Mode Stress Intensity Factors in Functionally Graded Materials," Int. J. Solids Struct., **39**(9), pp. 2557–2574.

[9] Rao, B. N., and Rahman, S., 2003, "Mesh-Free Analysis of Cracks in Isotropic Functionally Graded Materials," Eng. Fract. Mech., **70**(1), pp. 1–27.

[10] Kim, J.-H., and Paulino, G. H., 2003, "An Accurate Scheme for Mixed-Mode Fracture Analysis of Functionally Graded Materials Using the Interaction Integral and Micromechanics Models," Int. J. Numer. Methods Eng., **58**(10), pp. 1457–1497.

[11] Kim, J.-H., and Paulino, G. H., 2003, "T-Stress, Mixed-Mode Stress Intensity Factors, and Crack Initiation Angles in Functionally Graded Materials: A Unified Approach Using the Interaction Integral Method," Comput. Methods Appl. Mech. Eng., **192**(11–12), pp. 1463–1494.

[12] Kim, J.-H., and Paulino, G. H., 2004, "T-Stress in Orthotropic Functionally Graded Materials: Lekhnitskii and Stroh Formalisms," Int. J. Fract., **126**(4), pp. 345–389.

[13] Eischen, J. W., 1987, "Fracture of Non-Homogeneous Materials," Int. J. Fract., **34**(1), pp. 3–22.

[14] Gu, P., Dao, M., and Asaro, R. J., 1997, "A Simplified Method for Calculating the Crack-Tip Field of Functionally Graded Materials Using the Domain Integral," ASME J. Appl. Mech., **34**(1), pp. 1–17.

[15] Anlas, G., Santare, M. H., and Lambros, J., 2000, "Numerical Calculation of Stress Intensity Factors in Functionally Graded Materials," Int. J. Fract., **104**(2), pp. 131–143.

[16] Marur, P. R., and Tippur, H. V., 2000, "Numerical Analysis of Crack-Tip Fields in Functionally Graded Materials With a Crack Normal to the Elastic Gradient," Int. J. Solids Struct., **37**(38), pp. 5353–5370.

[17] Bao, G., and Cai, H., 1997, "Delamination Cracking in Functionally Graded Coating/Metal Substrate Systems," Acta Mech., **45**(3), pp. 1055–1066.

[18] Bao, G., and Wang, L., 1995, "Multiple Cracking in Functionally Graded Ceramic/Metal Coatings," Int. J. Solids Struct., **32**(19), pp. 2853–2871.

[19] Kim, J.-H., and Paulino, G. H., 2002, "Finite Element Evaluation of Mixed-Mode Stress Intensity Factors in Functionally Graded Materials," Int. J. Numer. Methods Eng., **53**(8), pp. 1903–1935.

[20] Kim, J.-H., and Paulino, G. H., 2002, "Mixed-Mode Fracture of Orthotropic Functionally Graded Materials Using Finite Elements and the Modified Crack Closure Method," Eng. Fract. Mech., **69**(14–16), pp. 1557–1586.

[21] Kim, J.-H., and Paulino, G. H., 2003, "Mixed-Mode J-Integral Formulation and Implementation Using Graded Finite Elements for Fracture Analysis of Nonhomogeneous Orthotropic Materials," Mech. Mater., **35**(1–2), pp. 107–128.

[22] Williams, M. L., 1957, "On the Stress Distribution at the Base of a Stationary Crack," ASME J. Appl. Mech., **24**(1), pp. 109–114.

[23] Becker, T. L., Jr., Cannon, R. M., and Ritchie, R. O., 2001, "Finite Crack Kinking and T-Stresses in Functionally Graded Materials," Int. J. Solids Struct., **38**(32–33), pp. 5545–5563.

[24] Erdogan, F., 1995, "Fracture Mechanics of Functionally Graded Materials," Composites Eng., **5**(7), pp. 753–770.

[25] Noda, N., 1999, "Thermal Stresses in Functionally Graded Materials," J. Therm. Stresses, **22**(4–5), pp. 477–512.

[26] Paulino, G. H., Jin, Z. H., and Dodds, R. H., Jr., 2003, "Failure of Functionally Graded Materials." *Comprehensive Structural Integrity*, B. Karihaloo and W. G. Knauss, eds., Elsevier Science, New York, Vol. 2, Chap. 13, pp. 607–644.

[27] Delale, F., and Erdogan, F., 1983, "The Crack Problem for a Nonhomogeneous Plane," ASME J. Appl. Mech., **50**(3), pp. 609–614.

[28] Erdogan, F, and Wu, B. H., 1997, "The Surface Crack Problem for a Plate With Functionally Graded Properties," ASME J. Appl. Mech., **64**(3), pp. 449–456.

[29] Chan, Y.-S., Paulino, G. H., and Fannjiang, A. C., 2001, "The Crack Problem for Nonhomogeneous Materials Under Antiplane Shear Loading—A Displacement Based Formulation," Int. J. Solids Struct., **38**(17), pp. 2989–3005.

[30] Delale, F., and Erdogan, F., 1988, "On the Mechanical Modeling of an Interfacial Region in Bonded Half-Planes," ASME J. Appl. Mech., **55**(2), pp. 317–324.

[31] Gu, P., and Asaro, R. J., 1997, "Cracks in Functionally Graded Materials," Int. J. Solids Struct., **34**(1), pp. 1–17.

[32] Shbeeb, N. I., Binienda, W. K., and Kreider, K. L., 1999, "Analysis of the Driving Forces for Multiple Cracks in an Infinite Nonhomogeneous Plate, Part I: Theoretical Analysis," ASME J. Appl. Mech., **66**(2), pp. 492–500.

[33] Shbeeb, N. I., Binienda, W. K., and Kreider, K. L., 1999, "Analysis of the Driving Forces for Multiple Cracks in an Infinite Nonhomogeneous Plate, Part II: Numerical Solutions," ASME J. Appl. Mech., **66**(2), pp. 501–506.

[34] Honein, T., and Herrmann, G., 1997, "Conservation Laws in Nonhomogeneous Plane Elastostatics," J. Mech. Phys. Solids, **45**(5), pp. 789–805.

[35] Ozturk, M., and Erdogan, F., 1997, "Mode I Crack Problem in an Inhomogeneous Orthotropic Medium," Int. J. Eng. Sci., **35**(9), pp. 869–883.

[36] Ozturk, M., and Erdogan, F., 1999, "The Mixed Mode Crack Problem in an Inhomogeneous Orthotropic Medium," Int. J. Fract., **98**(3–4), pp. 243–261.

[37] Sih, G. C., Paris, P. C., and Irwin, G. R., 1965, "On Cracks in Rectilinearly Anisotropic Bodies," Int. J. Fract. Mech., **1**(2), pp. 189–203.

[38] Eftis, J., Subramonian, N., and Liebowitz, H., 1977, "Crack Border Atress and Displacement Equations Revisited," Eng. Fract. Mech., **9**(1), pp. 189–210.

[39] Ting, C. T. C., 1996, *Anisotropic Elasticity: Theory and Applications*, Oxford University Press, Oxford.

[40] Lekhnitskii, S. G., 1968, *Anisotropic Plates*, Gordon and Breach Science, New York.

[41] Michell, J. H., 1900, "Elementary Distributions of Plane Stress," Proc. London Math. Soc., **32**, pp. 35–61.

[42] Rice, J. R., 1968, "A Path-Independent Integral and the Approximate Analysis of Strain Concentration By Notches and Cracks," ASME J. Appl. Mech., **35**(2), pp. 379–386.

[43] Kim, J.-H., 2003, "Mixed-Mode Crack Propagation in Functionally Graded Materials," PhD thesis, University of Illinois at Urbana-Champaign.

[44] Paulino, G. H., and Kim, J.-H., 2004, "A New Approach to Compute T-Stress in Functionally Graded Materials Using the Interaction Integral Method," Eng. Fract. Mech., **71**(13–14), pp. 1907–1950.

[45] Kim, J.-H., and Paulino, G. H., 2002, "Isoparametric Graded Finite Elements for Nonhomogeneous Isotropic and Orthotropic Materials," ASME J. Appl. Mech., **69**(4), pp. 502–514.

[46] Santare, M. H., and Lambros, J., 2000, "Use of Graded Finite Elements to Model the Behavior of Nonhomogeneous Materials," ASME J. Appl. Mech., **67**(4), pp. 819–822.

[47] Konda, N., and Erdogan, F., 1994, "The Mixed Mode Crack Problem in a Nonhomogeneous Elastic Medium," Eng. Fract. Mech., **47**(4), pp. 533–545.

[48] Paulino, G. H., and Dong, Z. (unpublished).

[49] Cook, R. D., Malkus, D. S., Plesha, M. E., and Witt, R. J., 2001, *Concepts and Applications of Finite Element Analysis*, 4th ed., Wiley, New York.

[50] Wawrzynek, P. A., 1987, "Interactive Finite Element Analysis of Fracture Processes: An Integrated Approach," MS thesis, Cornell University.

[51] Wawrzynek, P. A., and Ingraffea, A. R., 1991, "Discrete Modeling of Crack Propagation: Theoretical Aspects and Implementation Issues in Two and Three Dimensions," Report 91-5, School of Civil Engineering and Environmental Engineering, Cornell University.

Lewis Thigpen
Department of Mechanical Engineering,
Howard University,
2300 Sixth Street NW,
Washington, D.C. 20059
e-mail: thigpen@scs.howard.edu
Fellow ASME

Patrick T. Reardon
Material Science and Technology Division,
Los Alamos National Laboratory,
P.O. Box 1663, MS P946,
Los Alamos, NM 87545
e-mail: reardon@lanl.gov
Member ASME

Jeremy W. Leggoe
Department of Chemical Engineering,
Texas Tech University,
P.O. Box 43121,
Lubbock, TX 79409-3121
e-mail: Jeremy.Leggoe@ttu.edu

Alan L. Graham
Engineering Sciences and Applications Division,
Los Alamos National Laboratory,
P.O. Box 1663, MS P946,
Los Alamos, NM 87545
e-mail: graham@lanl.gov

Mark Fitzgerald
Department of Mathematics,
University of Colorado at Denver,
P.O. Box 173364, Campus Box 170,
Denver, CO 80217-3364
e-mail: fitz@carbon.cudenver.edu

A Method to Determine the Effect of Microscale Heterogeneities on Macroscopic Web Mechanics

The effect of a spatially heterogeneous density distribution on the development of defects during the transport of nonwoven webs through roller systems has been investigated numerically. A modeling approach has been developed by which the spatial heterogeneity in web mechanical properties can be characterized statistically and recreated for use in finite element simulations. The approach has been applied to model the transport of a carded nonwoven web, consisting of an agglomeration of polypropylene fibers bound together by a regular array of thermal bond points. The web was scanned optically to obtain a gray scale light distribution representing the local material density. Analysis of the local density distribution permitted the generation of "virtual webs" for use in heterogeneous finite element models, in which local mechanical properties were governed by local density. Virtual web response was investigated under two loading configurations; simple tensile testing, and web transport under tension through a three-roller system. The modeling approach provided results that were in good agreement with experimentally observed web mechanics, failure mechanisms, and processing instabilities. Spatial heterogeneity in material properties was found to strongly influence both general web behavior and the tendency for the web to incur manufacturing defects during transport through roller systems. [DOI: 10.1115/1.1876396]

1 Introduction

Knowledge of the relationship between microstructure and macroscopic properties is essential to understanding the mechanical behavior of materials. Nonwoven webs, which are used in the manufacture of a wide range of consumer goods, provide an example where microstructural heterogeneity can significantly influence macroscopic behavior in ways that are not well understood. The fiber laydown processes used to manufacture nonwoven webs make it difficult to precisely control local material density. Understanding the effects of this nonuniform density distribution is of particular importance in the assessment of product manufacturing processes, where the webs are commonly transported under tension through a variety of roller and guide systems.

It is well known that deformation and stress fields within webs during transport strongly influence the incidence of manufacturing defects [1–3]. Spatial heterogeneity may contribute to the initiation of instabilities in the manufacturing process and thereby generate product defects, resulting in processing machine downtime. This investigation explored the critical defect known as "foldover," in which a bending induced trough or wrinkle folds over at a roller and remains trapped during subsequent transport, creating aesthetic or dimensional defects in the finished product. Industrial interest in identifying factors contributing to foldover events, and thus establishing criteria to guide the tailoring of web material properties, motivated the present investigation. The objective was to develop a modeling approach by which the spatial heterogeneity in mechanical properties across the web could be characterized, and then incorporated in finite element simulations in a statistically reproducible manner.

The transport of nonwoven webs has not been widely discussed in the open literature. Nonwoven webs are based on loose assemblies of discrete fibers or continuous filaments, which are consolidated via thermal or chemical bonding, mechanical entanglement, or a combination of these approaches. The fibers may be randomly distributed or preferentially oriented via dynamic combing ("carding") or hydrodynamic methods (Ko and Du [4]). The current investigation considered the behavior of a thermally bonded carded web, in which the polypropylene fibers were preferentially aligned with the longitudinal, or "machine," direction of the web. Consolidation of the web is achieved via "calendaring," a process in which the loose fiber mat is passed through heated rollers, one of which is engraved to yield the desired pattern of bond points. Bonding is achieved by fusing fibers together under the combined effect of pressure and temperature at these bond points. An SEM micrograph depicting the resulting microstructure is presented in Fig. 1.

Understanding the mechanical behavior of these materials is essential to the formulation of novel material and product designs

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, March 11, 2003; final revision, July 20, 2004. Editor: R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

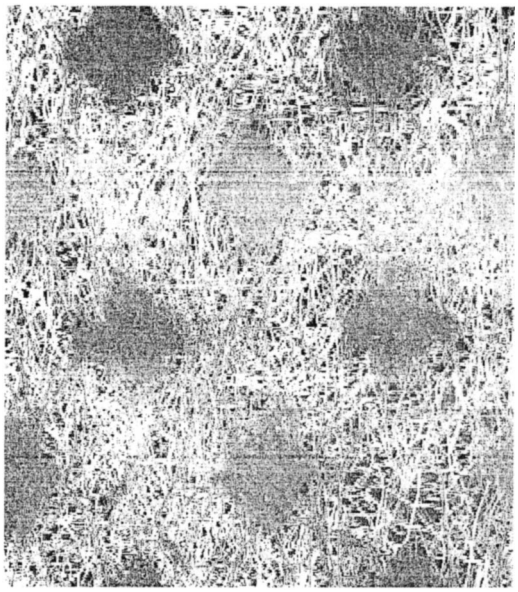


Fig. 1 Secondary electron SEM micrograph illustrating typical web microstructure. The diamond-shaped dark regions are bond points, where fibers have fused together under temperature and pressure during calendaring.

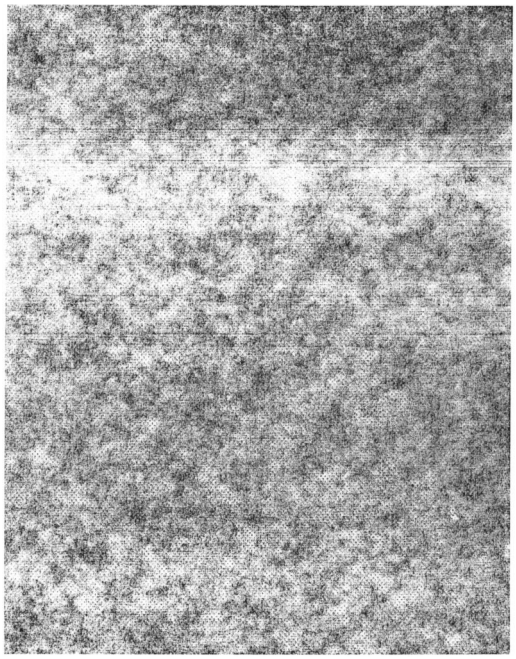


Fig. 2 Gray scale TIFF image resulting from optically scanning a web sample

that reduce product defects, machine downtime, and time to market. In the web mechanics literature, the web is often treated as a homogeneous continuum, neglecting both the effects of discontinuity and the effects of spatial nonuniformity (Stack et al. [5]; Good et al. [6]; Swanson [7]; Lin and Mote [8–10]). The emphasis has been to identify boundary or initial conditions in the transport process that can cause manufacturing defects. The web is typically treated as a beam or plate, and the tendency of the web to develop a defect is related to the stress field.

The deformation and failure of heterogeneous materials have been widely investigated. Materials considered include geologic materials, porous metals, composites, woven and braided textile materials, and foam. Investigations range from phenomenological continuum modeling to investigations at the micromechanical level. The effect of a nonuniform distribution of porosity on flow localization and failure in porous metal alloys was analyzed numerically by Becker [11], and Khvastunkov et al. [12]. Nakamura and Suresh [13] and Brockenbrough et al. [14] used finite element analysis to determine the effects of fiber distribution and local microstructure on deformation and stress in metal-matrix composites. Leggoe et al. [15], and Chen and Mai [16] have carried out three-dimensional finite element analyses to study the effects of microstructural heterogeneity on deformation and effective stress in composites. All of these studies have demonstrated the influence of microstructural heterogeneity on macroscale mechanical behavior, the effects being most pronounced for phenomena that are highly sensitive to localized effects (such as failure).

One of the challenges in simulating the stochastic nature of material properties lies in characterizing the natural spatial heterogeneity of materials. If spatial property distributions were truly random, this would be a relatively trivial matter; once the mean and standard deviation of the variation is determined, the generation of simulated materials having equivalent randomly distributed properties may be accomplished using codes that employ random number generation. Leggoe et al. [15] provide an example of an investigation where this was successfully accomplished. In actual materials, however, there is often an underlying spatial pattern that arises during material synthesis, so that the property distribution must properly be decomposed into two components; the underlying pattern, and a Gaussian "noise" superimposed on the underlying distribution.

This paper describes the development of a methodology for simulating the transport of heterogeneous nonwoven webs through a roller system as part of a manufacturing process, the objective being to identify material and system conditions that lead to product defects and machine downtime. In order to characterize the natural spatial heterogeneity of the carded web of interest, samples of the web were optically scanned to determine the spatial variation in density. Following the technique of Thigpen et al. [17], a covariance function was fit to the images, which was then used to generate simulated "virtual" webs having spatial characteristics equivalent to those of the real material.

The virtual webs were then used to formulate finite element models in which the local mechanical properties of the web were assumed to be governed by the local web density. Models were formulated representing two common loading configurations: (1) conventional tensile testing of rectangular web specimens, and (2) web transport under tension over a three-roller system. The tensile test models revealed that heterogeneous webs were weaker than homogeneous webs of equivalent mean density. In three roller models, the heterogeneous web exhibited behavior that qualitatively matched experimental observations, indicating that spatial heterogeneity in material density may potentially contribute to the initiation of process instability and product defects.

2 Simulation of the Spatial Heterogeneity of Nonwoven Web Fabrics

Creation of the "virtual" web was based on grayscale optical scanning of samples of the carded web. The specimen was scanned against a "black" background, ensuring that only the light reflected by the web itself was returned to the scanner; the recorded grayscale level should therefore reflect the local web density. The scanning process yielded TIFF images of the web, an example of which is provided in Fig. 2. These image files were then converted into text files containing the grayscale levels for each individual pixel.

It is assumed that the web density is a two-dimensional isotropic random field. Analysis of the gray scale level data thus enables

the formulation of a covariance matrix describing the spatial distribution of web density. To create the virtual web, the covariance matrix was fit to an analytical function;

$$C(\mathbf{h}) = \sigma^2 \left\{ 1 + \left(\frac{\|\mathbf{h}\|^2}{b^2} \right) \right\}^{-\beta}, \quad \mathbf{h} \in R^d, \quad \beta > 0, \tag{1}$$

where $b=1.268465$, $\beta=0.540665$, the separation vector between spatial locations $\mathbf{x}_i$ and $\mathbf{x}_j$ on the web is $\mathbf{h}=\mathbf{x}_i-\mathbf{x}_j$, σ^2 is the variance, R^d is the spatial domain, and the number of dimensions $d=2$. The normalized covariance matrix is then obtained by dividing $C(\mathbf{h})$ by the variance σ^2.

The objective is to create a "virtual" web having a spatial distribution of density, $Z(\mathbf{x})$, which is statistically equivalent to the measured characteristics of the actual web. It is assumed that the process to be simulated is Gaussian, is second-order stationary with zero mean, and yields the covariogam $C(\mathbf{h})$. From Eq. (1), $C(\mathbf{0})=\sigma^2$, and must by definition be greater than zero. A process with nonzero mean is easily obtained by adding the required mean to the simulated zero-mean process. The covariance matrix of a second-order stationary stochastic process $Z(\mathbf{x})$ is positive definite and bounded (Karlin and Taylor [18]), so that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |C(\mathbf{h})| d\mathbf{h} < \infty. \tag{2}$$

The covariance matrix defined in Eq. (1) is an even function of $\mathbf{h}$. Therefore, the covariance matrix has the spectral representation

$$C(\mathbf{h}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos(\boldsymbol{\omega}^T \mathbf{h}) s(\boldsymbol{\omega}) d\boldsymbol{\omega} \tag{3}$$

where $s(\boldsymbol{\omega})>0$ is called the spectral function. Dividing by the variance yields the spectral density $s/C(\mathbf{0})$, where $C(\mathbf{0})$ represents the variance (Cressie [19]). The spectral density function of the process, $Z(\mathbf{x})$, is given by

$$s(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(\mathbf{h}) \cos(\boldsymbol{\omega}^T \mathbf{h}) d\mathbf{h}. \tag{4}$$

Shinozuka [20] and Shinozuka and Jan [21] suggested simulating the stochastic process, $Z(\mathbf{x})$, by the discrete cosine transform

$$Z_N(\mathbf{x}) = \sigma \left(\frac{2}{N} \right)^{1/2} \sum_{k=1}^{N} \cos(\omega_{1k} x_1 + \omega_{2k} x_2 + \phi_k) \tag{5}$$

with ϕ_k being independent random variables uniformly distributed between $-\pi$ and π. The random frequencies ω_{1k}, ω_{2k} are distributed according to the joint density function, $g(\omega_1, \omega_2) = s(\omega_1, \omega_2)/\sigma^2$, where

$$\sigma^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(\omega_1, \omega_2) \mathrm{d}\omega_1 \mathrm{d}\omega_2. \tag{6}$$

Shinozuka [20] and Shinozuka and Jan [21] showed that the random process given in Eq. (5) has zero mean and covariance, $C(\mathbf{h})$. As the number of the terms in the series N tends to infinity, Z_N converges to a Gaussian process (Cressie [19]).

To create virtual webs having spatial density distributions statistically equivalent to those derived from the gray scale scan data, the procedure was as follows:

(1) Compute the covariance matrix from the two-dimensional grid of gray scale data using the function given by Eq. (1), and normalized by the variance σ^2.
(2) Perform the inverse (cosine) Fourier Transform on the normalized covariance matrix generated in step (1) to obtain the joint spectral density function, $g(\boldsymbol{\omega})$.
(3) Convolve the square root of the joint spectral density function with cosine transformed Gaussian white noise, and transform to the spatial domain to obtain a random process, $Y_N(\mathbf{x})$, having zero mean and unit variance. Zero padding was necessary to eliminate pollution due to edge effects for a nonperiodic signal.
(4) Scale and location adjust the process, $Y_N(\mathbf{x})$, defined in step (3) to get the final process, $Z_N(\mathbf{x})$, given by

$$Z_N(\mathbf{x}) = \sigma \left(\frac{2}{N} \right)^{1/2} \sum_{k=1}^{N} \cos(\omega_{1k} x_1 + \omega_{2k} x_2 + \phi_k) + \mu \tag{7}$$

where σ is the desired standard deviation and μ is the desired mean of the density distribution.
(5) Output the simulated two-dimensional density distribution.

3 Experimental and Modeling Procedures

3.1 Tensile Testing. Rectangular web specimens were subjected to tensile testing using an Instron 1125 uniaxial loading unit equipped with self aligning grips and fixtures. The specimen ends were attached to aluminum plates using superglue to ensure that the applied load was distributed evenly across the full width of the specimen. In no instance was there any evidence of failure or distortion of the web in the attachment region. Loading was accomplished by displacement of the crosshead at a fixed velocity. The negligible stiffness of the web (compared to the fixtures and grips) enabled the extension of the specimen to be regarded as effectively equal to the crosshead displacement.

All specimens were prepared with an aspect ratio of 4:1, and has their major axis aligned with the longitudinal (machine) direction of the web. Specimens of varying dimension were tested to support future investigations of the effect of scale on the variability of specimen behavior. A series of 80×20 mm specimens were tested with the crosshead speed set to 0.025 m/min, giving an engineering strain rate of 0.0053 s^{-1}. A series of 320×80 mm longitudinal specimens were then tested, with the cross-head speed set to 0.101 m/min (a strain rate of 0.0053 s^{-1}). The typical strain rate experienced by the web during transport through roller systems is expected to be on the order of 10^{-2} s^{-1}.

3.2 Finite Element Model Formulation

3.2.1 Material Properties for Heterogeneous Webs. All finite element computations were performed with the ABAQUS/Explicit software package [22]. The use of finite element models dictated that the discontinuous web be approximated as a continuum. The web was modeled with quadrilateral, finite membrane-finite strain, shell (S4R) elements. Shell element characteristics were defined to ensure that the bending properties of the web (critical in determining the tendency of the web to wrinkle and create foldover defects) were represented accurately in the model.

The web was modeled as being an isotropic, strain-rate insensitive, linear elastic-perfectly plastic material. Tensile testing experiments have confirmed that the web displays some viscoelastic behavior, as expected given that it is an agglomeration of polypropylene fibers. However, in manufacturing processes, the speed of web transport is such that the total transport time is usually too short for relaxation to significantly affect the process. The strains encountered during web transport are generally low enough to ensure that the assumption of linear elastic behavior is reasonable.

The assumption of J_2 plasticity is also a significant approximation. The web is a mat of aligned fibers, and as such deviatoric stresses will actually have little direct effect on plastic deformation. In fact, the web may plastically deform under the influence of direct tensile stresses. The objective in introducing a yield stress was to simulate foldover defects, where some inelastic deformation must occur for the fold become a permanent "crease" in the web. J_2 plasticity represented the most practical method of accomplishing this for continuum elements.

Relating material properties to density was an important ele-

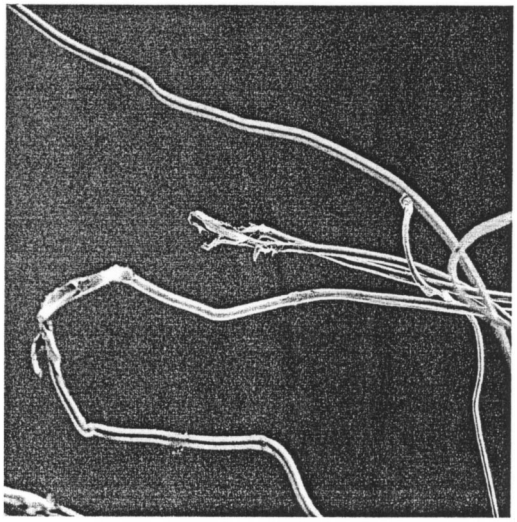


Fig. 3 Secondary electron SEM micrograph showing ruptured fibers in the vicinity of the nonwoven web fracture path

ment of the model formulation process. Experimental results indicate that in the vicinity of the fracture path the web reduces to the thickness of individual fibers, as is evident in the micrograph presented in Fig. 3. This implies that the mechanical properties should depend directly on the relative density of the web material. As the fiber number density increases within a region, the stiffness and yield strength of the web should also increase.

Appropriate relations for the Young's modulus, E, and yield strength, σ_{el}, were sought in the literature. A variety of relations are available for braided (Sun and Qiao [23]), woven (Gao et al. [24] and Scida et al. [25]), cellular (Christensen [26]; Ladd and Kinney [27]; Choi and Lakes [28]), and nonwoven materials (Wang [29]). Drawing from the behavior of cellular materials and experimental observations, it was assumed that E and σ_{el} were density dependent, and represented by the following equations:

$$\frac{E^*}{E_s} = C_1\left(\frac{\rho^*}{\rho_s}\right)^2 \tag{8}$$

$$\frac{\sigma_{el}^*}{E_s} = C_2\left(\frac{\rho^*}{\rho_s}\right)^2 \tag{9}$$

where the superscript $(*)$ denotes local material properties, and subscript (s) denotes the properties of the material at average density.

The values selected for the proportionality constants and physical properties were as follows:

$$C_1 = 1.00$$

$$C_2 = 0.05$$

$$\text{Average density, } \rho_s = 110 \text{ kg/m}^3$$

$$\text{Young's modulus at average density, } E_s = 1.6 \times 10^7 \text{ Pa}$$

$$\text{Poisson's ratio} = 0.4$$

$$\text{Yield strength at average density, } \sigma_{el} = 8.0 \times 10^5 \text{ Pa}.$$

It should be noted that the value selected for the constant C_2 will give a yield strength considerably lower than that observed in longitudinal testing. The purpose in assigning the material elastic-plastic constitutive behavior was to simulate foldover, which depends on bending in the transverse direction. Given that carded webs are significantly less resistant to bending than continuum

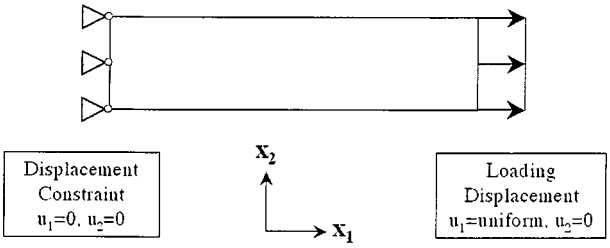


Fig. 4 Configuration and constraints for finite element models simulating tensile testing of heterogeneous nonwoven web specimens

shells of similar thickness, a relatively low average yield stress was sought. Given the relatively low longitudinal strains during transport (and thus the expected linear elastic behavior), this approximation should not adversely affect the response of the model webs during transport though the three roller system that is the focus of this investigation.

Regardless of the model configuration, the following procedure was used to create webs possessing spatially heterogeneous densities for use in finite element analyses;

(1) Apply a regular finite element mesh to a plane rectangular web of the desired dimensions.
(2) Generate a heterogeneous density distribution using the web simulation procedure described previously. The target mean and variance for the model may be derived from analysis of grayscale images or defined by the investigator.
(3) Overlay the heterogeneous density distribution on the finite element grid developed in step (1). The overlay is accomplished by manipulating the property definition section of the text input file used by the finite element analysis package (ABAQUS).
(4) Generate the final model geometry, including the application of any defects in the case of roller system models, using codes developed specifically for each of the models described in the next section.

In the models, the continuous set of real densities was approximated by assigning element densities to one of 65 discrete density values. This limit was imposed due to the necessity of creating a new property definition within the input file for every density that is assigned. Subsequent investigations (Khvastunkov et al. [12]) have indicated that as few as 11–12 distinct values may be sufficient to accurately represent the behavior that would arise in heterogeneous materials given the continuous set of real densities.

3.2.2 Conventional Tensile Test Models. The first loading configuration modeled represented a conventional tensile test, in which a rectangular web specimen is loaded to eventual failure. The model configuration is illustrated schematically in Fig. 4. The length and width of the specimens were selected to match the aspect ratio of the experimental specimens, and fall in the middle of the range of specimen dimensions tested in experiments. The specimen thickness was selected in conjunction with the average elastic modulus to ensure that the specimen exhibited an elastic tensile response approximately equivalent to that of the experimental specimens. The resulting dimensions were as follows:

$$\text{Length } L = 0.16 \text{ m}$$

$$\text{Width } W = 0.04 \text{ m}$$

$$\text{Thickness } T = 0.00025 \text{ m}.$$

The model web was loaded by the incremental application of a uniform displacement in the x-direction to the one end of the web, while the other end of the web was constrained to have zero

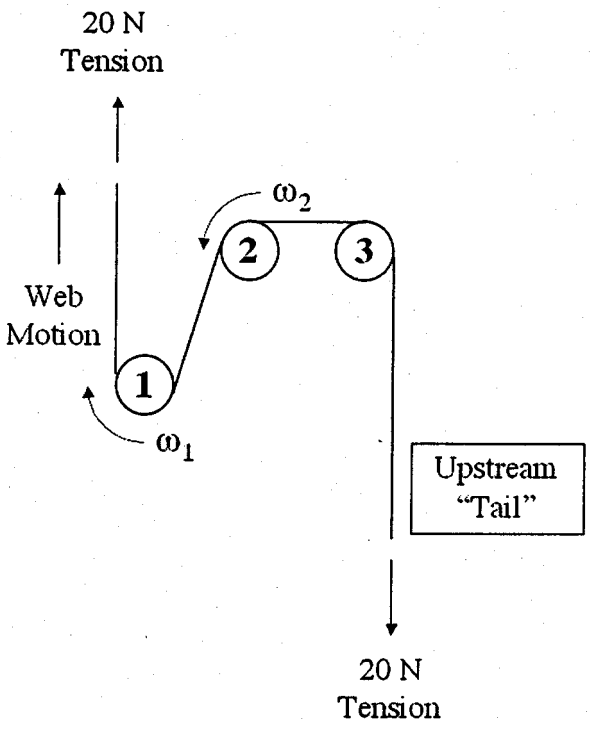


Fig. 5 Schematic illustration of the geometry of a three-roller web transport model. The shaded rollers (1 and 2) are driven at a fixed angular velocities; the unshaded roller (3) is an idler. A tensile load of 20 N is maintained at each end of the web to simulate line conditions.

displacement in the longitudinal (x_1) direction. To recreate test conditions in which the specimens were rigidly glued to aluminum loading plates, both ends of the web were constrained to zero displacement in the transverse (x_2) direction.

3.2.3 Models of Web Transport Over a Misaligned Three-Roller System. The second model configuration represented a web being driven over rollers under tension, as illustrated schematically in Fig. 5. Rollers 1 and 2 were driven, while roller 3 was a free idler (only free idlers could be subjected to misalignment in ABAQUS). The Lagrangian nature of the finite element model requires that the model include a long tail "upstream" of roller 1, in order to enable a reasonable length of the web to pass through the roller system during a model run. The width used in the model corresponded to the typical dimensions of a web used in an example manufacturing process, and the thickness was once again selected to yield the desired tensile and bending response. The resulting dimensions of the web in a three roller model were as follows:

$$\text{Length } L = 8.0 \text{ m}$$

$$\text{Width } W = 0.30 \text{ m}$$

$$\text{Thickness } T = 0.00025 \text{ m}.$$

The finite element grid contained 10 660 elements. There were 30 elements across the width of the web, giving an element aspect ratio of 1.000625. This level of discretization was selected based on trials conducted to determine the number of elements needed to achieve converged results for web troughing under the action of a shearing displacement.

The rollers are rigid, and constrained at the roller center to zero displacements. The driven roller axes were aligned with the x_3 direction. The free roller (3) was allowed to have fixed angular misalignments about both the x_1 and x_2 axes, and was driven by friction only. The friction coefficient between all rollers and the web contact surface is assumed to be 0.80. The roller characteristics are summarized in Table 1.

The web initially leads roller 1 by 0.3 m, and leaves at an angle of −45 deg with respect to the positive x_1 axis. The tail of the web approaches roller 3 at an angle of 90 deg with the with respect to the x_1 axis. In order to create the natural stress and deformation state in the static web, the tensile loads and roller misalignment are ramped in during an initial preload step 0.1 s in duration. The web is ramped up to a tension of 20 Newtons during the preload step. The tensioned web is then ramped up to the line velocity of 5.08 m/s by ramping rollers 1 and 2 up to angular speeds of −25.4 rad/s and 50.8 rad/s, respectively, and ramping the speeds of the ends of the web up to the desired line velocity.

4 Results and Discussion

4.1 Tensile Testing: Experiments and Finite Element Models. The results obtained from tensile testing a set of ten 320×80 mm "longitudinal" specimens (specimens cut so that the major axis coincided with the machine direction of the web) are presented in Fig. 6. The results are plotted in terms of tensile force per unit width; given that the thickness of a nonwoven web is difficult to define, this provides a more fundamental measure of the load supported by the web than would a stress based on an artificially defined thickness. The tensile response is approximately linear at low strains (less than 0.05), and becomes increasingly non-linear with increasing strain. Final failure was relatively abrupt, with little evidence of necking. There is significant scatter in the load to failure and elongation at failure. This is to be expected, given the natural heterogeneity in local density—each specimen gives rise to a unique failure path and location. The 80×20 mm specimens exhibited similar behavior, with an increased degree of scatter in the failure properties, as would be expected given the increased characteristic dimension of the heterogeneity relative to the specimen dimensions (this type of size effect is commonly observed in the failure of ductile alloys [30]).

Finite element models were formulated using material properties drawn from Eqs. (8) and (9). A set of models was analyzed to determine the effect of grid refinement on model response. The results indicated that a mesh containing 3600 elements (30 elements across the width $\times$ 120 elements along the length) yielded converged results. Two calculations were then undertaken in order to evaluate the effects of spatial heterogeneity on web response;

Table 1 Roller characteristics used in finite element models

Roller number	Roller center, m		Radius, m	Mass, kg	State	Over speed
	X_1	X_2				
1	0.0	0.0	0.2	1.00	Driven	1.0
2	0.4	0.2	0.1	10.0	Driven	1.0
3	0.8	0.2	0.1	0.50	Idler	1.0

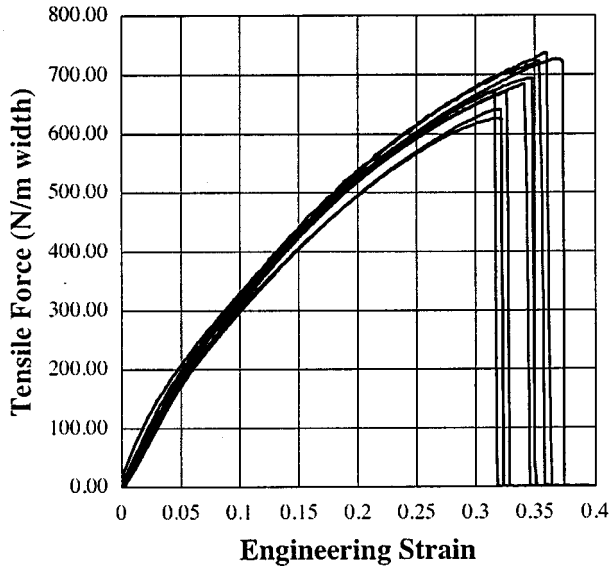


Fig. 6 Tensile test results for 320×80 mm "longitudinal" specimens, with load plotted in units of Newtons per meter specimen width

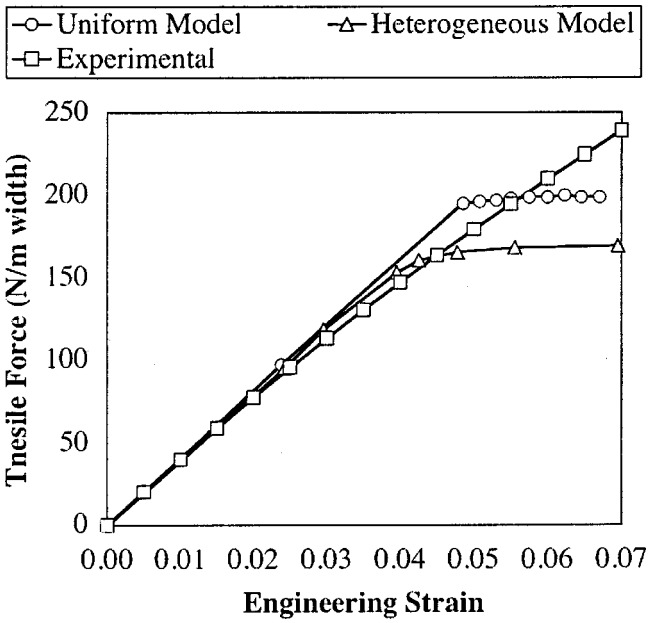


Fig. 7 Comparison of tensile responses of finite element models representing uniform and heterogeneous web specimens with experimental results obtained for "longitudinal" specimens

one model representing a uniform web, and one representing an artificially generated spatially heterogeneous web. The model characteristics are summarized in Table 2. Note that the disparity between the average overall densities in the uniform and heterogeneous models results from the random nature of virtual web creation; each realization of the virtual web will have a unique overall average density.

The load-displacement response for the models is compared with the averaged response of the experimental specimens (both 320×80 and 80×20) in Fig. 7. Although the web material was modeled as a linear elastic, perfectly plastic material, the overall response of both finite element models exhibits slightly nonlinear post yield behavior. For the uniform model, this is a natural result of the constraint imposed on lateral displacement at the ends of the web. For the heterogeneous model, the nonlinearity is more pronounced, and is particularly noticeable at the onset of plastic deformation. The additional effect can be attributed to the variation in element yield stress associated with local density variations; weaker elements yield earlier than the average element, creating nonlinearity as the number of elements participating in plastic deformation gradually increases. The finite element models provide good agreement with the linear elastic portion of the experimental data, indicating that the combination of elastic modulus and web thickness was chosen appropriately. The low yield stress introduced into the models to simulate foldover results in the models yielding at a significantly lower stress than the experimental specimens, as expected.

Examination of the onset of yielding in the heterogeneous specimen can provide insight into the evolution of failure in heterogeneous materials. The heterogeneous specimen exhibits a significantly lower yield strength than the uniform specimen, despite the comparable mean density of the two specimens. The plastic strain contours in Fig. 8(*a*) show that during initial loading, plastic strain initiates in regions of low density, where the material is weak, creating a locally heterogeneous strain field. Although the deformation was not homogeneous, no region within the mesh was excessively distorted at this stage in the deformation. As straining continued (Fig. 8(*b*)), it is evident from the contours of plastic strain that deformation becomes concentrated in a narrow band, as is typically observed in necking processes. Here the contours of low plastic strain are obscured by the extreme plastic strain in the region of strain localization. Continued loading resulted in effective failure of the material within the localization band, and an eventual complete loss of load carrying capacity.

It should be noted that the location and evolution of this localization band is extremely sensitive to the spatial property distribution. During the convergence tests, it was noted that with different property distributions, the localization band emerged at a different location in each specimen. This variation can also lead to significant differences in the overall yield strength predicted by the models, as has previously been observed by Leggoe et al. [15] and Khvastunkov et al. [12].

The deformed mesh for the uniform web is shown in Fig. 9. The necking pattern is typical of the behavior expected from uniform specimens conforming to J_2 plasticity. Plastic deformation can be seen to localize in two symmetrically located regions, corresponding to locations where the constraint imposed by the boundary conditions generates the highest deviatoric stresses. Bands of localized deformation emerge at 45 deg to the loading direction.

The failure patterns in the heterogeneous and uniform models may be compared with the fracture path observed in the 320

Table 2 Parameters for tensile test models

Model	Mean density (kg/m^3)	Maximum density (kg/m^3)	Minimum density (kg/m^3)	Standard deviation (kg/m^3)	Number of elements
Heterogeneous	108.56	131.00	85.00	8.81	3600
Uniform	110.00	110.00	110.00	0.00	3600

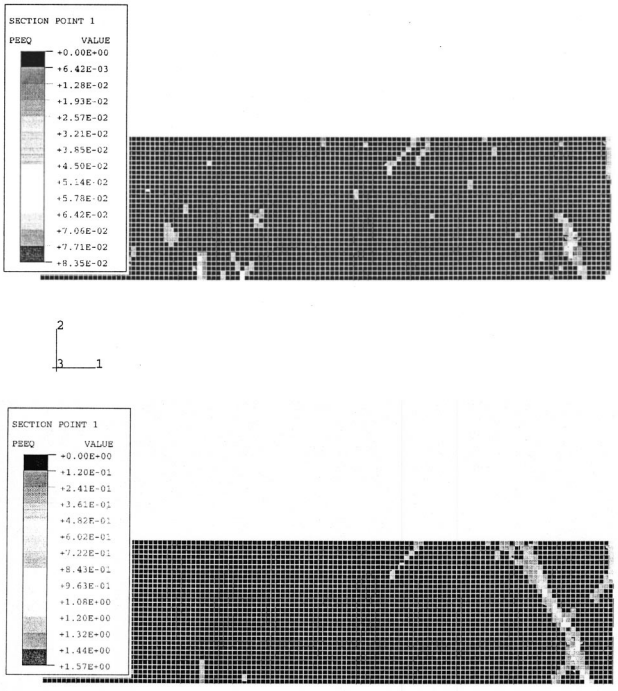


Fig. 8 (a) Plot showing contours of equivalent plastic strain at an applied displacement of 0.00625 m (representing an overall applied strain of 0.02). (b) Plot showing contours of equivalent plastic strain at an applied displacement of 0.0112 m (representing an overall applied strain of 0.035). Final necking of the web has initiated.

$\times 80$ "longitudinal" test specimens illustrated in Fig. 10. Failure in both the web specimen and the finite element model follows an irregular path across the width of the specimen. There are, however, qualitative differences between the model and actual fracture patterns. The actual fracture path is noticeably more jagged than the path in the model, with long sharp steps aligned parallel to the longitudinal direction of the web. These steps most likely correspond to locations where adjacent regions of accumulated damage link up during the final stages of fracture. The alignment of these steps is consistent with the preferred direction of fiber alignment.

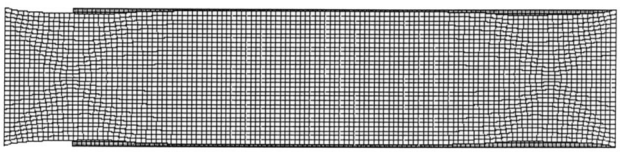


Fig. 9 Deformed mesh for a model representing a uniform web at an applied displacement of 0.0112 (representing an overall applied strain of 0.035). The web is constrained to zero lateral (x_2) displacement at the ends to simulate being glued to loading plates. The shaded web illustrates the undeformed web path.

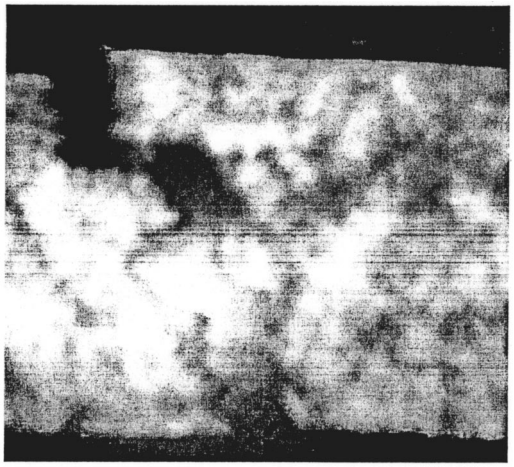


Fig. 10 Fracture path for a 320×80 mm "longitudinal" specimen of the carded nonwoven. The irregular failure path is similar in form to that arising in the heterogeneous web model of Fig. 8(*b*).

These differences may be considered in light of the model formulation. The model was loaded only until the onset of strain localization; the material model did not include a failure criterion. Even if the model had been formulated to continue to final failure, the use of a J_2 plasticity model is likely to have distorted the observed fracture pattern. As a fibrous mat, it is probable that the web will yield in response to direct tensile stresses rather than shearing stresses. The authors have observed in SEM investigations that fibers tend to align with the straining direction as deformation proceeds. With stress being supported solely by the fibers, this renders the use of a J_2 plasticity model to represent large deformations questionable. For future investigations interested in web failure, it would be advisable to formulate a novel material model that more accurately captures the large strain behavior of the web material.

4.2 Web Transport Under Tension Through a Three-Roller System. Five models were formulated to investigate web behavior during transport under tension through a three-roller system, the objective being to examine how spatial heterogeneity and changes in the average material properties contributed to manufacturing defects. The defect introduced was a misalignment of the free roller (idler) about its center point. In the manufacturing processes of interest, a maximum misalignment of 0.0436 rad (2.5 deg) about each axis is maintained; thus the worst case is represented by a simultaneous misalignment of 2.5 deg about each axis.

The individual model characteristics are summarized in Table 3. In the first two models, a uniform web was transported; in model

Table 3 Model parameters for transport of web under tension

Model	Mean density (kg/m^3)	Maximum density (kg/m^3)	Minimum density (kg/m^3)	Standard deviation (kg/m^3)	Angular defect (radians)	
					x_1	x_2
WEBHND	110.00	110.00	110.00	0.00	0.0000	0.0000
WEBHWD	110.00	110.00	110.00	0.00	0.0436	0.0436
WEBRND	88.93	110.00	65.00	9.31	0.0000	0.0000
WEBRWD1	88.93	110.00	65.00	9.31	0.0436	0.0436
WEBRWD2	112.27	145.00	76.00	8.89	0.0436	0.0436

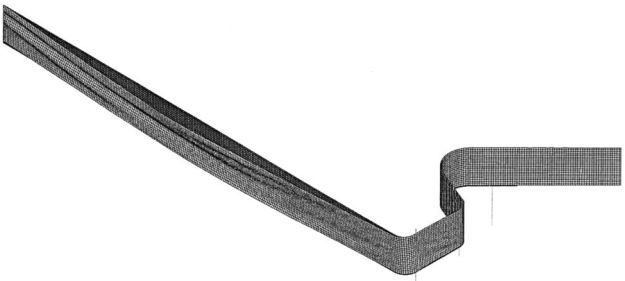


Fig. 11 Deformation of a uniform web during transport through a three-roller system in which the idler roller (roller 3) is misaligned (model WEBHWD) after 1.0 s of web motion. The shaded web illustrates the undeformed web path.

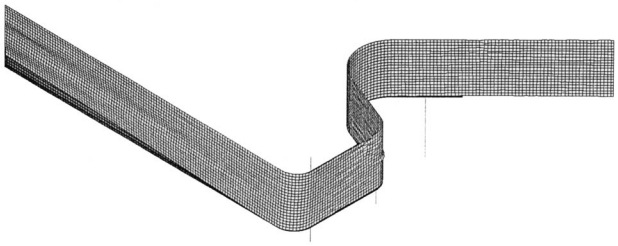


Fig. 12 Deformation of a spatially heterogeneous web during transport through a three-roller system in which the idler roller (roller 3) is misaligned after 0.8 s of web motion, for a web having a mean density of 88.89 kg/m^3 and standard deviation in element density of 9.31 kg/m^3 (model WEBRWD1). The shaded web illustrates the undeformed web path.

WEBHND the rollers were perfectly aligned, while in model WEBHWD a worst case misalignment of roller three about the x_1 and x_2 axes was introduced. Models WEBRND and WEBRWD1 transported a lower density heterogeneous web through systems with and without misalignment (respectively), while model WEBRWD2 transported a heterogeneous web of normal density through a misaligned roller system.

The manufacturing defect of interest was web foldover. A foldover arises when a "wrinkle" (shear-induced trough), reaches a roller and becomes permanently folded upon itself. Once initiated, this defect is propagated along the length of the manufacturing line, and production must be stopped while the defect is corrected. Roller misalignment is one of the principal causes of foldover.

The uniform web in the perfectly aligned model WEBHND did not undergo plastic deformation during the transport process, and no wrinkling occurred. The web more or less followed the ideal path described by the undeformed mesh through the roller system. The introduction of a roller misalignment significantly affected the web, as illustrated in Fig. 11. Distinct ridges developed in the trailing end of the web (to the left side of Fig. 11) during the initial pretensioning step. As the web accelerated to the target velocity, the web oscillated due to the misalignment of roller 3, though no wrinkling developed downstream of roller 3 and no foldover arose.

The results for the perfectly aligned heterogeneous web model WEBRND were similar to those obtained for the uniform web, and once again no defects arose. Small isolated areas of plastic deformation three or four elements wide occurred during the application of the tensile load. As the web ramped up in velocity during the second stage of loading, a few additional pockets of plastic deformation occurred. However, the plastic strains reached a maximum early in this stage of loading and no additional plastic strains occurred. This behavior is not typical of the real web material, and may be attributed to the artificially low yield strength assigned to the web in this model. Even in the presence of this plastic deformation, no wrinkling or folding was observed in the web downstream of roller 3.

The results for WEBRWD1 are shown in Fig. 12. Under the influence of the roller misalignment, ridging in the tail of the web and isolated pockets of plastic strain arise during the pretensioning stage of loading. As the web accelerated, additional regions of plastic strain formed between rollers two and three. As roller three oscillates, contours of plastic strain show that this deformation alternates from the top edge of the web to the bottom due to extra tension as the web oscillates. New regions of plastic straining develop through the motion, with localized bands of plastic deformation arising and causing the web to develop wrinkles, as is clearly visible at roller two (the middle roller) in Fig. 12.

Figure 13 shows the deformation pattern for model 5, WEBRWD2 after 1.0 s of web motion. The deformation patterns show behavior similar to that of previous models. Although some plastic deformation developed during loading, the web as a whole was able to support the loads throughout its transport without developing localized regions of plastic deformation. The web in this model was substantially stronger than the web in models WEBRND and WEBRWD1, and better represents the response of the real web than the artificially weak web used in those models. Despite the development of substantial troughs in the upstream tail, no wrinkling or folding defects were observed.

Table 4 summarizes some of the results of the five cases for web transport over the three-roller system. For future investigations, it will be important to investigate the effect of anisotropy on foldover. Carded nonwoven webs are strongly anisotropic, and can have transverse elastic moduli as much as a factor of 5 lower than the longitudinal modulus. Foldovers typically occur along lines aligned parallel to the longitudinal direction, so it would be expected that transverse bending properties would strongly influence foldover. The development of anisotropic elastic-plastic ma-

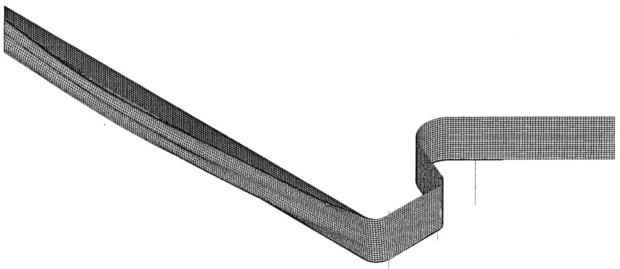


Fig. 13 Deformation of a spatially heterogeneous web during transport through a three-roller system in which the idler roller (roller 3) is misaligned after 1.0 s of web motion, for a web having a mean density of 112.27 kg/m^3 and standard deviation in element density of 8.89 kg/m^3 (model WEBRWD2). The shaded web illustrates the undeformed web path.

Table 4 Some results of the finite element analyses for the three-roller system

Model	Strain localization	Ridge in trailing end of web	Oscillations	Wrinkles
WEBHND	NO	NO	NO	NO
WEBHWD	NO	YES	YES	NO
WEBRND	NO	NO	NO	NO
WEBRWD1	YES	YES	YES	YES
WEBRWD2	NO	YES	YES	NO

terial models representing the unique characteristics of nonwoven webs for implementation in ABAQUS will be necessary to enable further investigation of this effect.

5 Summary and Conclusions

The primary objective of this work was to develop a method to analyze the effects of random material property variations on development of manufacturing defects such as folds and wrinkles during web transport. The spatial heterogeneity of a carded nonwoven web has been simulated by transforming grayscale distributions, obtained by optically scanning real webs, into spectral density matrices in the frequency domain using FFT, ultimately providing the covariance matrix. The covariance matrix was fit to an analytical function, and a computer code was created to generate "virtual webs" possessing a density distribution having the same mean, variance and spatial characteristics as the real web. Finite-element models were then developed using idealized models relating material property variations to the density distribution in order to investigate the effect of spatial heterogeneity on web transport.

In models of conventional tensile testing, isolated regions of plastic strain arose in the weaker elements of the models during initial loading. As loading continued, competing regions of elevated plastic strain developed until localization formed bands of extreme strain spanning the width of the web. For web transport through a three roller system, wrinkles were associated with plastic straining and plastic strain localization. It is therefore believed that spatial heterogeneity in material properties can strongly influence the tendency of a web to develop manufacturing defects. It should be noted that a unique web with its own properties and transport characteristics is generated each time a virtual web is created. Future investigations should investigate the degree of variation in web response across a set of virtual web specimens.

Acknowledgments

The authors would like to acknowledge the support of Los Alamos National Laboratory, operated by the University of California for the National Nuclear Security Administration, of the U.S. Department of Energy under Contract No. W-7406-ENG-36.

References

[1] Shelton, J. J., 1997, "Effects of Web Camber on Handling," in Proceedings of the Fourth International Conference on Web Handling, Oklahoma State University, Stillwater, Oklahoma, June 1–4, pp. 248–263.
[2] Good, J. K., 1997, "Shear in Multispan Web Systems," in Proceedings of the Fourth International Conference on Web Handling, Oklahoma State University, Stillwater, Oklahoma, June 1–4, pp. 264–286.
[3] Onno, F., and Elias, A., 1997, "Optimization of Roll Mechanical Profiles in a Continuous Annealing Line For Steel Strips," in Proceedings of the Fourth International Conference on Web Handling, Oklahoma State University, Stillwater, Oklahoma, June 1–4, pp. 430–445.
[4] Ko, F. K., and Du, G. W., 1998, "Textile Preforming," *Handbook of Composites*, 2nd ed., edited by S. T. Peters, Chapman and Hall, New York, pp. 397–424.
[5] Stack, K., Perconti, J., Jeans, A., LaFleche, J., and Benson, R., 1997, "A Nonlinear Finite Element Model for Web Spreading," in Proceedings of the Fourth International Conference on Web Handling, Oklahoma State University, Stillwater, Oklahoma, June 1–4, pp. 445–461.
[6] Good, J. K., Kedl, D. M., and Shelton, J. J., 1997, "Shear Wrinkling in Isolated Spans," in Proceedings of the Fourth International Conference on Web Handling, Oklahoma State University, Stillwater, Oklahoma, June 1–4, pp. 462–480.
[7] Swanson, R. P., 1997, "Testing and Analysis of Web Spreading and Anti-Wrinkle Devices," in Proceedings of the Fourth International Conference on Web Handling, Oklahoma State University, Stillwater, Oklahoma, June 1–4, pp. 414–429.
[8] Lin, C. C., and Mote, C. D., Jr., 1995, "Equilibrium Displacement and Stress Distribution in a Two-Dimensional, Axially Moving Web Under Transverse Loading," ASME J. Appl. Mech., **62**, pp. 772–779.
[9] Lin, C. C., and Mote, C. D., Jr., 1996, "The Wrinkling of Rectangular Webs Under Nonlinearly Distributed Edge Loading," ASME J. Appl. Mech., **63**, pp. 665–669.
[10] Lin, C. C., and Mote, C. D., Jr., 1996, "The Wrinkling of Thin, Flat, Rectangular Webs," ASME J. Appl. Mech., **63**, pp. 774–779.
[11] Becker, R., 1987, "The Effect of Porosity Distribution on Ductile Failure," J. Mech. Phys. Solids, **35**, No. 5, pp. 577–599.
[12] Khvastunkov, M., and Leggoe, J. W., 2004, "Adapting cellular automata to model failure in spatially heterogeneous ductile alloys," Scr. Mater., **51**, pp. 309–314.
[13] Nakamura, T., and Suresh, S., 1993, "Effects of Thermal Residual Stresses and Fiber Packing on Deformation of Metal-Matrix Composites," Acta Metall. Mater., **41**, pp. 1665–1681.
[14] Brockenbrough, J. R., Suresh, S., and Wienecke, H. A., 1991, "Deformation of Metal-Matrix Composites with Continuous Fibers: Geometrical Effects of Fiber Distribution and Shape," Acta Metall. Mater., **39**, pp. 735–752.
[15] Leggoe, J. W., Mammoli, A. A., Bush, M. B., and Hu, X. Z., 1998, "Finite Element Modelling of Deformation in Particulate Reinforced Metal-Matrix Composites with Random Local Microstructure Variation," Acta Mater., **46**, pp. 6075–6088.
[16] Chen, X., and Mai, Y., 1998, "Micromechanics of Rubber-Toughened Polymers," J. Mater. Sci., **33**, pp. 3529–3539.
[17] Thigpen, L., Reardon, P., Fitzgerald, M., Leggoe, J., and Graham, A., 2000, "Numerical Simulations of the Effect of Local Material Property Variation on Deformation of Webs During Loading," LACP-0034, Los Alamos National Laboratory, Los Alamos, New Mexico.
[18] Karlin, S., and Taylor, H. M., 1975, *A First Course in Stochastic Processes*, 2nd ed., Academic, New York.
[19] Cressie, N. A. C., 1991, *Statistics for Spatial Data*, Wiley, New York.
[20] Shinozuka, M., 1971, "Simulation of Multivariate and Multidimensional Random Processes," J. Acoust. Soc. Am., **49**, pp. 357–368.
[21] Shinozuka, M., and Jan, C.-M., 1972, "Digital Simulation of Random Processes and its Applications," J. Sound Vib., **25**, pp. 111–128.
[22] ABAQUS, 1998, Users Manual, Version 5.8, Hibbitt, Karlsson and Sorensen, Inc., Pawtucket, RI.
[23] Sun, H. Y., and Qiao, X., 1997, "Prediction of the Mechanical Properties of Three-Dimensionally Braided Composites," Compos. Sci. Technol., **57**, pp. 623–629.
[24] Gao, F., Boniface, L., Ogin, S. L., Smith, P. A., and Greaves, R. P., 1999, "Damage Accumulation in Woven-Fabric CFRP Laminates Under Tensile Loading: 2. Modelling the Effect of Damage on Macro-Mechanical Properties," Compos. Sci. Technol., **59**, pp. 137–145.
[25] Scida, D., Aboura, Z., Benzeggagh, M. L., and Bocherens, E., 1999, "A Micromechanics Model for 3D Elasticity and Failure of Woven-Fibre Composite Materials," Compos. Sci. Technol., **59**, pp. 505–517.
[26] Ladd, A. J. C., and Kinney, J. H., 1998, "Numerical Errors and Uncertainties in Finite-Element Modeling of Trabecular Bone," J. Biomech., **31**, pp. 941–945.
[27] Choi, J. B., and Lakes, R. S., 1995, "Analysis of Elastic Modulus of Conventional Foams and Re-Entrant Foam Materials with a Negative Poisson's Ratio," Int. J. Mech. Sci., **27**, pp. 51–59.
[28] Wang, Y., 1999, "Effect of Consolidation Method on the Mechanical Properties of Non-woven Fabric Reinforced Composites," Appl. Compos. Mater., **6**, pp. 19–34.
[29] Christensen, R. M., 1995, "The Hierarchy of Microstructures for Low Density Materials," ZAMP, **46**, Special Issue, pp. 5506–5521.
[30] Besson, J., Devillers-Guerville, L., and Pineau, A., 2000, "Modeling of scatter and size effect in ductile fracture: application to thermal embrittlement of duplex stainless steels," Eng. Fract. Mech., **67**, pp. 169–190.

Eric M. Mockensturm[1]
e-mail: emm10@psu.edu
Mem. ASME

Jianping Guo
Graduate Student

Department of Mechanical and Nuclear Engineering,
Pennsylvania State University,
157 Hammond Building,
University Park, PA 16802

Nonlinear Vibration of Parametrically Excited, Viscoelastic, Axially Moving Strings

The dynamic response of parametrically excited, axially moving viscoelastic belts is investigated in this paper. Results are compared to previous work in which the partial, not material, time derivative was used in the viscoelastic constitutive relation. It is found that this added "steady state" dissipation greatly affects both the existence and amplitudes of nontrivial limit cycles. The discrepancy increases with increasing translation speed. To limit the comparison to the additional physics included in the model, the solution procedure of Zhang and Zu [1,2], who applied the method of multiple scales to the governing equations prior to discretization, is retained. The excitation here is provided by physically stretching the belt. In this case, viscoelastic behavior and excitation frequency also affects the amplitude of the tension fluctuations. [DOI: 10.1115/1.1827248]

Introduction

Axially moving materials are present in many engineering systems from material processing machines to power transmission belts. For this reason, a great many investigators have studied the dynamic behavior of these structures; see [1–10] and the references therein. One prominent use of axially moving materials in power transmission is the serpentine belt drive found in nearly every automobile produced in the last 15 years. Because the crankshaft pulley, which drives the serpentine belt, does not rotate at a steady angular velocity over one revolution and the torques on the accessory pulleys are periodic, the tension in each belt span varies periodically. Front end accessory drive designers then must be aware that tension fluctuations may cause a parametric resonance in the belt span. The belt drive can then be designed such that engine speeds (excitation frequencies) which cause resonances can be moved away from frequent operating ranges.

To better understand the effect of parametric excitation in individual belt spans, Mockensturm et al. [10] used a string model to investigate the regions where principal, and first summation and difference parametric resonances may occur. Discretization was performed using the eigenfunctions of an axially moving string with constant tension. Modal damping was introduced and first order nonlinear terms were included to estimate the vibration amplitude in the resonance regions.

In a series of subsequent papers by Zu and co-authors [1,2,11–16], a better understanding of the mechanical energy dissipation mechanisms was attempted by using a viscoelastic material model for the belt. A variety of constitutive models and excitation sources were investigated. Weakly nonlinear terms were retained as the viscoelastic effect only appears in the nonlinear terms of the string model used. While these studies provided a systematic method to include material damping in the analysis, an important material dissipation mechanism was not included in the derivation of the equations of motion.

In what follows, this mechanism is included and a discussion of its affects on the equations of motion is provided. A more physically motivated excitation source is also introduced. The effects on the previous results for a parametrically excited belt span are then studied. As the translation speed increases the discrepancy between the previous and present results grows.

Equation of Motion

Here the weakly nonlinear equations of motion for an axially moving string composed of a viscoelastic material and under parametric excitation are derived. It is important to make clear the source of the parametric excitation so it can be modeled correctly. The usual source of parametric excitation in a tensioned string is a moving support which causes the tension in the string to change but has negligible effect on the total length between the supports. One could also assume that an external field is acting on the string to alter the tension directly while holding the supports fixed. Such excitation does not occur in most mechanical systems and is not how tension fluctuations arise in a serpentine belt. The motion of the crankshaft is assumed to be prescribed, with the accessory drive system having no effect on its motion. Thus, in the following, it is assumed tension fluctuations are caused by belt stretching.

The primary difference between the following and previous derivations is the generalization to a viscoelastic material model. In previous work, the viscous material model did not include a ''steady dissipation'' term due to the axial motion of the string; the elastic modulus was simply replaced by an operator such that $E \rightarrow \hat{E}[\bullet] = E(\bullet) + \eta(\bullet)_t$ where E is the elastic modulus, η is the viscous material constant, and a subscript denotes partial differentiation. This transformation (model) neglects the dissipation present when the belt is undergoing steady motions. A more physical transformation (model) would be $E \rightarrow \hat{E}[\bullet] = E(\bullet) + \eta(\dot{\bullet})$ where a dot denotes material, not partial, time differentiation. For an axially moving string, the material and partial time derivatives are not identical and related by

$$(\dot{\bullet}) = (\bullet)_t + v(\bullet)_x, \tag{1}$$

where v is the axial velocity of a steadily translating reference configuration.

In addition to the effects of translation speed on material viscous dissipation, one must also be careful when determining the effects of material damping on the parametric excitation. If the string tension was being varied directly (by some means), then

[1]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, June 6, 2003; final revision, November 3, 2004. Associate Editor: O. O'Reilly. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

viscous effects would not alter the parametric excitation. However, if the tension fluctuations are a result of stretching, the change in constitutive model will alter the parametric excitation. For an elastic string, the tension is approximated as $EA\varepsilon$ where A is the referential cross-sectional area. If the belt is being periodically stretched, ε and, thus, tension are periodic functions of time. If the belt is composed of a simple viscoelastic material, the tension is related to the strain by

$$T = A(E\varepsilon + \eta\dot{\varepsilon}) = A[E\varepsilon + \eta(\varepsilon_t + v\varepsilon_x)]. \tag{2}$$

If the stretching is uniform, the partial with respect to x is zero and the tension does not depend on the translation speed. The tension, however, does still depend on the strain rate. If the strain in the belt is varying harmonically as $\varepsilon = \Delta \sin(\Omega t)$, the dynamic tension in the belt is given by

$$T/A = E\Delta \sin(\Omega t) + \eta\Delta\Omega \cos\Omega t = \Lambda \sin(\Omega t + \phi), \tag{3}$$

where $\Lambda^2 = (E\Delta)^2 + (\eta\Delta\Omega)^2$ and the phase, ϕ, is not important if only steady state behavior is studied. The excitation amplitude, as well as the frequency, then affects the amplitude of parametric excitation.

Labeling the axial and transverse displacements U and W, respectively, the Lagrangian strain in the string is $\varepsilon(x,t) = U_x(t) + W_x^2(x,t)/2$. For the viscoelastic material the tension is

$$T/A = \hat{E}[\varepsilon] = E[U_x + \tfrac{1}{2}W_x^2] + \eta[U_{xt} + W_xW_{xt} + vW_xW_{xx}]. \tag{4}$$

The equation governing the transverse motion of the string ($T_x = \rho A\ddot{W}$) becomes

$$\rho W_{tt} + 2\rho v W_{xt} + (\rho v^2 - EU_x - \eta U_{xt})W_{xx} = \{W_x[\tfrac{1}{2}EW_x^2 + \eta(W_xW_{xt} + vW_xW_{xx})]\}_x, \tag{5}$$

where ρ is the referential mass per unit volume.

The spatially constant axial strain U_x is assumed to be additively composed of a temporally constant strain, ε_0, and a harmonically varying strain, $\varepsilon_1 \cos\Omega t$, where Ω is the frequency of excitation.

With the nondimensional parameters

$$w = \frac{W}{L},\quad \xi = \frac{X}{L},\quad \tau = t\sqrt{\frac{E\varepsilon_0}{\rho L^2}},\quad c = v\sqrt{\frac{\rho}{E\varepsilon_0}},$$
$$a = \frac{\varepsilon_1}{\varepsilon_0},\quad \omega = \Omega\sqrt{\frac{\rho L^2}{E\varepsilon_0}},\quad \zeta = \eta\sqrt{\frac{1}{\rho E\varepsilon_0 L^2}},\quad \varphi = \frac{1}{\varepsilon_0} = \frac{EA}{T_0}, \tag{6}$$

the nondimensional equation of the transverse motion is

$$w_{\tau\tau} + 2cw_{\tau\xi} + (c^2 - 1 - a\cos\omega\tau + \underline{\zeta\omega a \sin\omega\tau})w_{\xi\xi} = \hat{N}[w], \tag{7}$$

where the nonlinear operator $N[w]$ is defined as

$$\hat{N}[w] = \tfrac{3}{2}\varphi w_\xi^2 w_{\xi\xi} + \zeta w_\xi w_{\xi\xi} w_{\xi\tau} + \zeta w_\xi (w_{\xi\xi}w_{\xi\tau} + w_\xi w_{\xi\xi\tau}) + \underline{c\zeta w_\xi (2w_{\xi\xi}^2 + w_\xi w_{\xi\xi\xi})}. \tag{8}$$

If underlined terms are ignored Zhang and Zu's model [1,2] is recovered. Note, in this formulation the viscoelastic constant appears in both the parametric excitation in Eq. (7) and the nonlinear terms in Eq. (8).

Following Wickert and Mote [17], Eq. (7) can be written in canonical operator form by introduction of the operators

$$\hat{M}[\bullet] = (\bullet) \quad \hat{G}[\bullet] = 2c(\bullet)_\xi \quad \hat{K}[\bullet] = (c^2 - 1)(\bullet)_{\xi\xi}, \tag{9}$$

as

$$\hat{M}[w_{\tau\tau}] + \hat{G}[w_\tau] + \hat{K}[w] = \hat{N}[w] + a(\cos\omega\tau - \zeta\omega\sin\omega\tau)w_{\xi\xi}. \tag{10}$$

Approximate Solution

Because the goal of this work is to investigate the importance of the mechanics neglected in previous work and not to compare solution methods, the solution procedure used by Zhang and Zu [1,2] is employed. The perturbation method is applied directly to the partial differential Eq. (10) and the sequence of perturbation equations discretized using the traveling string eigenfunctions.

Thus, an approximation is sought in the form

$$w(\xi,\tau,\varepsilon) = w_0(\xi,T_0,T_1) + \varepsilon w_1(\xi,T_0,T_1) + \ldots, \tag{11}$$

where $T_0 = \tau$ is a fast time scale that, in steady state, characterizes motions occurring at ω. $T_1 = \varepsilon\tau$ is a slow time scale on which the amplitudes and phases depend.

If the amplitude of parametric excitation, a, is of order ε, the equations for the leading order and first correction terms in Eq. (11) are

$$\hat{M}[\overset{\circ\circ}{w_0}] + \hat{G}[\overset{\circ}{w_0}] + \hat{K}[w_0] = 0, \tag{12}$$

$$\hat{M}[\overset{\circ\circ}{w_1}] + \hat{G}[\overset{\circ}{w_1}] + \hat{K}[w_1] = -2\hat{M}[\overset{\circ}{w_0'}] - \hat{G}[w_0'] + \hat{N}[w_0] + \tilde{a}\sin(\omega T_0 + \phi)(w_0)_{\xi\xi}, \tag{13}$$

where $(\overset{\circ}{\bullet})$ and $(\bullet)'$ denote a partial differentiation with respect to T_0 and T_1, respectively. The dimensionless amplitude of parametric excitation is then $\tilde{a}^2 = a^2(1 + \zeta^2\omega^2)$. Using the standard procedure to study resonances that may occur when the excitation frequency is near a summative combination of any two system natural frequencies, a detuning parameter, μ, is introduced such that

$$\omega = \omega_n + \omega_l + \varepsilon\mu, \tag{14}$$

in which ω_n and ω_l are natural frequencies of the linear system.

Solutions of Eq. (12) can be expressed as

$$w_0 = \psi_n(\xi)A_n(T_1)e^{i\omega_n T_0} + \psi_l(\xi)A_l(T_1)e^{i\omega_l T_0} + cc, \tag{15}$$

where only the nth and lth complex eigenfunctions, $\psi_n(\xi)$ and $\psi_l(\xi)$, are retained to study combination parametric resonances. To reduce equation length the notation cc is used to denote the complex conjugate of all preceding terms in the expression. Functions A_n and A_l are found by eliminating the secular terms from the equation governing w_1.

Substituting Eqs. (14) and (15) into Eq. (13) and expressing the trigonometric functions in exponential form results in

$$\begin{aligned}\hat{M}\left[\overset{\circ\circ}{w_1}\right] + \hat{G}\left[\overset{\circ}{w_1}\right] + \hat{K}[w_1] = \text{NST} + \Bigg\{ &-2i\omega_n A_n'\hat{M}[\psi_n] - A_n'\hat{G}[\psi_n] \\ &+ \frac{\tilde{a}\overline{A_l}}{2}\frac{\partial^2\overline{\psi_l}}{\partial\xi^2}e^{i\mu T_1} + [M_{2n}(3\varphi \\ &+ 2i\omega_n\zeta) + cM_{3n}\zeta]A_n^2\overline{A_n}\Bigg\}e^{i\omega_n T_0} \\ + \Bigg\{ &-2i\omega_l A_l'\hat{M}[\psi_l] - A_l'\hat{G}[\psi_l] \\ &+ \frac{\tilde{a}\overline{A_n}}{2}\frac{\partial^2\overline{\psi_n}}{\partial\xi^2}e^{i\mu T_1} + [M_{2l}(3\varphi \\ &+ 2i\omega_l\zeta) + cM_{3l}\zeta]A_l^2\overline{A_l}\Bigg\}e^{i\omega_l T_0} + cc,\end{aligned} \tag{16}$$

where

$$M_{2k}=\frac{1}{2}\left[\left(\frac{\partial\psi_k}{\partial\xi}\right)^2\frac{\partial^2\overline{\psi}_k}{\partial\xi^2}+2\frac{\partial\psi_k}{\partial\xi}\frac{\partial\overline{\psi}_k}{\partial\xi}\frac{\partial^2\psi_k}{\partial\xi^2}\right], \tag{17}$$

$$M_{3k}=4\frac{\partial\psi_k}{\partial\xi}\frac{\partial^2\psi_k}{\partial\xi^2}\frac{\partial^2\overline{\psi}_k}{\partial\xi^2}+2\frac{\partial\psi_k}{\partial\xi}\frac{\partial\overline{\psi}_k}{\partial\xi}\frac{\partial^3\psi_k}{\partial\xi^3}+2\frac{\partial\overline{\psi}_k}{\partial\xi}\left(\frac{\partial^2\psi_k}{\partial\xi^2}\right)^2+\left(\frac{\partial\psi_k}{\partial\xi}\right)^2\frac{\partial^3\overline{\psi}_k}{\partial\xi^3}, \tag{18}$$

for $k=n,\ l$ and the phase of the excitation has been neglected. The nonsecular terms in Eq. (16) are unimportant in subsequent analysis and combined into the term NST. The term M_{3k} does not appear in previous work.

Equation (16) has a solution only if a solvability condition is satisfied. This requires that the right side of Eq. (16) be orthogonal to every solution of the homogeneous equation. For the general case where internal resonances do not exist, the solvability condition can be expressed as

$$-2i\omega_nA_n'm_n-A_n'g_ni+[(3\varphi+2i\omega_n\zeta)m_{2n}-ic\zeta m_{3n}]A_n^2\overline{A}_n+\frac{\tilde{a}\overline{A}_l}{2}m_{ln}e^{i\mu T_1}=0, \tag{19}$$

$$-2i\omega_lA_l'm_l-A_l'g_li+[(3\varphi+2i\omega_l\zeta)m_{2l}-ic\zeta m_{3l}]A_l^2\overline{A}_l+\frac{\tilde{a}\overline{A}_n}{2}m_{nl}e^{i\mu T_1}=0, \tag{20}$$

in which

$$m_k=\langle\hat{M}[\psi_k],\psi_k\rangle,\quad g_k=-i\langle\hat{G}[\psi_k],\psi_k\rangle,$$

$$m_{2k}=\langle M_{2k},\psi_k\rangle,\quad m_{3k}=i\langle M_{3k},\psi_k\rangle, \tag{21}$$

$$m_{nl}=\left\langle\frac{\partial^2\overline{\psi}_n}{\partial\xi^2},\psi_l\right\rangle,\quad m_{ln}=\left\langle\frac{\partial^2\overline{\psi}_l}{\partial\xi^2},\psi_n\right\rangle$$

and the notation $\langle\bullet,\bullet\rangle$ represents the standard inner product of two complex functions over $\xi\in(0,1)$.

Referring to Wickert and Mote [17], the kth natural frequency and mass normalized eigenfunction of translating strings are $\omega_k=k\pi(1-c^2)$ and $\psi_k=\sqrt{2}\sin(k\pi\xi)e^{ik\pi c\xi}$, respectively. Substituting these eigenvalues and eigenfunctions into Eqs. (19) and (20) leads to

$$g_k=2k\pi c^2\quad m_{2k}=-\frac{1}{4}\pi^4k^4(3+2c^2+3c^4)$$

$$m_{3k}=\frac{1}{2}\pi^5k^5c(7+6c^2+3c^4) \tag{22}$$

$$\frac{m_{nl}}{4\pi n^2l^2c}=\frac{-(-1)^{n+l}\sin[(n+l)\pi c]+i\{1-(-1)^{n+l}\cos[(n+l)\pi c]\}}{(n+l)[(n+l)^2c^2-(n-l)^2]}. \tag{23}$$

Equation (23) is slightly different than that derived by Zhang and Zu [1,2]. In particular, the sign toggling terms $(-1)^{n+l}$ were incorrectly neglected in their derivation. Thus, when $n+l$ is even—for all the principal parametric resonances—*and* ζ is zero the results here agree with those of Zhang and Zu. However, when $n+l$ is odd, for some summation parametric resonances, the results here differ from Zhang and Zu but agree with Mockensturm et al. [10] when ζ is zero. Thus, in addition to neglecting the dynamic dissipation, a mathematical mistake corrupted the results of Zhang and Zu.

To solve the nonlinear Eqs. (19) and (20), express A_n and A_l in polar form

$$A_n=\frac{1}{2}\alpha_ne^{i\beta_n},\quad A_l=\frac{1}{2}\alpha_le^{i\beta_l}. \tag{24}$$

Note that α_k and β_k represent the amplitude and the phase of the response, respectively. Substituting Eq. (24) into Eqs. (19) and (20) and separating the resulting equation into real and imaginary parts yields

$$\alpha_n'=\frac{2\omega_nm_{2n}-cm_{3n}}{8n\pi}\zeta\alpha_n^3+\frac{\tilde{a}\alpha_l}{4n\pi}[m_{nl}^I\cos(\mu T_1-\beta_n-\beta_l)+m_{nl}^R\sin(\mu T_1-\beta_n-\beta_l)], \tag{25}$$

$$\alpha_l'=\frac{2\omega_lm_{2l}-cm_{3l}}{8l\pi}\zeta\alpha_l^3+\frac{\tilde{a}\alpha_n}{4l\pi}[m_{nl}^I\cos(\mu T_1-\beta_n-\beta_l)+m_{nl}^R\sin(\mu T_1-\beta_n-\beta_l)], \tag{26}$$

$$\alpha_n\beta_n'=-\frac{3m_{2n}\varphi}{8n\pi}\alpha_n^3-\frac{\tilde{a}\alpha_n}{4n\pi}[m_{nl}^R\cos(\mu T_1-\beta_n-\beta_l)-m_{nl}^I\sin(\mu T_1-\beta_n-\beta_l)], \tag{27}$$

$$\alpha_l\beta_l'=-\frac{3m_{2l}\varphi}{8l\pi}\alpha_l^3-\frac{\tilde{a}\alpha_l}{4l\pi}[m_{nl}^R\cos(\mu T_1-\beta_n-\beta_l)-m_{nl}^I\sin(\mu T_1-\beta_n-\beta_l)], \tag{28}$$

where $m_{nl}^I=\mathrm{Im}(m_{nl})$ and $m_{nl}^R=\mathrm{Re}(m_{nl})$.

Limit Cycles

Equations (25)–(28) govern the slowly varying amplitudes and phases of the modal coordinates. If the amplitudes are not changing with time, steady motions (limit cycles) are obtained. The trivial solution ($\alpha_l=\alpha_n=0$) always exists, as expected. Since the equations for the amplitudes and phases depend explicitly on (slow) time, it is clear that if the amplitudes are constant the phases must be functions of T_1; in particular, the phases must be such that $\beta_n'+\beta_l'=\mu$.

In the following, nontrivial steady motions are obtained for both elastic and viscoelastic belts. Multiplying Eq. (27) by α_l and Eq. (28) by α_n and summing the results provides

$$\mu=-3\varphi\left(\frac{m_{2n}\alpha_n^2}{8\pi n}+\frac{m_{2l}\alpha_l^2}{8\pi l}\right)-\left(\frac{\tilde{a}}{4n\pi}+\frac{\tilde{a}}{4l\pi}\right)[m_{nl}^R\cos\theta-m_{nl}^I\sin\theta], \tag{29}$$

where it has been assumed both α_n and α_l are nonzero, and the constant $\theta=\mu T_1-\beta_n-\beta_l$.

Elastic Limit Cycles. For the case when no viscous material effects are considered ($\zeta=0$), nontrivial limit cycles only occur when, from Eqs. (25) or (26), $m_{nl}^I \cos\theta + m_{nl}^R \sin\theta = 0$. Finding $\sin\theta$ and $\cos\theta$ in terms of the system parameters, m_{nl}^R and m_{nl}^I, and inserting the result into Eq. (29) gives

$$n^3\alpha_n^2 + l^3\alpha_l^2 = \frac{8}{\varphi(9+6c^2+9c^4)}\left[\frac{4\mu}{\pi^3} \mp \tilde{a}\left(\frac{1+n}{nl\pi^4}\right)\sqrt{(m_{nl}^I)^2+(m_{nl}^R)^2}\right], \tag{30}$$

where

$$\sqrt{(m_{nl}^R)^2+(m_{nl}^I)^2} = \frac{4\pi n^2 l^2 c\sqrt{2-2(-1)^{n+l}\cos[(n+l)\pi c]}}{(n+l)[(n+l)^2c^2-(n-l)^2]}. \tag{31}$$

For a principle parametric resonance ($l=n$), the limit cycle amplitudes are given by

$$\alpha_n^2 = \frac{8}{n^3\varphi(9+6c^2+9c^4)}\left[\frac{2\mu}{\pi^3} \mp \frac{\tilde{a}}{n\pi^4}\sqrt{(m_{nl}^I)^2+(m_{nl}^R)^2}\right]. \tag{32}$$

Real values for α_n exist only if the right-hand side of Eq. (30) is positive, or

$$\mu \geqslant \pm\frac{\tilde{a}}{4\pi}\left(\frac{1}{n}+\frac{1}{l}\right)\sqrt{(m_{nl}^I)^2+(m_{nl}^R)^2}. \tag{33}$$

Zhang and Zu [1,2] give a relationship between α_n and α_l for combination resonances. However, it is not clear how this was derived and numerically integrating Eqs. (25)–(28) shows the relationship is not valid.

Steady-State Response of Viscoelastic Moving Belts. When the viscoelastic effects of the belt are not negligible, one finds the steady state amplitudes of the excited modes are related by

$$\alpha_l^2 = \frac{n^2}{l^2}\sqrt{\frac{n}{l}}\,\alpha_n^2. \tag{34}$$

Using this and Eq. (25) to eliminate θ from Eq. (29), gives

$$C_1\alpha_n^6 + C_2\alpha_4^4 + C_3\alpha_n^2 = 0, \tag{35}$$

where C_1, C_2, and C_3 are time independent constants given by

$$C_1 = AB + P^2, \quad C_2 = 2P\mu,$$
$$\text{and} \quad C_3 = \mu^2 - B[(m_{nl}^I)^2+(m_{nl}^R)^2], \tag{36}$$

with

$$A = \frac{(2\omega_n m_{2n} - cm_{3n})\zeta l^{5/4}}{2\tilde{a}n^{5/4}},$$
$$B = \left\{\frac{\tilde{a}}{4\pi}\left[\frac{1}{l}\left(\frac{n}{l}\right)^{1/4} + \frac{1}{n}\left(\frac{l}{n}\right)^{1/4}\right]\right\}^2,$$
$$P = \frac{3\varphi}{8}\left(\frac{m_{2n}}{n\pi} + \frac{m_{2l}n^{5/2}}{l^{7/2}\pi}\right). \tag{37}$$

These expressions are identical to those of Zhang and Zu [1,2] if m_{3n} is set to zero.

Equation (35) obviously admits the trivial solution $\alpha_n = 0$. Nontrivial solutions simply have amplitudes governed by

$$\alpha_{n0}^2 = \frac{-C_2 \pm \sqrt{C_2^2 - 4C_1C_3}}{2C_1}. \tag{38}$$

The amplitudes of a general summation parametric resonance are then given by Eqs. (34) and (38). For the special case of a principal parametric resonance, the amplitude simplifies to

$$\frac{\alpha_{n0}^2}{2n\pi} = \frac{-6\beta c\mu m_{2n} \pm \sqrt{(3\tilde{a}m_{2n}\varphi\sin n\pi c)^2 - \zeta^2(2\omega_n m_{2n} - cm_{3n})^2(4\mu^2c^2 - \tilde{a}^2\sin^2 n\pi c)}}{c[\zeta^2(2\omega_n m_{2n} - cm_{3n})^2 + 9\varphi^2 m_{2n}^2]}. \tag{39}$$

Nontrivial limit cycles can then only exist if α_n is real or

$$C_2^2 - 4C_1C_3 > 0 \quad \text{and} \quad -C_2 \pm \sqrt{C_2^2 - 4C_1C_3} > 0. \tag{40}$$

Substituting the expressions for C_1, C_2, and C_3 into Eq. (40) leads to the conclusions that the first limit cycle of the viscoelastic system exists if

$$\pm\left(\sqrt[4]{\frac{n}{l^5}} + \sqrt[4]{\frac{l}{n^5}}\right)\frac{\sqrt{(m_{nl}^I)^2+(m_{nl}^R)^2}}{4\pi} \leqslant \frac{\mu}{\tilde{a}} \leqslant \frac{-2\sqrt{[(m_{nl}^I)^2+(m_{nl}^R)^2]C_1}}{(2\omega_n m_{2n} - cm_{3n})\zeta}\left(\frac{n}{l}\right)^{5/4}. \tag{41}$$

Results and Discussions

Although the energy dissipated during steady state motion of the belt may not seem significant, only slightly modifying the governing equations, the effect on the parametric resonances can be substantial. As is expected, after correcting an algebra mistake, the present analysis agrees with that of Zhang and Zu [1,2] when the belt is not translating or steady state dissipation is absent. However, as the speed of the belt increases, the consequences of not including steady state dissipation become significant. In the following these differences will be highlighted. Stability results are not presented in the present work since they are as would be expected. The same dimensionless values of the elastic and viscoelastic moduli used by Zhang and Zu ($\varphi=400$, $\zeta=10$) are also used here.

Results for primary parametric resonances can be compared directly with those of Zhang and Zu. However, due to the algebraic mistake mentioned previously, the combination parametric resonance ($n=1$ and $l=2$) for a viscoelastic belt presented here does not agree with the Zhang and Zu result even if steady state dissipation is neglected. When results for this parametric resonance are presented, the mathematical error was corrected so that only the effects of steady state dissipation are compared.

As the excitation frequency approaches a summation of any two system natural frequencies the trivial solution becomes unstable and stable nontrivial solutions branch from this point. The amplitude-phase coordinates undergo a pitchfork bifurcation. As the frequency increases, the amplitude of the response grows as expected for a system that stiffens with increasing deformation. At some point above the summation of the natural frequencies, the trivial solution again becomes stable and an unstable nontrivial branch appears. Increasing the excitation frequency still further, for the dissipative system the stable and unstable nontrivial branches coalesce at a saddle-node bifurcation in amplitude-phase space. For conservative systems the nontrivial amplitudes continue to grow with increasing excitation frequency, without annihilating each other in a saddle-node bifurcation.

The stability diagrams are presented as projections of the sur-

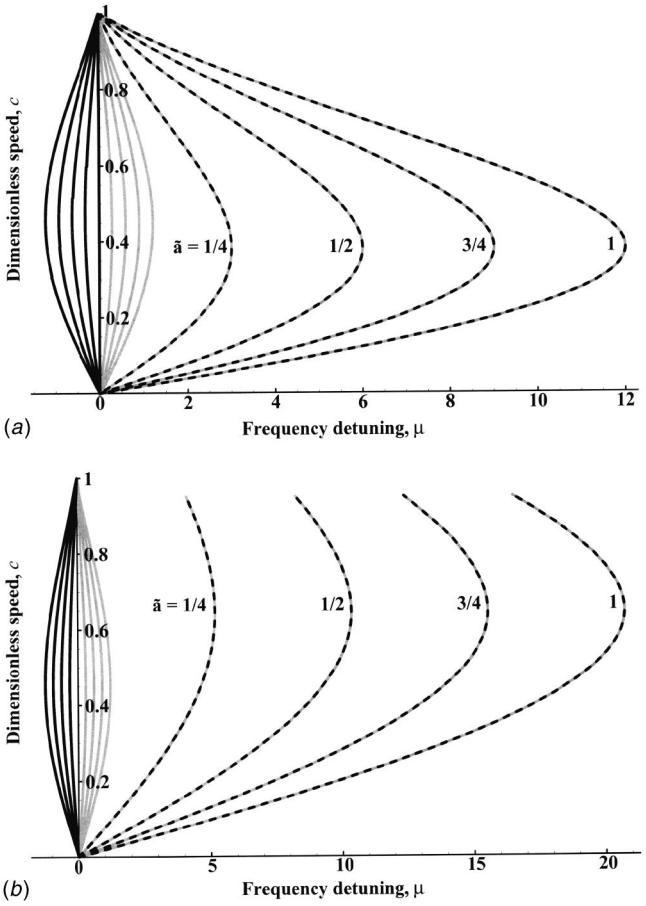


Fig. 1 Stability boundaries of the first combination parametric resonance ($n=1$, $l=2$) obtained with (*a*) and without (*b*) steady state dissipation with $\zeta=10$ and $\varphi=400$

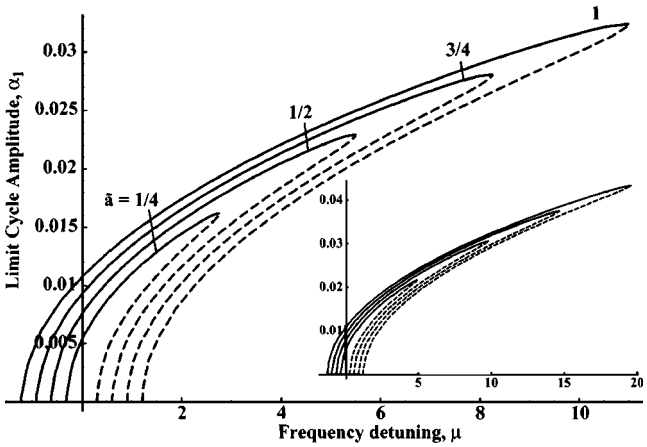


Fig. 2 The amplitude of the nontrivial limit cycles in the first combination parametric resonance ($n=1$ and $l=2$) for $c=0.5$, $\zeta=10$, and $\varphi=400$. Inset shows solution neglecting steady state dissipation.

faces in the $\tilde{a}$-μ-c parameter space onto the μ-c plane. The area between the dark solid and dashed lines corresponds to the region in which the stable nontrivial solution exists. These regions expand for both excitation frequencies above and below the nominal value with increasing excitation amplitude. The regions between the solid dark and light lines is where the trivial solution is unstable. Again, this region grows with increasing excitation amplitude. Finally, the regions between the solid light and dashed lines enclose areas in which the unstable nontrivial limit cycles exist.

For the first combination parametric resonance ($n=1$ and $l=2$), significant difference is predicted by the present analyses. Figures 1(*a*) (the present results) and 1(*b*) (Zhang and Zu's results) illustrate how the extra dissipation alters the detuning frequency at which the two nontrivial branches coalesce and only the trivial solution remains. The curves where these branches emanate from the trivial solution are identical in both analyses; there is no steady state dissipation when the belt is traveling without deformation. In both cases, this combination resonance region is closed when the translation speed is zero, as expected. For translation speeds significantly lower than the critical speed, the differences are slight. This is expected since the steady state dissipation is small when c is small. However, at just 20% of critical speed and for all values of $\tilde{a}$, the detuning at which the unstable and stable branches coalesce is 8.9% less with steady state dissipation. This difference grows to 44% at half critical speed, and 100% at critical speed, where the present analysis predicts this combination resonance disappears as it does when no dissipation is present [2]. This is not predicted by the Zhang and Zu analysis. The largest region of nontrivial limit cycles occurs at $c=0.38$ presently and at $c=0.65$ in prior work. The limit cycle amplitudes for present and previous (inset) analyses at half critical speed are shown in Fig. 2. For all excitation amplitudes and $c=0.5$ the maximum amplitude is predicted to be 25% less in the present analysis. The effect of the added dissipation is as expected; the regions were nontrivial solutions exist are smaller when steady state dissipation is included.

Qualitatively, steady state dissipation has much the same effect on the first two principal parametric resonance regions as on the first combination resonance; see Figs. 3–6. Again, the solutions approach each other as the translation speed goes to zero and both models predict the resonance regions are largest for a stationary belt. Both models again predict the collapse of the second primary resonance when $c=0.5$. For the first principal resonance, the detuning frequency at which the saddle-node bifurcation occurs is predicted by the present model to again be 8.9% and 44% lower for all $\tilde{a}$, and $c=0.2$ and $c=0.5$, respectively. These values are 8.8% and 31% for the second principal resonance at speeds $c=0.2$ and $c=0.4$, respectively. Again, the present model predicts these resonance regions close when the belt is traveling at its critical speed. When steady state dissipation is neglected the regions where the trivial solution is unstable closes but the saddle-node bifurcations do not approach $\mu=0$ as shown in Figs. 3 and 5. When $\omega\approx 2\omega_1$, at half the critical speed and for all $\tilde{a}$ the peak limit cycle amplitude is again 25% lower in the present analysis; see Fig. 4. When $\omega\approx 2\omega_2$, the peak amplitude is 7.1% and 50% lower in the present analysis for all $\tilde{a}$ and $c=0.25$ (Fig. 6) and $c=0.75$, respectively.

The source of the parametric excitation is extremely important when viscous effects are included in the material model. Zhang and Zu assumed the tension in the belt varies with a given amplitude and frequency. The physical source of this tension variation was not discussed. However, in most systems and certainly belt drives, the source of the tension fluctuation is forced periodic motion of material points at the boundary. In belt drives, this motion is transferred from the pulleys and the length of the span does not change. If this is the case, as noted previously, the excitation frequency and viscoelastic material constant help determine the excitation amplitude. Assuming the system is being excited in a summation parametric resonance, the ratio of the excitation amplitude, $\tilde{a}$, to that given by Zhang and Zu is $\tilde{a}/a=\sqrt{1+(n+l)^2\zeta^2\pi^2(1-c^2)^2}$. As expected, this ratio goes to one as the translation speed approaches the critical speed where all natural frequencies go to zero. The ratio is greatest when the belt is not translating. For a stationary belt, excited in the first principal parametric resonance and using $\zeta=10$ (Zhang and Zu), the amplitude ratio is 62.8. Thus, if the dynamic strain amplitude is a

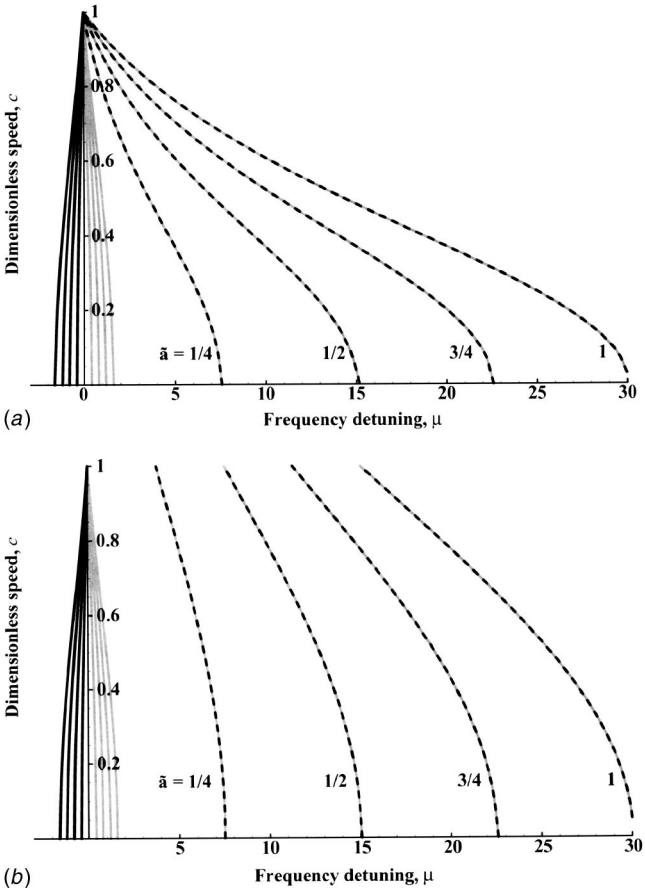


Fig. 3 Stability boundaries of the first principle parametric resonance ($n=l=1$) obtained with (*a*) and without (*b*) steady state dissipation with $\zeta=10$ and $\varphi=400$

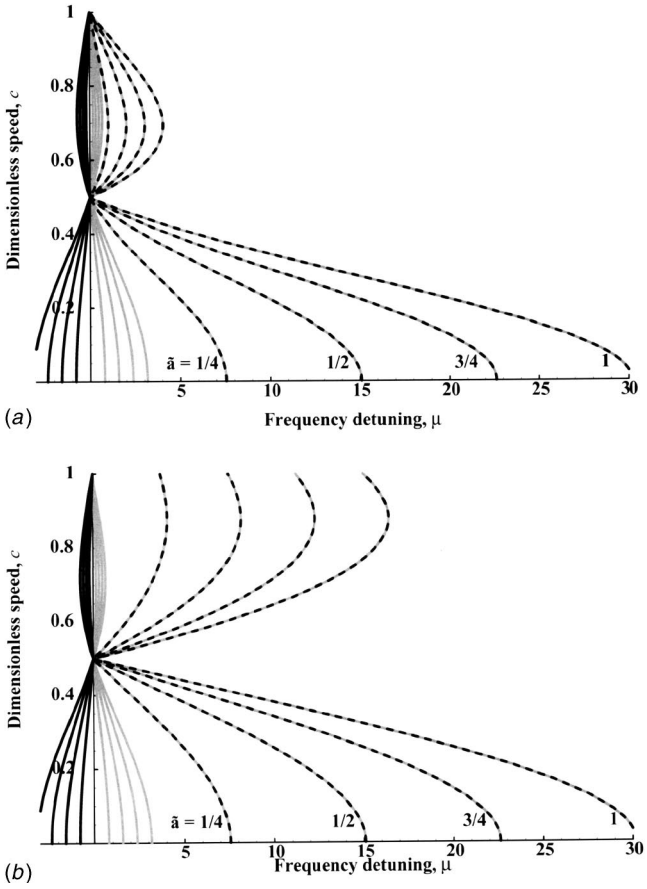


Fig. 5 Stability boundaries of the second principle parametric resonance ($n=l=2$) obtained with (*a*) and without (*b*) steady state dissipation with $\zeta=10$ and $\varphi=400$

fourth of the static strain, the elastic dynamic tension amplitude is a fourth the elastic static tension but the viscoelastic dynamic tension amplitude is 15.7 times the static tension; the belt is then in compression much of the time. For the belt to remain in tension at this driving frequency, the ratio of the dynamic to static strains must be less than 1.6%; at the first combination and second principal resonances, this ratio decreases to 1.1% and 0.80%, respectively. If the material viscoelastic constant is reduced to 10% of that used by Zu ($\zeta=1$), these ratios are 16%, 11%, and 7.9% for the first three resonance regions, respectively. The viscosity substantially alters the amplitude of the parametric excitation.

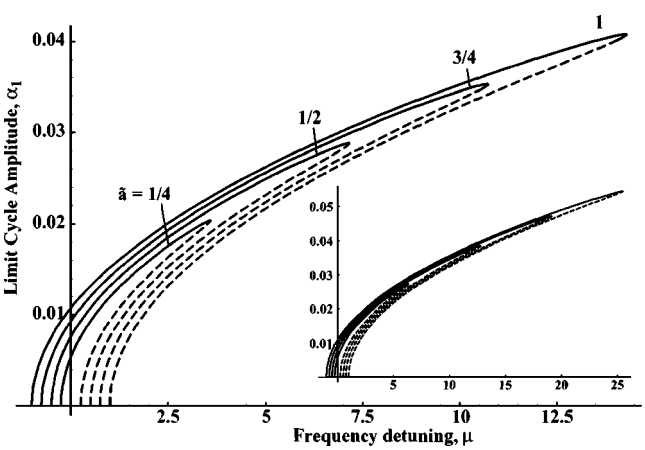


Fig. 4 The amplitude of the nontrivial limit cycles in the first principle parametric resonance ($n=l=1$) for $c=0.5$, $\zeta=10$, and $\varphi=400$. Inset shows solution neglecting steady state dissipation.

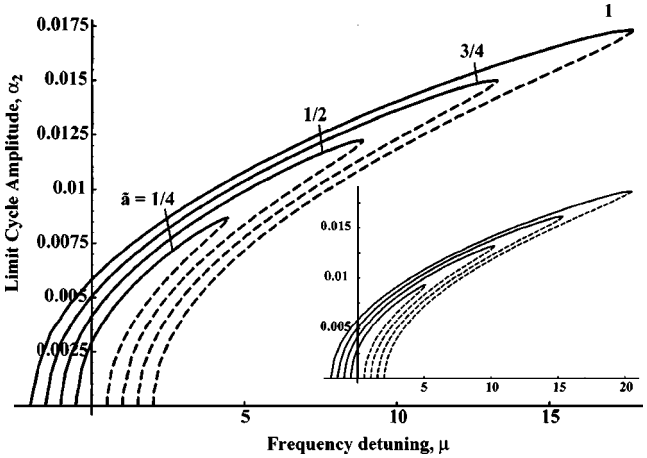


Fig. 6 The amplitude of the nontrivial limit cycles in the second principle parametric resonance ($n=l=2$) for $c=0.25$, $\zeta=10$, and $\varphi=400$. Inset shows solution neglecting steady state dissipation.

Conclusions

While the effects of viscous material behavior on translating strings have been studied previously, the way that it was incorporated in the mathematical model is only correct if the string is not translating. When the string is translating, the material time derivative in the constitutive relation is not simply the partial time derivative; a steady state component, proportional to the translation speed and the axial strain gradient, also appears. This steady state dissipation significantly effects the frequencies at which nontrivial limit cycles exist and also the amplitudes they obtain. The regions in parameter space in which the trivial solution is unstable is not affect by steady state dissipation; the strain gradient is zero in this trivial state.

The viscoelastic material model also significantly alters the excitation amplitude if the dynamic belt tensioning is caused by physically stretching the belt. In this case, the magnitude of the tension fluctuation depends on the frequency of the excitation.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CMS-0134064. 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

[1] Zhang, L., and Zu, J. W., 1999, ''Nonlinear Vibration of Parametrically Excited Viscoelastic Moving Belts, Part I: Dynamic Response,'' J. Appl. Mech., **66**, pp. 396–402.

[2] Zhang, L., and Zu, J. W., 1999, ''Nonlinear Vibration of Parametrically Excited Viscoelastic Moving Belts, Part II: Stability Analysis,'' J. Appl. Mech., **66**, pp. 403–409.

[3] Jha, R. K., and Parker, R. G., 2000, ''Spatial Discretization of Axially Moving Media Vibration Problems,'' J. Vibr. Acoust., **122**, pp. 290–294.

[4] Oz, H. R., 2001, ''On the Vibrations of an Axially Travelling Beam on Fixed Supports With Variable Velocity,'' J. Sound Vib., **239**, pp. 556–564.

[5] Oz, H. R., Pakdemirli, M., and Boyaci, H., 2001, ''Nonlinear Vibrations and Stability of an Axially Moving Beam With Time-Dependent Velocity,'' Int. J. Non-Linear Mech., **36**, pp. 107–115.

[6] Parker, R. G., 1999, ''Supercritical Speed Stability of the Trivial Equilibrium of an Axially-Moving String on an Elastic Foundation,'' J. Sound Vib., **221**, pp. 205–219.

[7] Parker, R. P., 1998, ''On the Eigenvalues and Critical Speed Stability of Gyroscopic Continua,'' J. Appl. Mech., **65**, pp. 134–140.

[8] Pakdemirli, M., and Ulsoy, A. G., 1997, ''Stability Analysis of an Axially Accelerating String,'' J. Sound Vib., **203**, pp. 815–832.

[9] Lengoc, L., and McCallion, H., 1996, ''Transverse Vibration of a Moving String: A Comparison Between the Closed-Form Solution and the Normal-Mode Solution,'' J. Syst. Eng., **6**, pp. 72–78.

[10] Mockensturm, E. M., Perkins, N. C., and Ulsoy, A. G., 1996, ''Stability and Limit Cycles of Parametrically Excited, Axially Moving Strings,'' J. Vibr. Acoust., **118**, pp. 346–351.

[11] Zhang, L., and Zu, J. W., 1999, ''Nonlinear Vibration of Parametrically Excited Moving Belts, Part I: Dynamic Response,'' J. Appl. Mech., **66**, pp. 396–402.

[12] Chen, L. Q., Zhang, N. H., and Zu, J. W., 2003, ''The Regular and Chaotic Vibrations of an Axially Moving Viscoelastic String Based on Fourth Order Galerkin Truncation,'' J. Sound Vib., **261**, pp. 764–773.

[13] Zhang, L., and Zu, J. W., 1998, ''Nonlinear Vibrations of Viscoelastic Moving Belts, Part I: Free Vibration Analysis,'' J. Sound Vib., **216**, pp. 75–91.

[14] Zhang, L., and Zu, J. W., 1998, ''Nonlinear Vibrations of Viscoelastic Moving Belts, Part II: Forced Vibration Analysis,'' J. Sound Vib., **216**, pp. 93–105.

[15] Chen, L. Q., Zhang, N. H., and Zu, J. W., 2002, ''Bifurcation and Chaos of an Axially Moving Viscoelastic String,'' Mech. Res. Commun., **29**, pp. 81–90.

[16] Hou, Z. C., and Zu, J. W., 2002, ''Nonlinear Free Oscillations of Moving Viscoelastic Belts,'' Mech. Mach. Theory, **37**, pp. 925–940.

[17] Wickert, J. A., and Mote, Jr., C. D., 1990, ''Classical Vibration Analysis of Axially Moving Continua,'' J. Appl. Mech., **57**, pp. 738–744.

Modeling of Threshold Strength in Cylindrical Ceramic Structures

Fjóla Jónsdóttir
Department of Mechanical and Industrial Engineering,
University of Iceland,
Reykjavik, Iceland
e-mail: fj@hi.is

Glenn E. Beltz
e-mail: beltz@engineering.ucsb.edu

Robert M. McMeeking
e-mail: mcm@engineering.usb.edu

Department of Mechanical and Environmental Engineering,
University of California,
Santa Barbara, CA 93106-5070

Recently, three-dimensional structured ceramic composites with large threshold strengths (i.e., stress below which there is zero probability of failure) have been fabricated utilizing an architecture consisting of relatively stress-free, elongated prismatic domains, separated by thin compressive walls. We build upon prior work on laminate architectures, with the common feature that these structures are all susceptible to fracture. Typically, these three-dimensional structures consist of thin shells of mullite that surround alumina. Cracks, originating from large flaws within the ceramic body, are arrested by the surrounding compressive layers until a specific stress level is attained (i.e., the threshold strength), resulting in a truncation of the strength distribution in the flaw region. A preliminary stress intensity solution has shown that this arrest is caused by a reduction of the crack driving force by the residual compression in the compressive walls. This solution also predicts that the threshold strength is dependent not only on the magnitude of the residual compression in the walls but also on the dimensions of both phases. A finite element model is presented that utilizes a penny-shaped crack in the interior of such a structure or half-penny-shaped crack emanating from the edge of such a structure. Ongoing analytical and experimental work that is needed to more fully understand this arrest phenomenon and its application towards the development of reliable, damage-tolerant ceramic components are discussed. [DOI: 10.1115/1.1831296]

1 Introduction

The major drawback of ceramics as structural materials is their brittleness. Brittle materials contain an unknown variety of cracks and flaws that are inadvertently introduced during processing and surface machining [1,2]. The high brittleness makes ceramic parts extremely prone to impact damage, often resulting in catastrophic failure. A concept which plays a central role in the study of crack arrest in brittle materials is the threshold strength—that is, a stress below which the probability of failure vanishes. This phenomenon increases the damage tolerance of ceramics and will allow engineers to design reliable ceramic components for structural applications. It has been shown by Rao et al. [3] and Hbaieb and McMeeking [4,5], that thin compressive layers, within a laminar ceramic, arrest large cracks, and produce a threshold strength. These laminates have been shown to exhibit threshold strengths in accordance with finite element modeling of the crack propagation process. Typically, the laminar plates are composed of alternating layers of Al_2O_3 and a mullite/Al_2O_3 mixture. Residual compressive stresses in the layers can arise due to differential strain between the layers caused by one or more of the following: differential thermal contraction (or expansion) during cooling (or heating), a change in volume due to a crystallographic phase transformation, or molar volume change associated with the formation of a reaction product.

Since laminates are simply two-dimensional structures, they are only effective at arresting a crack in *one* direction. Three-dimensional structured ceramic composites have been fabricated that yield a threshold strength in other dimensions. This is being accomplished by assembling fibers and spheres using colloidal processing techniques and coating these geometries with another material. The coated fibers and spheres are then consolidated to produce a material with a periodic structure that includes layers in compression that can arrest cracks propagating in all three dimensions [6]. One type of three-dimensional structure consists of relatively stress-free, elongated prismatic domains, separated by thin compressive walls (see Fig. 1). This is the central idea underlying the present study. To understand the mechanics of fracture in this architecture, a simple analytic model and several finite element analyses are carried out to study the threshold strengths for different configurations.

Our analysis proceeds in the following way. First, the three-dimensional ceramic structure is simplified as an infinitely long cylindrical structure and a crack is modeled as a penny-shaped crack in the interior. The longitudinal axis of the cylinder is perpendicular to the plane of the crack and is assumed to nominally be in tension. The problem is then extended to a semi-infinite structure with a half-penny-shaped crack emanating from the edge of such a structure. Stress intensity factors are calculated for these two configurations as a crack grows from the tensile section into the compressive section. The stress intensity factors are used to determine the threshold stress, that is, the level needed to extend the crack through the compressive layers to produce catastrophic failure.

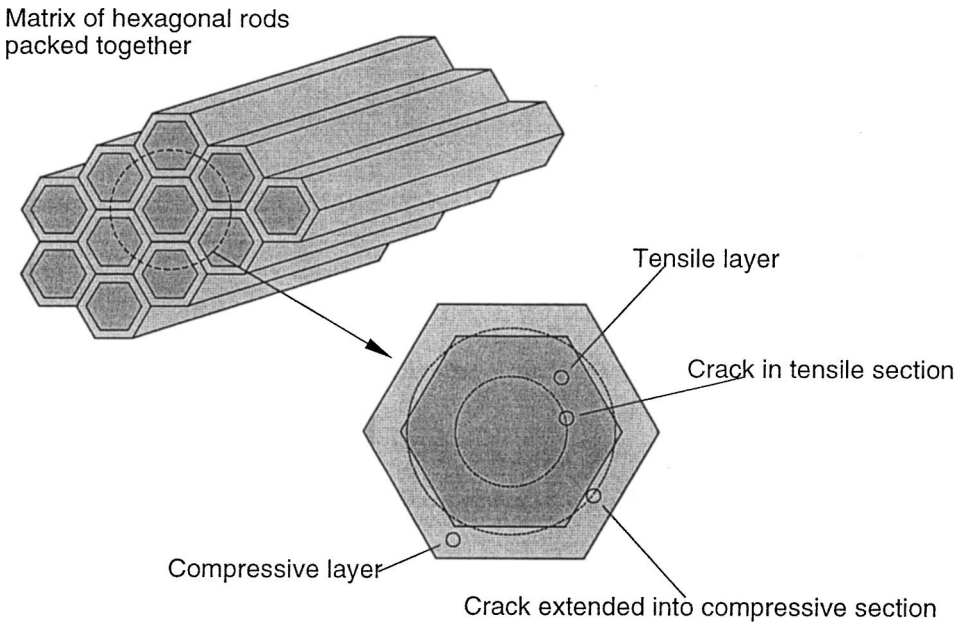
2 Problem Formulation

The physical system that provides the basis for the following discussion is a three-dimensional architecture consisting of elongated prismatic domains, separated by thin compressive walls, as shown schematically in Fig. 1.

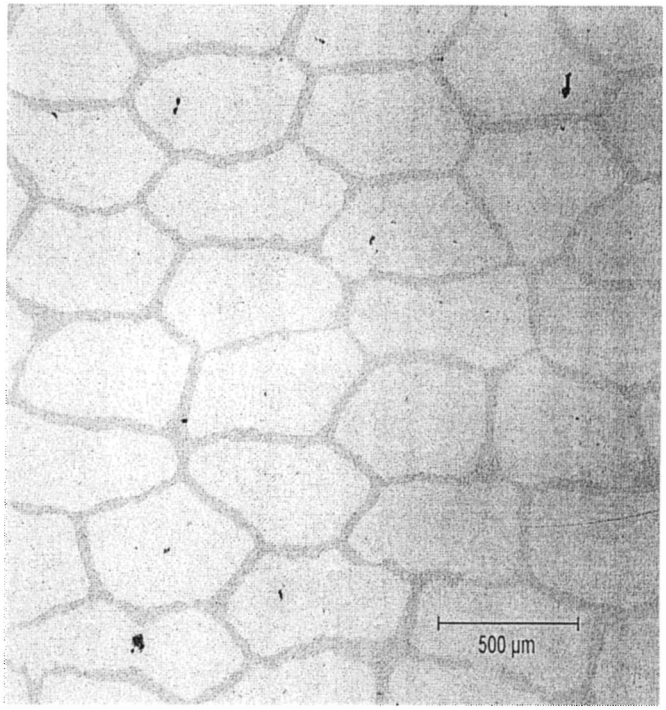
2.1 Analytic Model for an Idealized Cylindrical Structure. As an approximation, we assume the three-dimensional architecture to be an infinitely long cylindrical structure. The structure consists of concentric cylinders, alternating between tensile and compressive zones, the innermost being a tensile zone. As a representative model, Fig. 2 shows three concentric cylinders with radii r_a, r_b, and r_c, respectively. The thickness of the compressive layer is given by $t=r_b-r_a$. Assume a preexisting penny-shaped crack of diameter Fig. 2(*a*) spans the diameter of the tensile layer. In the following analyses, we determine the stress intensity factors for a crack when it extends into the compressive zone, that is, for $r_a<a<r_b$. The stress intensity factors are used to determine the applied threshold stress, σ_{thr}, needed to extend the crack through the compressive layers to produce catastrophic failure.

A stress intensity factor K can be determined by superimposing the two stress fields: the applied stress field and the residual stress

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, June 30, 2003; final revision, July 9, 2004. Associate Editor: K. M. Liechti.



(a)



(b)

Fig. 1 (*a*) **Schematic and** (*b*) **micrograph of a three-dimensional ceramic architecture. As a material example, the solid core consists of alumina** (Al_2O_3), **while the thinner, compressive coating-like phase consists of a mixture of mullite and** Al_2O_3 (**micrograph courtesy of M. Snyder**).

field, as depicted in Fig. 3. Each stress field is applied to the same penny-shaped crack of diameter $2a$ and each has its own known stress intensity factor.

Before we carry out the superposition, let us assume that the cylindrical structure is infinitely long and subject only to uniform tension, σ_a, at the remote boundary. The stress intensity factor for this case is readily available and given by Tada [7] as:

$$K_{\text{applied}} = 2\sigma_a \sqrt{\frac{a}{\pi}} F(a/r_c), \tag{1}$$

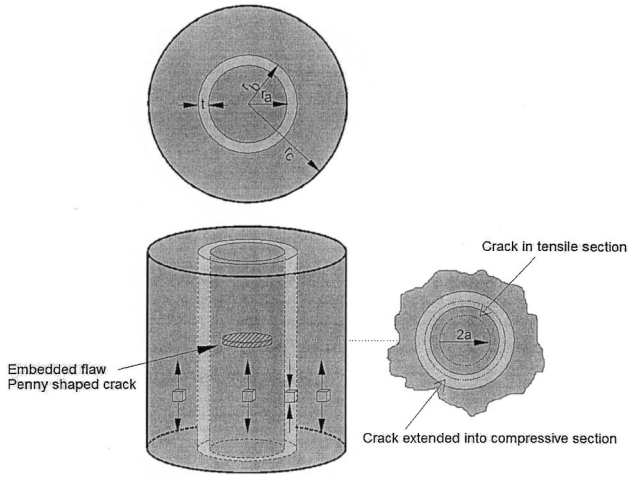


Fig. 2 Schematic of an infinite cylindrical structure containing a penny-shaped crack in its interior (tensile) phase

where $F(a/r_c)$ is a dimensionless correction function given by:

$$F(a/r_c)\approx \frac{1-0.5\dfrac{a}{r_c}+0.148\left(\dfrac{a}{r_c}\right)^3}{\sqrt{1-\dfrac{a}{r_c}}}. \quad (2)$$

The correction function depends on geometry of the structure and, as Eq. (2) shows, as radial dimension of the cylinder approaches infinity ($r_c\rightarrow\infty$), the correction function approaches unity, that is $F(a/r_c)\rightarrow 1$. For purposes of this study, it is not feasible to use an infinite value for r_c and we must accept a finite value, which is reflective of the real three-dimensional structure. For our geometries, the magnitude of $F(a/r_c)$ ranges from about 1.01 to about 1.375. The arbitrariness in the choice of r_c, which sets the background stress, is one disadvantage of the using cylindrical model to represent the stresses in what really is a periodic structure.

We now return our attention to the superposition scheme outline in Fig. 3. The applied stress is σ_a, the magnitude of the residual compression is σ_c (defined to be a positive number) and the residual tension is denoted as σ_t. The first stress field on the right-hand side of the "equality" is a tensile stress of magnitude ($\sigma_a-\sigma_c$), applied at the remote boundary, to a cracked specimen that does *not* contain residual stresses (with $F\approx 1$). The stress intensity factor for this stress is given by the first term on the right side of Eq. (3). The second stress field is a tensile stress of magnitude ($\sigma_c+\sigma_t$), applied only to the crack within the tensile region. Its stress intensity factor is given by the second term on the right side of Eq. (3). The two superimposed stress fields sum to that shown on the left-hand side of Fig. 3. The stress intensity factor for the two superimposed stress fields is thus given by:

$$K=(\sigma_a-\sigma_c)\frac{2}{\pi}\sqrt{\pi a}+\frac{2}{\sqrt{\pi a}}(\sigma_c+\sigma_t)\int_0^{r_a}\frac{\xi}{\sqrt{a^2-\xi^2}}d\xi. \quad (3)$$

The integrand in Eq. (3) is due to a ring load of radius ξ which is integrated with intensity $\sigma_t+\sigma_c$ up to a radius r_a. Evaluating the integral and simplifying gives:

$$K=2\sigma_a\sqrt{\frac{a}{\pi}}+2\sigma_t\sqrt{\frac{a}{\pi}}-2(\sigma_c+\sigma_t)\sqrt{\frac{a}{\pi}}\sqrt{1-\left(\frac{r_a}{a}\right)^2}. \quad (4)$$

The first term in Eq. (4) is recognized as the stress intensity factor for a penny-shaped crack in an applied tensile field, while the remainder of the expression is negative. Thus, the stress intensity factor initially decreases when the crack extends into the compressive shell of the material, and fracture resistance correspondingly increases. The analytical result in Eq. (4), for the stress intensity factor, is compared with calculated stress intensity factor, in Section 3.

Using elasticity theory, it can be shown that the magnitude of the axial tensile stress ($0<r<r_a$ and $r_b<r<r_c$) is given by:

$$\sigma_t=\frac{E'\Delta\alpha\Delta T t(t+2r_a)}{r_c^2}, \quad (5)$$

and, similarly, the magnitude of the axial compressive stress ($r_a<r<r_b$) is given by:

$$\sigma_c=\frac{E'\Delta\alpha\Delta T(r_c^2-t^2-2tr_a)}{r_c^2}, \quad (6)$$

where $E'=E/(1-\nu)$, E is Young's modulus, ν is Poisson's ratio, $\Delta\alpha$ is the difference in thermal expansion between the two materials, and ΔT is the temperature relative to a datum at which the

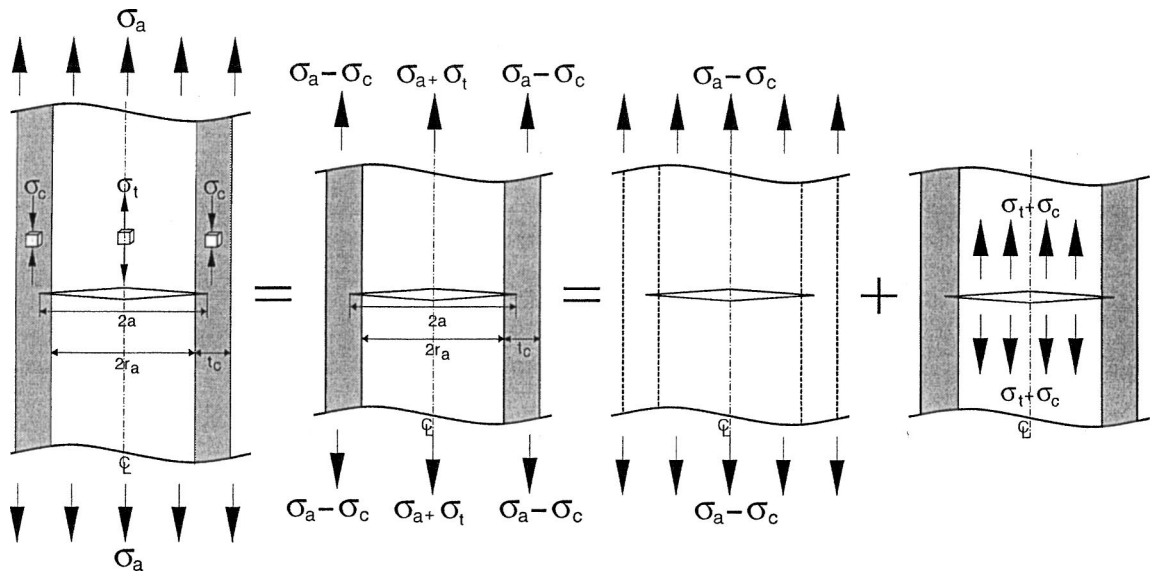


Fig. 3 Stresses in a loaded cylindrical ceramic architecture can be obtained via superposition

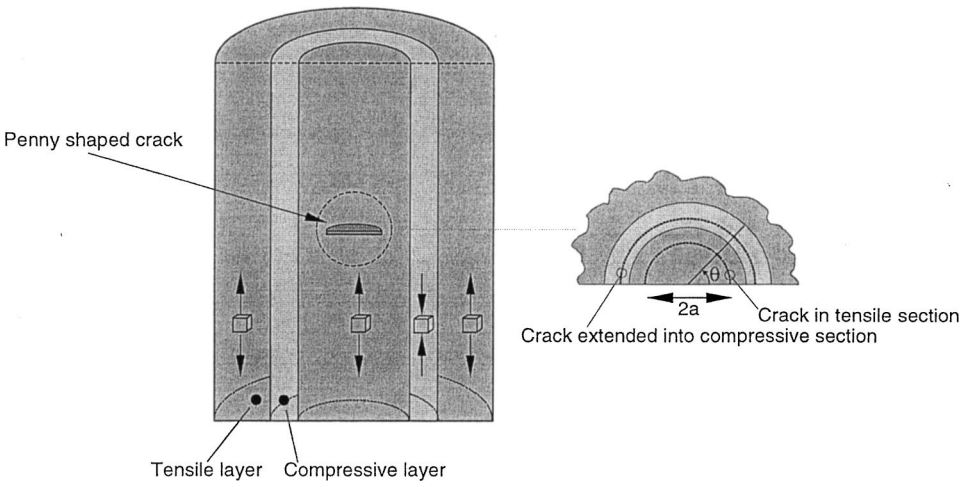


Fig. 4 Schematic of the semi-infinite cylindrical structure with a half-penny shaped crack emanating from the free surface

thermal residual stresses are zero. In Eqs. (5) and (6), $\Delta\alpha\Delta T$ is taken to be a positive number. The derivation of Eqs. (5) and (6) is given in Appendix A.

In general, the stress intensity factor in Eq. (4) decreases monotonically as the crack grows through the compressive layer and, hence, the crack grows stably until it reaches the interface with the next tensile zone as the applied stress is increased. However, it can be shown that above a critical value of applied stress, the function $K(a)$ reaches a local minimum in the compressive zone. If the crack were to reach this location, it would continue to grow unstably until reaching the interface with the tensile layer. Further discussion of this critical applied stress is given in Appendix B. In this paper, we avoid parameter regimes that lead to a minimum in $K(a)$ prior to the crack reaching the tensile zone.

Assuming the threshold stress occurs when the crack has reached the interface with the tensile zone, one can identify K with K_c and a with r_a+t and solve for σ_a to arrive at:

$$\sigma_{thr}=\frac{K_c}{2}\sqrt{\frac{\pi}{r_a+t}}+\frac{\sigma_c t(t+2r_a)}{t(t+2r_a)-r_c^2}+\sigma_c\left(\frac{r_c^2}{r_c^2-t(t+2r_a)}\right)\sqrt{1-\left(\frac{1}{1+t/r_a}\right)^2}. \quad (7)$$

By substituting Eq. (6) into (7) we can write the normalized threshold stress as:

$$\frac{\sigma_{thr}}{E'\Delta\alpha\Delta T}=\frac{K_c}{2E'\Delta\alpha\Delta T\sqrt{r_a}}\sqrt{\frac{\pi}{1+t/r_a}}-\left(\frac{r_a}{r_c}\right)^2\left(\frac{t}{r_a}\right)\left(\frac{t}{r_a}+2\right)+\sqrt{1-\left(\frac{1}{1+t/r_a}\right)^2}. \quad (8)$$

Equations (7) and (8) show that the threshold strength for a cylindrical composite increases with the fracture toughness of the thin-layer material, the magnitude of the compressive stress and the thicknesses of the various layers. These expressions are analogous to those worked out for laminate architectures in earlier work [3–5], in that they give very similar trends with regard to the variation of σ_{thr} with crack geometry. Most importantly, they allow one to design cylindrical ceramic architectures with the knowledge that failure will not occur below this value of stress.

This theoretical model ceases to apply when a variety of realistic effects prevail. For example, elastic mismatch is not accounted for; that is, we assume the effective Young's modulus E' is identical in both phases. In addition, we assume the crack propagates radially on its original plane through the various phases. While this straight crack propagation has been observed to occur in many experiments, a phenomenon known as bifurcation can alternatively occur, where the crack may branch from its original plane after penetrating into the compressive layer [3]. This effect has been shown to increase the threshold strength beyond what is calculated here, but does imply that the physical mechanisms considered in this section are not universally applicable. While the finite element method discussed in the next section can be extended to consider cases that involve elastic modulus mismatch and bifurcation, we submit that the results presented in this work still provide invaluable guidance on the design of three-dimensional architectures that are fracture resistant. In addition, the current work provides a level of confidence before extending the FEM model to more complex geometries.

2.2 Cracks Emanating From a Surface. The second case we consider is a half of a concentric cylindrical structure with a half-penny-shaped crack emanating from the edge, as shown schematically in Fig. 4. The motivation for this geometry is that ceramic composites of this type are typically tested in bending, with surface cracks initiating from the surface in maximum tension.

For a half-space with a half-penny-shaped crack emanating from the edge and subject to tensile loading at the remote boundary, the stress intensity factor is well known and is given by Tada [7] as:

$$K=\frac{2}{\pi}\sigma_a\sqrt{\pi a}F(\theta), \quad (9)$$

where $F(\theta)$ is given by:

$$F(\theta)\approx 1.211-0.186\sqrt{\sin\theta}\quad (10°<\theta<170°). \quad (10)$$

Given that we are using a finite value for the cylindrical diameter, the result by Tada can only be used as an approximation to our results.

Equations (9) and (10) show the stress intensity factor is dependent on the angle θ, measured from the edge of the structure. However, this dependence is relatively weak. For a crack emanating from a free surface, the state of stress varies from plane strain in the interior of the plate to plane stress at the surface. Hence, using a crack-opening displacement method to calculate the stress intensity factors can give erroneous results so Eq. (10) is limited to internal angles. Raju and Newman [8] use a nodal-force method, which requires no prior assumption of either plane stress or plane strain, to obtain the stress intensity factors of semielliptical surface cracks. Their results seem to suggest that the stress

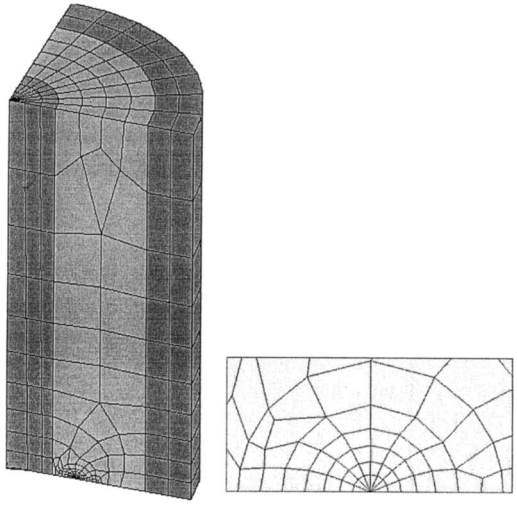


Fig. 5 Typical finite element model with singular elements around the crack tip

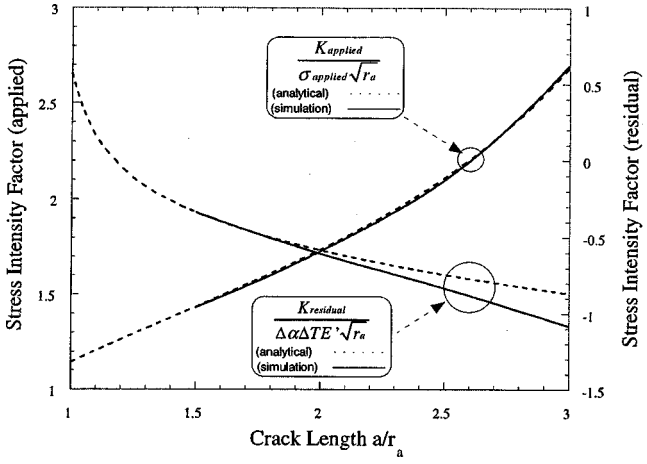


Fig. 6 Comparison of calculated and theoretical stress intensity factors. Tensile and compressive layers have equal thicknesses. Thickness of tensile layer is $2r_a=200\ \mu$m.

intensity factor varies strongly near the surface, indicating the need for substantial mesh refinement in this regime. Because of this complication we defer consideration of cracks at the point where they intersect the surface to future work.

For the case of residual stresses in a semi-infinite cylindrical structure there is no analytical solution available for stress intensity factors and we must rely solely on finite element analysis.

3 FEM Models for Stress Intensity Factors

Stress intensity factors are calculated using the commercial finite element code ANSYS [9]. Recall the two cases considered; that is, an infinitely long cylindrical structure with a penny-shaped crack in the interior and a semi-infinite structure with a half-penny-shaped crack emanating from the edge.

3.1 Full Penny-Shaped Crack. We first consider the structure of concentric cylinders with an embedded penny-shaped crack, as shown in Fig. 2. Figure 5 shows a typical finite element model with eight wedges. Given the symmetry of the problem, only one-eighth of the body is modeled. The 20-node brick elements are used in the analysis. The first row of elements around the crack tip is modeled with singular elements, with the midside nodes placed at the quarter points, to account for the $r^{-1/2}$ singularity in stresses and strains at the crack tip. The stress intensity factors are calculated with a displacement extrapolation method as outlined in the ANSYS theory manual [9].

A typical dimension for the Al_2O_3 tensile cells in the three-dimensional architecture is $2r_a=450\ \mu$m. For the compressive layers, a mixture of mullite and Al_2O_3, typical dimensions range from 23 to 90 μm [10]. Two configurations are considered here: one in which the thickness of the compressive layer is equal to the diameter of the tensile cell, that is, $t=2r_a$, and with $2r_a=200\ \mu$m; one in which the compressive layer is one tenth the diameter of the tensile cell, that is, $t=2r_a/10$, for a thickness of tensile layer $2r_a=450\ \mu$m. In the former, $t/2r_a=1$, and in the latter, $t/2r_a=1/10$. In both configurations, the elastic constants of the tensile and compressive zones are considered to be identical. The reason we choose two different thickness ratios is that the smaller one is comparable to the experimental dimensions used by Lange et al. [6,10] and the larger one is comparable to ratios used in finite element studies on laminates [4].

Figure 6 shows the calculated stress intensity factors for a thickness ratio of $t/2r_a=1$. The results are plotted versus normalized crack length, as the crack extends from the tensile layer through the compressive layer. The results are shown separately for the stress intensity factor K_{applied} due to the externally applied load, and for the stress intensity factor K_{residual} due to the residual stress caused by thermal mismatch. The theoretical results for K_{applied} and K_{residual}, given by Eq. (4), are plotted on the same graph for comparison. As the figure shows, there is good agreement between the theoretical and computed results.

In Fig. 7, we show the calculated stress intensity factors for a thickness ratio of $t/2r_a=1/10$. Again, the results are shown separately for K_{applied} due to the externally applied load, and K_{residual} due to the material mismatch. The theoretical results for K_{applied} and K_{residual} are plotted on the same graph for comparison. As in the previous case, there is good agreement between the theoretical and calculated results.

3.2 Half-Penny-Shaped Crack Emanating From the Surface. We next consider the second geometry—half of a cylindrical structure with a half-penny-shaped crack emanating from the edge, as shown in Fig. 4. The same two configurations are considered as for the full-penny-shaped crack, that is, a configuration where the thickness of the compressive layer is equal to the diameter of the tensile cell, and a configuration where the compressive layer is one tenth the diameter of the tensile cell. The finite element calculations are carried out in a similar fashion as in the previous section. Only one-fourth of the body is modeled, given the symmetry of the problem.

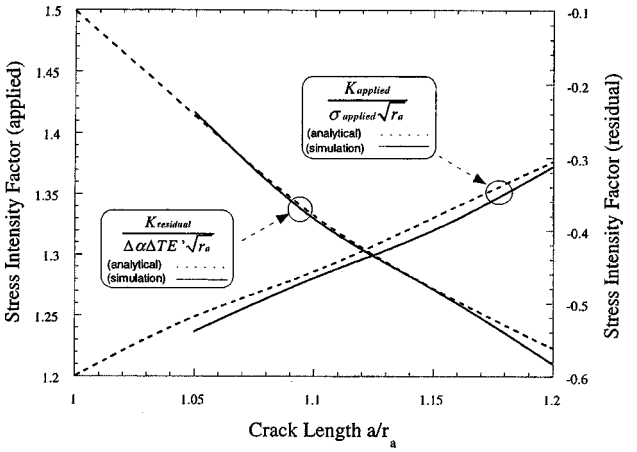


Fig. 7 Comparison of calculated and theoretical stress intensity factors. Thickness of compressive layer is one tenth the diameter of the tensile zone. Diameter of tensile zone is $2r_a=450\ \mu$m.

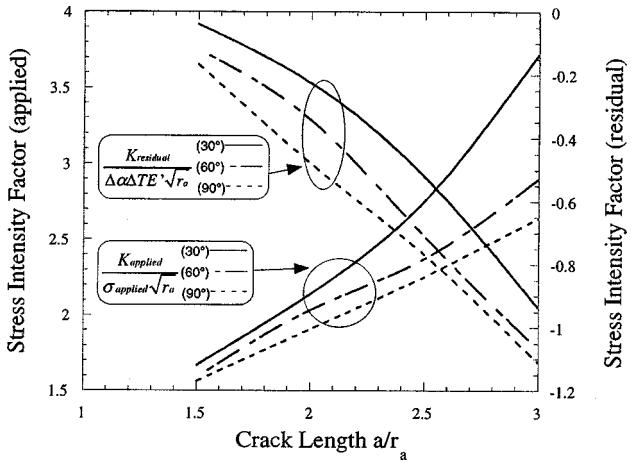


Fig. 8 Comparison of calculated and theoretical stress intensity factors. Tensile and compressive layers have equal thicknesses. Thickness of tensile layer is $2r_a = 200\ \mu$m.

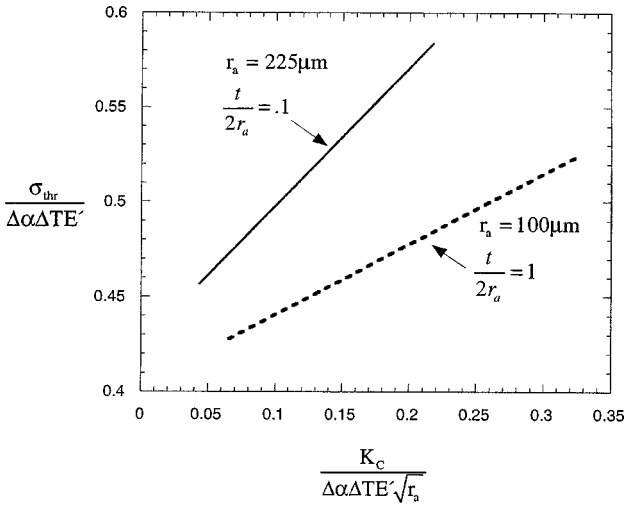


Fig. 10 Threshold strength versus fracture toughness for a full-penny shaped crack

Figure 8 shows the calculated stress intensity factors for the thickness ratio of $t/2r_a = 1$. Again, the results are shown separately for K_{applied} due to the externally applied load, and K_{residual} due to the material mismatch. The results are shown for three values of angle θ, 30 deg, 60 deg, and 90 deg. As expected, and suggested by Eq. (9), the stress intensity factor decreases as the angle θ increases.

Figure 9 shows the calculated stress intensity factors for the thickness ratio of $t/2r_a = 1/10$, also for the same three values of angle θ, 30 deg, 60 deg, and 90 deg.

Now that the stress intensity factors have been calculated, the next step in our analysis is the determination of threshold strength, which we take up in the following section.

4 Discussion of Threshold Strength

As discussed in Sec. 2.1, the stress intensity factor generally decreases as the crack extends into the compressive layers. Thus, the maximum stress needed to drive the crack through the compressive layers occurs when the crack is at the interface between the compressive and tensile zones, that is, when $a = r_a + t = r_b$. In the context of the superposition concept introduced in Sec. 2.1, we can set $K = K_{\text{applied}} + K_{\text{residual}} = K_c$; hence, $K_{\text{applied}} = K_c - K_{\text{residual}}$, and then solve for the applied stress σ_a, which appears as a linear prefactor in K_{applied}. Thus, the largest stress needed to extend the crack through the compressive zone is given in a normalized form as:

$$\frac{\sigma_{thr}}{\Delta\alpha\Delta TE'} = \frac{\dfrac{K_c}{\Delta\alpha\Delta TE'\sqrt{r_a}} - \dfrac{K_{\text{residual}}(r_b)}{\Delta\alpha\Delta TE'\sqrt{r_a}}}{\dfrac{K_{\text{applied}}(r_b)}{\sigma_{\text{applied}}\sqrt{r_a}}}. \tag{11}$$

If the initial crack size in the tensile layer is less than r_a and the stress needed to extend it is less than σ_{thr}, the crack will be arrested by the compressive layer. However, if the crack is very small and extends at a stress greater than σ_{thr}, it will extend through the compressive layer and lead to catastrophic failure without being arrested. Thus, Eq. (11) is rigorously thought of as a threshold stress. If the applied stress is less than σ_{thr}, the body should not fail when tensile stress is applied along the fibers. As previously noted, this enables load-bearing components to be designed with the foreknowledge that failure is unlikely to occur below that stress.

As expected, Eq. (11) shows that the threshold strength increases with the fracture toughness of the compressive layers, K_c. The normalized threshold strengths are plotted in Figs. 10 and 11, against the normalized fracture toughness K_c of the compressive layer. Values of K_c are chosen between 1 and 5 MPa$\sqrt{\text{m}}$, a range which is typical of ceramic materials. Values of other material parameters are taken as $E = 300$ GPa, $\nu = 0.3$, $\Delta\alpha = 2.795 \times 10^{-6}\,\text{C}^{-1}$ and $\Delta T = -1200°\text{C}$, for purposes of setting the ranges of these plot axes. Taking $K_c = 3$ MPa$\sqrt{\text{m}}$ and $t/2r_a = 1/10$, typical of the cylindrical structure by Snyder [10], the threshold strength we arrive at is ~800 MPa. This modestly exceeds what has been observed in that system, but other effects, such as edge cracking and crack branching into other propagation planes (as well as the fact that modulus mismatch is not accounted for here) are being considered as mechanisms that are coming into play in the experimental system.

In addition to the effect of intrinsic fracture toughness, the finite element results reveal the effect of mismatch strain $\Delta T\Delta\alpha$, albeit in an indirect way due to the normalization we have chosen to use in Figs. 10 and 11. Inspection of Eq. (11), coupled with the fundamental result that we expect K_{residual} to vary linearly with $\Delta T\Delta\alpha$ (and that the K_{residual} in the numerator of the equation is actually expected to be negative), yields the intuitive result that σ_{thr} increases with mismatch strain.

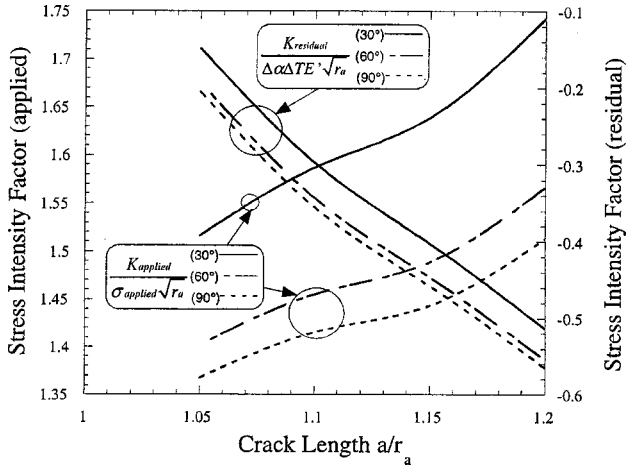


Fig. 9 Comparison of calculated and theoretical stress intensity factors. Thickness of compressive layer is one tenth the diameter of the tensile zone. Diameter of tensile layer is $2r_a = 450\ \mu$m.

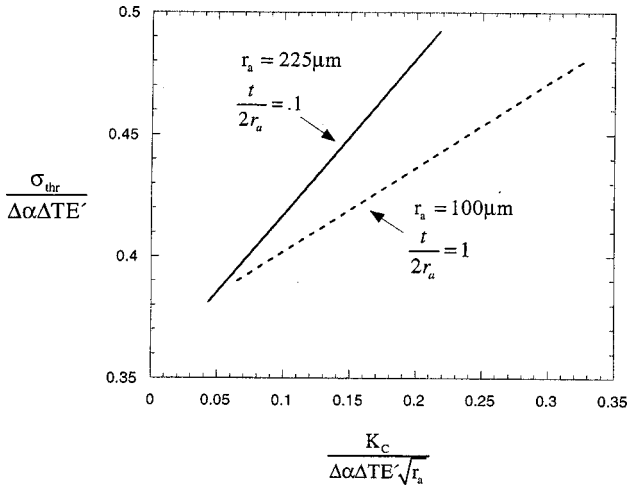


Fig. 11 Threshold strength versus fracture toughness for a half-penny shaped crack. Results are shown for angle θ=60°.

The effect of tensile region size $2r_a$ and compressive layer thickness t, and the ratio $t/2r_a$, are even less transparent in the finite element results, simply because we did not perform an extensive parametric study in this space. Nevertheless, useful insight can be gained from the simple analytical model through Eq. (8) which shows that the threshold strength decreases with the thicknesses of the various layers. The ratio $t/2r_a$ has a modest effect; in addition the absolute size of the tensile zone, r_a, impacts the threshold stress as well. The latter effect is being exploited by Paranjpye et al. [11] through microelectromechanical systems (MEMS) processing technology to achieve threshold stresses in laminate systems in excess of 1 GPa.

As to be expected, the threshold strength depends on the elastic moduli of the tensile and compressive layers. With everything else held fixed, if the tensile layer were more stiff than the compressive layer, the magnitude of the residual stress rises and hence the threshold strength increases as is apparent from Eq. (7). While we have not performed a systematic study of cases where the compressive layer elastic properties differ from those in the tensile zones, the good agreement that has been observed in this study between the FEM results and the analytic results provide the necessary confidence necessary to build elastic mismatch into future implementations of this FEM model. In addition, more sophisticated procedures, such as considering a periodic structure based on a hexagonal compressive layer configuration (depicted in Fig. 1), and using the J-integral to calculate stress intensity factors, should be explored.

5 Summary

The finite element method was used to predict threshold strengths in a model system consisting of a cylindrical jacket under residual compression, surrounded by regions of tensile material, subject to tensile loading aligned with the cylindrical major axis. The model system has relevance to ceramic composites that have been fabricated by consolidating fibers of one phase in another at high temperature, followed by cooling, resulting in residual compression in the phase surrounding the original fibers. The architecture offers superior mechanical response, in that cracks which originate in the cylindrical zones may be arrested by the surrounding compressive layers, resulting in a truncation of the strength distribution with respect to flaw size and an associated design threshold strength. A simple fracture mechanics model, valid for similar elastic properties is presented, and the finite element results are in good agreement with that analysis. Moreover, the finite element model is extended to the case of a half-penny crack emanating from a traction-free surface. As expected from prior experience with laminate systems, the threshold strength is shown to depend on the mismatch strain (through the thermal expansion coefficient mismatch and temperature change), the intrinsic toughness of the constituent materials, and the thickness ratio. The results are in modest agreement with experimental results.

Acknowledgments

The authors would like to thank Fred Lange, Mark Snyder, and Geoff Fair for illuminating discussions.

Appendix A

In this Appendix, we overview the derivation of the background stress field that drives crack propagation in the cylindrical, axisymmetric problem discussed in this paper. Consider the geometry depicted in Fig. 2, albeit without a crack. The elastic moduli are taken as identical in all three layers. The coefficient of thermal expansion in the interior layer (of thickness t) is taken as α_b, and that in the remaining layers is taken as α_a. The stress equilibrium equations, written in cylindrical coordinates and assuming no body forces, reduce to

$$\frac{\partial \sigma_{rr}}{\partial r} + \frac{1}{r}(\sigma_{rr} - \sigma_{\theta\theta}) = 0$$
$$\frac{\partial \sigma_{zz}}{\partial z} = 0. \tag{A1}$$

Moreover, compatibility requirements dictate that

$$\frac{\partial}{\partial r}(r\varepsilon_{\theta\theta}) = \varepsilon_{rr}. \tag{A2}$$

Equation (A2) follows from the fact that displacements in the θ direction vanish, and the displacement component in the r direction may only depend on r. In addition, we insist that ε_{zz} remain constant throughout the structure. Hooke's Law is written as

$$\varepsilon_{rr} = \frac{1}{E}[\sigma_{rr} - \nu(\sigma_{\theta\theta} + \sigma_{zz})] + \alpha_i \Delta T$$
$$\varepsilon_{\theta\theta} = \frac{1}{E}[\sigma_{\theta\theta} - \nu(\sigma_{rr} + \sigma_{zz})] + \alpha_i \Delta T, \tag{A3}$$
$$\varepsilon_{zz} = \frac{1}{E}[\sigma_{zz} - \nu(\sigma_{\theta\theta} + \sigma_{rr})] + \alpha_i \Delta T$$

with the subscript on α taken to coincide with the appropriate phase. The symmetry of the deformation dictates that all shear quantities vanish.

Inserting ε_{rr} and $\varepsilon_{\theta\theta}$ from Eq. (A3) into Eq. (A2), and eliminating $\sigma_{\theta\theta}$ via Eq. (A1), yields a linear ordinary differential equation for σ_{rr} that leads to a general solution of the form

$$\sigma_{rr} = C_1 + C_2/r^2; \quad \sigma_{\theta\theta} = C_1 - C_2/r^2, \tag{A4}$$

where the constants C_1 and C_2 must be determined separately for each phase, resulting in six (6) unknowns. We note that Eq. (A4) is constant with a piecewise constant solution for σ_{zz}, consistent with the second part of Eq. (A1). Additional boundary conditions are imposed in order to determine the constants: (1) the stress components must remain finite as $r \to 0$; hence, C_2 vanishes for the inner phase; (2) the outer surface of the structure is free of traction; hence, σ_{rr} is taken as zero at $r = r_c$; (3) continuity of $\varepsilon_{\theta\theta}$ is enforced at both $r = r_a$ and $r = r_b$; and (4) continuity of σ_{rr} is enforced at both $r = r_a$ and $r = r_b$.

There remains a seventh unknown, the constant value of ε_{zz}, which is found through a macroscopic force balance. That is, St. Venant's principle is exploited to write

$$\pi r_a^2 \sigma_{zz}(0 \leqslant r < r_a) + \pi(r_b^2 - r_a^2)\sigma_{zz}(r_a \leqslant r < r_b) + \pi(r_c^2 - r_b^2)\sigma_{zz}(r_b \leqslant r \leqslant r_c) = 0. \quad (A5)$$

With the unknowns in hand, the third of Eq. (*A*3) may be used to write the longitudinal stress in each phase:

$$\sigma_{zz} = \begin{cases} \dfrac{E\Delta T(\alpha_b - \alpha_a)(r_b^2 - r_a^2)}{(1-\nu)r_c^2} & \text{for } 0 \leqslant r < r_a \\ \dfrac{E\Delta T(\alpha_b - \alpha_a)(r_b^2 - r_a^2 - r_c^2)}{(1-\nu)r_c^2} & \text{for } r_a \leqslant r < r_b. \\ \dfrac{E\Delta T(\alpha_b - \alpha_a)(r_b^2 - r_a^2)}{(1-\nu)r_c^2} & \text{for } r_b \leqslant r \leqslant r_c \end{cases} \quad (A6)$$

The solution maps to Eqs. (5) and (6) by substituting $\Delta\alpha = \alpha_b - \alpha_a$, $r_b = r_a + t$, and noting that $\sigma_c \equiv -\sigma_{zz}$ for the sandwiched (compressive) layer.

Appendix B

When calculating the threshold strength, it is usually assumed that Eq. (4) decreases continuously throughout the regime $r_a < a < r_b$. Taking the first derivative of $K(a)$ and insisting that it must remain negative yields the following restriction:

$$\frac{(\sigma_a + \sigma_t)a\sqrt{a^2 - r_a^2}}{(\sigma_c + \sigma_t)(a^2 + r_a^2)} < 1. \quad (B1)$$

We note that the left-hand side of Eq. (*B*1) increases monotonically in *a*, so the first location for which the derivative ceases to be negative, at some prescribed level of σ_a, would occur at $a = r_a + t$ (i.e., the farthest extent within the compressive zone). Solving Eq. (*B*1) as an identity gives the critical value:

$$\sigma_{\text{crit}} = \frac{(\sigma_c + \sigma_t)(2r_a^2 + 2r_a t + t^2)}{(r_a + t)\sqrt{t^2 + 2r_a t}} - \sigma_t. \quad (B2)$$

Thus, the stress intensity factor undergoes a minimum within the compressive zone for $\sigma_a > \sigma_{\text{crit}}$, and Eq. (7) becomes invalid for the threshold strength. By equating Eq. (*B*2) with Eq. (7), a restriction on material parameters that guarantees stable crack growth across the entire compressive zone can be obtained. McMeeking and Hbaieb [5] have derived similar results for the case of a two-dimensional laminar composite.

References

[1] Green, D. J., 1998, *Introduction to Mechanical Properties of Ceramics*, Cambridge University Press, Cambridge, UK.
[2] Lange, F. F., 1989, "Powder Processing Science and Technology for Increased Reliability," J. Am. Ceram. Soc., **72**, pp. 3–15.
[3] Rao, M. P., Sanchez-Herencia, A. J., Beltz, G. E., McMeeking, R. M., and Lange, F. F., 1999, "Laminar Ceramics That Exhibit a Threshold Strength," Science, **286**, pp. 102–105.
[4] Hbaieb, K., and McMeeking, R. M., 2002, "Threshold Strength Predictions for Laminar Ceramics With Cracks That Grow Straight," Mech. Mater., **34**, pp. 755–772.
[5] McMeeking, R. M., and Hbaieb, K., 1999, "Optimal Threshold Strength of Laminar Ceramics," Z. Metallkd., **90**, pp. 1031–1036.
[6] Lange, F. F., Snyder, M., and Fair, G. (unpublished).
[7] Tada, H., 1985, *The stress analysis of cracks handbook*, Paris Productions Inc., St. Louis, MO.
[8] Raju, I. S., and Newman, J. C., 1979, "Stress-Intensity Factors for a Wide Range of Semi-Elliptical Surface Cracks in Finite-Thickness Plates," Eng. Fract. Mech., **11**, pp. 817–829.
[9] ANSYS, V.5.7, ANSYS Inc., Canonsburg, PA.
[10] Snyder, M. (private communication).
[11] Paranjpye, A., MacDonald, N. C., and Beltz, G. E., "Modeling and Simulation in Materials Science and Engineering," in press.

P. B. N. Prasad
Postdoctoral Researcher

Norio Hasebe
Professor

X. F. Wang
Research Associate

Department of Civil Engineering,
Nagoya Institute of Technology,
Gokiso-Cho, Showa-Ku,
Nagoya 466, Japan

Y. Shirai
Engineer
Shizuoka Prefectural Government,
Shizuoka, Japan

Green's Function of a Bimaterial Problem With a Cavity on the Interface—Part I: Theory

The problem of a point dislocation interacting with an elliptical hole located on a bimaterial interface is examined. Analytical solution is obtained by employing the techniques of complex variables and conformal mapping. A rational mapping function is used to map a half-plane with a semielliptical notch onto a unit circle. In the first part of this paper, complex potentials for the bimaterial system with an elliptical hole on the interface is derived when a point dislocation is present in the upper half-plane without loss of generality. The solution derived can be used as Green's function to study internal cracks interacting with an elliptical interfacial cavity. [DOI: 10.1115/1.1876432]

1 Introduction

Composites are widely employed in present day applications. Their tailor made properties enable them to be used in varied environments. Interface between dissimilar materials is an essential feature of composites. Interfaces are relatively weak compared to adjoining materials and may debond easily. Debonding of interfaces leads to the formation of cavities and cracks. Interfacial cracks/cavities interacting with subinterfacial cracks is also a common feature in composite materials.

Williams [1] used eigenfunction expansion method to solve the problem of a crack lying along the interface of two bonded dissimilar elastic materials and derived the stress fields near the crack tip. Rice and Sih [2], England [3], and Erdogan [4] employed complex variables to obtain analytical solution for the problem of an interfacial crack along two bonded dissimilar elastic half-planes. Analytical solution reveals oscillating stress fields ahead of the crack tip and interpenetration of crack faces behind the crack tip but these anomalies were found to be confined to a small region and hence could be neglected [5]. To overcome the above anomalies Comninou [6,7] proposed a model by treating interface crack as a continuous distribution of dislocations.

The problems of internal cracks approaching interfaces have been examined using the distributed dislocation method and solving the resulting singular equations to obtain stress intensity factors [8,9]. Several problems pertaining to interfacial cracks have been summarized in Murakami et al. [10]. Recently, Oda et al. [11] used the body force method to study the problem of an interface crack interacting with an internal crack.

Hasebe et al. [12] and Okumura et al. [13] modeled the interfacial cavity as an elliptical hole and solved the problem of the bonded bimaterial plane with an elliptical hole at the interface and loaded it parallel to and normal to the interface, respectively. They employed the rational mapping function approach and studied stress concentration effects due to interfacial cavities. It should be noted that crack is a degenerate case of an elliptical hole.

The problem of a point dislocation interacting with an elliptical hole at the interface of dissimilar materials is not solved yet. It is well known that point dislocation solutions can be employed to simulate cracks by means of a distributed dislocation technique. In this paper, the solution for the problem of a point dislocation in the upper half-plane interacting with an elliptical hole on the interface of bonded dissimilar materials is obtained. Mapping of a half-plane with a semielliptical notch onto a unit circle is done by means of a rational mapping function and the problem is solved by the complex variable method and a closed form solution is obtained.

2 Derivation of Solution

The problem under consideration consists of two half-planes with semielliptical notches bonded along the common boundary as shown in Fig. 1(*a*). Material 1 occupies $Y \geqslant 0$ and material 2 occupies $Y \leqslant 0$. Bonded parts of the interface is denoted by M and the unbonded parts by L_j where $j=1,2$ denote materials 1 and 2, respectively. Shear modulus and Poisson's ratio for materials 1 and 2 are denoted by μ_j and ν_j, respectively. Materials 1 and 2 are considered separately in z_1 and z_2 planes as shown in Figs. 1(*b*) and 1(*c*). The z_1 plane is obtained by rotating material 1 by 180 deg about the X-axis while material 2 is undisturbed in the z_2 plane. Since the geometry of z_j planes are identical, same mapping function is used to map z_j planes onto unit circles as shown in Fig. 1(*c*). S^+ and S^- denote the regions inside and outside the unit circle. Point dislocations $D_{01}=D_{x1}+iD_{Y1}$ and $-D_{01}$ are located in material 1 at z_{01} $(t_1=t_{01})$ and at infinity $(t_1=1)$, respectively. Dislocations, stresses and displacements in z_j planes are related to those in the original z-plane as

$$D_{X1} = -d_{x1} \quad D_{Y1} = d_{y1} \quad X_{X2} = d_{x2} \quad D_{Y2} = d_{y2}$$

$$U_1 = u_1 \quad V_1 = -v_1 \quad U_2 = u_2 \quad V_2 = v_2$$

$$\sigma_{X1} = \sigma_{x1} \quad \sigma_{Y1} = \sigma_{y1} \quad \tau_{XY1} = -\tau_{xy1}$$

$$\sigma_{X2} = \sigma_{x2} \quad \sigma_{Y2} = \sigma_{y2} \quad \tau_{XY2} = \tau_{xy2} \tag{1}$$

A mapping function that maps a half-plane with semielliptic notch onto a unit circle is given by

$$z_j = -ib\frac{\sqrt{1+t_j^2}}{1-t_j} - ia\frac{1+t_j}{1-t_j} + \text{constant}. \tag{2}$$

a,b are the dimensions of ellipse along x_j and y_j axes, respectively. A rational mapping function can be obtained from (2) as

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, August 5, 2003; final revision, November 24, 2003. Associate Editor: Z. Suo. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

$$z_j = \omega(t_j) = \frac{E_0}{1-t_j} + \sum_{k=1}^{N} \frac{E_k}{\zeta_k - t_j} + E_c, \quad j = 1,2. \tag{3}$$

E_0, E_k, E_c, and ζ_k are complex constants and $N=28$ in this analysis. The procedure for calculating the constants in (3) is explained in detail by Hasebe and Inohara [14].

The given problem can be split into two parts: problem (a) the solution of point dislocations and a free boundary in a half-plane with a semielliptical notch (see Fig. 1(*b*)) is solved and, problem (b) the continuity of stresses and displacements across the bonded part of half-planes is satisfied. Accordingly, complex potentials for the upper half-plane (material 1) can be written in the mapped plane as

$$\phi_1(t_1) = \phi_{1A}(t_1) + \phi_{d1}(t_1)$$

$$\psi_1(t_1) = \psi_{1A}(t_1) + \psi_{d1}(t_1) \tag{4}$$

where the suffix "*d*1" corresponds to problem (a) and suffix "1*A*" corresponds to problem (b).

2.1 Solution for Problem (a). The solution for the problem of a half-plane with a semielliptical notch subjected to a point dislocation at $t_1=t_{01}$ and $t_1=1$ ($z_1=\infty$) has been obtained by Hasebe et al. [15]. The solution is presented here for the sake of convenience,

$$\begin{aligned}\phi_{d1}(t_1) = &-\frac{D_1}{2\pi}\log(t_1 - t_{01}) + \frac{D_1}{2\pi}\log(t_1 - 1/\overline{t_{01}}) \\ &+ \frac{\overline{D_1}}{2\pi}\frac{\omega(t_{01}) - \omega(1/\overline{t_{01}})}{\overline{\omega'(t_{01})}}\frac{(1/\overline{t_{01}})^2}{t_1 - 1/\overline{t_{01}}} + \frac{\overline{D_1}}{2\pi}\sum_{k=1}^{N} \\ &\times\left(\frac{1}{\overline{\zeta_k'} - 1} - \frac{1}{\overline{\zeta_k'} - \overline{t_{01}}}\right)\frac{B_k}{t_1 - \zeta_k} + \sum_{k=1}^{N}\frac{\overline{A_{d1k}}B_k}{t_1 - \zeta_k},\end{aligned} \tag{5}$$

and

$$\psi_{d1}(t_1) = -\overline{\phi_{d1}}(1/t_1) - \frac{\overline{\omega}(1/t_1)}{\omega'(t_1)}\phi_{d1}'(t_1), \tag{6}$$

where $B_k \equiv E_k/\overline{\omega'(\zeta_k')}$, $\zeta_k' = 1/\overline{\zeta_k}$, and A_{d1k} can be evaluated as explained in [15].

2.2 Solution for Problem (b). Solution to problem (b) entails satisfying boundary conditions along the bonded and unbonded parts of the interface. Since the elliptical hole is traction free, analytic continuation leads to

$$\psi_j(t_j) = -\overline{\phi_j}(1/t_j) - \frac{\overline{\omega}(1/t_j)}{\omega'(t_j)}\phi_j'(t_j). \tag{7}$$

The boundary condition on the traction free elliptical hole becomes

$$\phi_j^+(\sigma) - \phi_j^-(\sigma) = 0 \quad \text{on} \quad L_j \tag{8}$$

where σ represents the point on the boundary; superscripts + and − denote the limiting values of the functions on the circumference when approached from inside and outside the unit circle, respectively.

Since tractions and displacements are continuous across the bonded part of the interface,

$$\phi_1^+(\sigma) - \phi_1^-(\sigma) = \overline{\phi_2^+(1/\overline{\sigma})} - \overline{\phi_2^-(1/\overline{\sigma})} \quad \text{on} \quad M, \tag{9}$$

$$\frac{1}{\mu_1}[\kappa_1\phi_1^+(\sigma) + \phi_1^-(\sigma)] = \frac{1}{\mu_2}[\kappa_2\overline{\phi_2^+(1/\overline{\sigma})} + \overline{\phi_2^-(1/\overline{\sigma})}] \quad \text{on} \quad M. \tag{10}$$

Writing

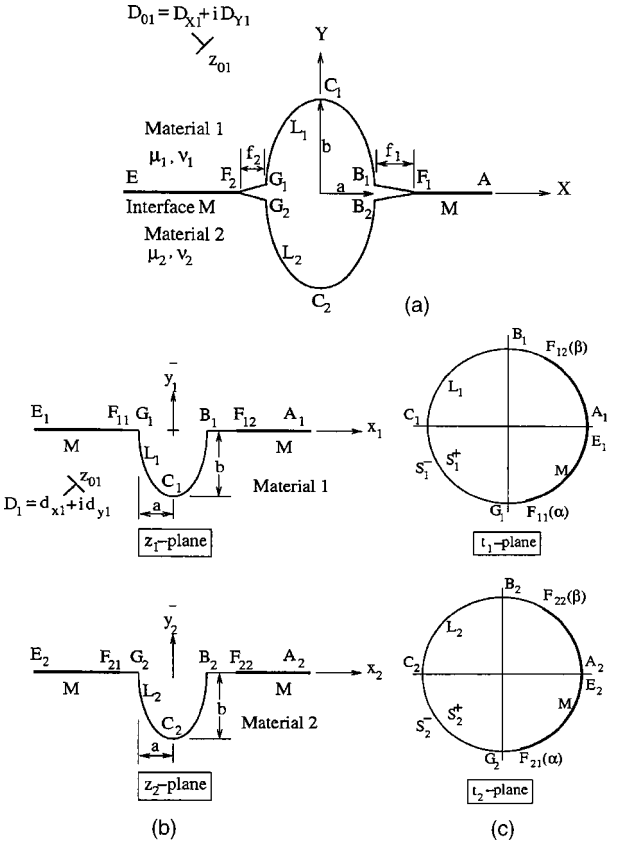


Fig. 1 Elliptical hole at a bimaterial interface

$$\phi_1(t_1) + \overline{\phi_2(1/\overline{t_1})} = \Theta(t_1), \tag{11}$$

Eq. (9) can be rewritten as

$$\Theta_1^+(\sigma) - \Theta_1^-(\sigma) = 0. \tag{12}$$

General solution of Eq. (12) is an arbitrary rational function [16]

$$\Theta(t_1) = \theta_1(t_1). \tag{13}$$

Since $\theta_1^+(\sigma) = \theta_1^-(\sigma) \equiv \theta_1(\sigma)$ on the boundary,

$$\overline{\phi_2^+(1/\overline{\sigma})} = -\phi_1^-(\sigma) + \theta_1(\sigma) \quad \text{and} \quad \overline{\phi_2^-(1/\overline{\sigma})} = -\phi_1^+(\sigma) + \theta_1(\sigma). \tag{14}$$

Substituting Eq. (14) into Eq. (10), the boundary condition on M can be written as

$$\phi_1^+(\sigma) + A_1\phi_1^-(\sigma) = B_1\theta_1(\sigma), \tag{15}$$

$$A_1 = \frac{\kappa_2\mu_1 + \mu_2}{\kappa_1\mu_2 + \mu_1} = \frac{1-\beta_D}{1+\beta_D}, \quad B_1 = \frac{\mu_1(\kappa_2+1)}{\kappa_1\mu_2 + \mu_1} = \frac{1-\alpha_D}{1+\beta_D}$$

where μ_1 and μ_2 denote the shear moduli of material 1 and 2, respectively; α_D and β_D are the Dundurs parameters given by

$$\alpha_D = \frac{(\kappa_1+1)\Gamma - (\kappa_2+1)}{(\kappa_1+1)\Gamma + (\kappa_2+1)}, \quad \beta_D = \frac{(\kappa_1-1)\Gamma - (\kappa_2-1)}{(\kappa_1+1)\Gamma + (\kappa_2+1)} = \Gamma\frac{\mu_2}{\mu_1}.$$

By repeating the above procedure for $\phi_2(t_2)$, the following boundary condition on M can be formulated

$$\phi_2^+(\sigma) + A_2\phi_2^-(\sigma) = B_2\theta_2(\sigma), \tag{16}$$

$$A_2 = \frac{\kappa_1\mu_2 + \mu_1}{\kappa_2\mu_1 + \mu_2} = \frac{1+\beta_D}{1-\beta_D}, \quad B_2 = \frac{\mu_2(\kappa_1+1)}{\kappa_2\mu_1+\mu_2} = \frac{1+\alpha_D}{1-\beta_D}.$$

The problem of obtaining the potentials $\phi_1(t_1)$ and $\phi_2(t_2)$ is thus reduced to finding solutions to the Riemann–Hilbert Eqs. (8), (15), and (16).

In the dislocation problem, $\phi_1(t_1)$ is expressed by Eq. (4). Substituting Eq. (4) in Eqs. (8) and (15) yields

$$\phi_{1A}^{+}(\sigma) - \phi_{1A}^{-}(\sigma) = 0 \quad \text{on} \quad L_1, \tag{17}$$

$$\phi_{1A}^{+}(\sigma) + A_1\phi_{1A}^{-}(\sigma) = B_1\theta_{1A}(\sigma) + C_1\phi_{d1}(\sigma) \quad \text{on} \quad M, \tag{18}$$

where

$$C_1 = \mu_2(\kappa_1+1)/(\kappa_1\mu_2+\mu_1) = (1+\alpha_D)/(1+\beta_D)$$

and $\theta_{1A}(t_1) = \phi_{1A}(t_1) + \overline{\phi_2(1/\bar{t}_1)}$ is a rational function. Similarly,

$$\phi_2^{+}(\sigma) - \phi_2^{-}(\sigma) = 0 \quad \text{on } L_2 \tag{19}$$

$$\phi_2^{+}(\sigma) + A_2\phi_2^{-}(\sigma) = B_2\theta_{2A}(\sigma) + C_2\overline{\phi_{d1}(1/\bar{\sigma})} \quad \text{on } M \tag{20}$$

where

$$C_2 = -\mu_2(\kappa_1+1)/(\kappa_2\mu_1+\mu_2) = -(1+\alpha_D)/(1-\beta_D)$$

and $\theta_{2A}(t_2) = \phi_2(t_2) + \overline{\phi_{1A}(1/\bar{t}_2)}$ is a rational function. The general solution to Eqs. (17) and (18) can be written as [16]

$$\phi_{1A}(t_1) = \frac{B_1\chi_1(t_1)}{2\pi i}\int_M \frac{\theta_{1A}(\sigma)d\sigma}{\chi_1^{+}(\sigma)(\sigma-t_1)} + \frac{C_1\chi_1(t_1)}{2\pi i}\int_M \frac{\phi_{d1}(\sigma)d\sigma}{\chi_1^{+}(\sigma)(\sigma-t_1)} + \chi_1(t_1)P_1(t_1), \tag{21}$$

where $P_1(t_1)$ is a rational function to be determined, $\chi_1(t_1) = (t_1-\alpha)^{m_1}(t_1-\beta)^{1-m_1}$, and $m_1 = 0.5 + i(\log A_1)/2\pi$. The behavior of $\chi_1(t_1)$ is given by

$$\chi_1^{+}(\sigma) = -A_1\chi_1^{-}(\sigma) \quad \text{on} \quad M,$$

$$\chi_1^{+}(\sigma) = \chi_1^{-}(\sigma) \quad \text{on} \quad L_1.$$

Similarly, the general solution to Eqs. (19) and (20) can be written as

$$\phi_2(t_2) = \frac{B_2\chi_2(t_2)}{2\pi i}\int_M \frac{\theta_{2A}(\sigma)d\sigma}{\chi_2^{+}(\sigma)(\sigma-t_2)} + \frac{C_2\chi_2(t_2)}{2\pi i}\int_M \frac{\overline{\phi_{d1}(1/\bar{\sigma})}d\sigma}{\chi_2^{+}(\sigma)(\sigma-t_2)} + \chi_2(t_2)P_2(t_2), \tag{22}$$

where

$$\chi_2(t_2) = (t_2-\alpha)^{m_2}(t_2-\beta)^{1-m_2}$$

, and $m_2 = 0.5 + i(\log A_2)/2\pi$. The second term in Eqs. (21) and (22) is evaluated using the residue theorem as

$$\frac{C_1\chi_1(t_1)}{2\pi i}\int_M \frac{\phi_{d1}(\sigma)}{\chi_1^{+}(\sigma)(\sigma-t_1)}d\sigma = \frac{C_1\chi_1(t_1)}{2\pi i(1+A_1)}\oint_M \frac{\phi_{d1}(\sigma)}{\chi_1(\sigma)(\sigma-t_1)}d\sigma = \frac{C_1}{2\pi(1+A_1)}\left[D_1\left\{\log(t_1 - 1/\bar{t}_{01}) - \log(t_1-t_{01}) + \chi_1(t_1)\int_{1/\bar{t}_{01}}^{t_{01}}\frac{d\sigma}{\chi_1(\sigma)(\sigma-t_1)}\right\} + \overline{D}_1\frac{\omega(t_{01}) - \omega(1/\bar{t}_{01})}{\overline{\omega'(t_{01})}}\left[1 - \frac{\chi_1(t_1)}{\chi_1(1/\bar{t}_{01})}\right]\frac{(1/\bar{t}_{01})^2}{t_1 - 1/\bar{t}_{01}} + \overline{D}_1\sum_{k=1}^{N}\left(\frac{1}{\bar{\zeta}_k' - 1} - \frac{1}{\bar{\zeta}_k' - \bar{t}_{01}}\right)\left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta_k)}\right]\frac{B_k}{t_1-\zeta_k} + 2\pi\sum_{k=1}^{N}\frac{\overline{A_{d1k}}B_k}{t_1-\zeta_k}\left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta_k)}\right]\right], \tag{23}$$

$$\frac{C_2\chi_2(t_2)}{2\pi i}\int_M \frac{\overline{\phi_{d1}(1/\bar{\sigma})}}{\chi_2^{+}(\sigma)(\sigma-t_2)}d\sigma = \frac{C_2\chi_2(t_2)}{2\pi i(1+A_2)}\oint_M \frac{\overline{\phi_{d1}(1/\bar{\sigma})}}{\chi_2(\sigma)(\sigma-t_2)}d\sigma = \frac{C_2}{2\pi(1+A_2)}\left[\overline{D}_1\left\{\log(t_2-t_{01}) - \log(t_2 - 1/\bar{t}_{01}) + \chi_2(t_2)\int_{t_{01}}^{1/\bar{t}_{01}}\frac{d\sigma}{\chi_2(\sigma)(\sigma-t_2)}\right\} - D_1\frac{\overline{\omega(t_{01})} - \overline{\omega(1/\bar{t}_{01})}}{\omega'(t_{01})}\left[1 - \frac{\chi_2(t_2)}{\chi_2(t_{01})}\right]\frac{1}{t_2-t_{01}} - D_1\sum_{k=1}^{N}\left(\frac{1}{\zeta_k'-1} - \frac{1}{\zeta_k' - t_{01}}\right) \times\left[1 - \frac{\chi_2(t_2)}{\chi_2(\zeta_k')}\right]\frac{\bar{B}_k\zeta_k'^2}{t_2-\zeta_k'} - 2\pi\sum_{k=1}^{N}\frac{A_{d1k}\bar{B}_k\zeta_k'^2}{t_2-\zeta_k'}\left[1 - \frac{\chi_2(t_2)}{\chi_2(\zeta_k')}\right]\right]. \tag{24}$$

The contour integrals of Eqs. (23) and (24) are carried out as outlined in Hasebe et al. [17]. The first derivatives of the integral terms in Eqs. (23) and (24) are expressed by terms without integrals [17].

The rational function $\theta_{1A}(t_1)$ can be expressed as a sum of irregular terms inside and outside the unit circle as

$$\theta_{1A}(t_1) = \sum_n \frac{a_{1n}}{\xi_{1n} - t_1} + \sum_n \frac{b_{1n}}{\eta_{1n} - t_1}, \tag{25}$$

where $|\xi_{1n}| > 1$, $|\eta_{1n}| < 1$, and a_{1n} and b_{1n} are complex constants to be determined.

Substituting (25) into (21) the first term is evaluated using residue theorem as

$$\frac{B_1\chi_1(t_1)}{2\pi i}\int_M \frac{\theta_{1A}(\sigma)d\sigma}{\chi_1^{+}(\sigma)(\sigma-t_1)} = \frac{B_1\chi_1(t_1)}{2\pi i(1+A_1)}\oint_M \frac{\theta_{1A}(\sigma)d\sigma}{\chi_1(\sigma)(\sigma-t_1)} = \frac{B_1}{1+A_1}\left\{\sum_n\left[1 - \frac{\chi_1(t_1)}{\chi_1(\xi_{1n})}\right]\frac{a_{1n}}{\xi_{1n}-t_1} + \sum_n\left[1 - \frac{\chi_1(t_1)}{\chi_1(\eta_{1n})}\right]\frac{b_{1n}}{\eta_{1n}-t_1}\right\}. \tag{26}$$

$P_1(t_1)$ is obtained from the analytical conditions on $\psi_{1A}(t_1)$. Substituting Eq. (4) into Eq. (7),

$$\psi_{1A}(t_1) = -\overline{\phi_{1A}}(1/t_1) - \frac{\overline{\omega}(1/t_1)}{\omega'(t_1)}\phi'_{1A}(t_1) - \psi_{d1}(t_1) - \overline{\phi_{d1}}(1/t_1) - \frac{\overline{\omega}(1/t_1)}{\omega'(t_1)}\phi'_{d1}(t_1). \quad (27)$$

The first term in Eq. (27) can be written from Eq. (21) as

$$\overline{\phi_{1A}}(1/t_1) = \overline{\chi_1}(1/t_1)\overline{P_1}(1/t_1) + \text{terms regular in } S^+.$$

The second term in Eq. (27) can expressed as

$$\frac{\overline{\omega}(1/t_1)}{\omega'(t_1)}\phi'_{1A}(t_1) = -\sum_{k=1}^{N}\frac{A_{1k}\overline{B}_k\zeta'^2_k}{\zeta'_k - t_1} + \text{terms regular in } S^+$$

where $A_{1k} \equiv \phi'_{1A}(\zeta'_k)$. From (6), the last three terms on the right-hand side of Eq. (27) disappear. Thus, $\psi_{1A}(t_1)$ is written as

$$\psi_{1A}(t_1) = -\overline{\chi_1}(1/t_1)\overline{P_1}(1/t_1) + \sum_{k=1}^{N}\frac{A_{1k}\overline{B}_k\zeta'^2_k}{\zeta'_k - t_1} + \text{terms regular in } S^+.$$

Since $\psi_{1A}(t_1)$ is regular in S^+, the irregular part should be cancelled out. Expanding the first term in Laurent series and equating terms with pole at $t_1 = \zeta'_k$ to zero, we get

$$P_1(t_1) = -\sum_{k=1}^{N}\frac{\overline{A_{1k}}B_k}{\chi_1(\zeta_k)(\zeta_k - t_1)}. \quad (28)$$

Similarly, $P_2(t_2)$ can be obtained as

$$P_2(t_2) = -\sum_{k=1}^{N}\frac{\overline{A_{2k}}B_k}{\chi_2(\zeta_k)(\zeta_k - t_2)}. \quad (29)$$

In order to evaluate the function $\Theta_1(t_1)$, the following relations are used:

$$\frac{\overline{\chi_2(1/\bar{t}_1)}}{\overline{\chi_2(1)}} = \frac{\chi_1(t_1)}{t_1\chi_1(1)}, \quad \frac{\overline{\chi_2(1/\bar{t}_1)}}{\overline{\chi_2(\xi_{2n})}} = \frac{\xi'_{2n}\chi_1(t_1)}{t_1\chi_1(\xi'_{2n})}, \quad \frac{\overline{\chi_2(1/\bar{t}_1)}}{\overline{\chi_2(\eta_{2n})}} = \frac{\eta'_{2n}\chi_1(t_1)}{t_1\chi_1(\eta'_{2n})}$$

where $\xi'_{2n} \equiv 1/\overline{\xi_{2n}}$, $\eta'_{2n} \equiv 1/\overline{\eta_{2n}}$, $A_2 = 1/A_1$, and $m_2 = \overline{m_1}$.

Using Eqs. (21) and (22)

$$\theta_{1A}(t_1) = \phi_{1A}(t_1) + \overline{\phi_2(1/\bar{t}_1)} = \frac{B_1}{1+A_1}\left\{\sum_n\left[1 - \frac{\chi_1(t_1)}{\chi_1(\xi_{1n})}\right]\frac{a_{1n}}{\xi_{1n} - t_1} + \sum_n\left[1 - \frac{\chi_1(t_1)}{\chi_1(\eta_{1n})}\right]\frac{b_{1n}}{\eta_{1n} - t_1}\right\} - \frac{B_2}{1+A_2}\left\{\sum_n\left[1 - \frac{\chi_1(t_1)}{\chi_1(\eta'_{2n})}\right]\frac{\overline{b_{2n}}\eta'^2_{2n}}{\eta'_{2n} - t_1} + \sum_n\left[1 - \frac{\chi_1(t_1)}{\chi_1(\xi'_{2n})}\right]\frac{\overline{a_{2n}}\xi'^2_{2n}}{\xi'_{2n} - t_1}\right\} - \sum_{k=1}^{N}\frac{\chi_1(t_1)\overline{A_{1k}}B_k}{\chi_1(\zeta_k)(\zeta_k - t_1)} + \sum_{k=1}^{N}\frac{\chi_1(t_1)A_{2k}\overline{B}_k\zeta'^2_k}{\chi_1(\zeta'_k)(\zeta'_k - t_1)}. \quad (30)$$

Since Eqs. (25) and (30) are identical, their poles must be the same and the coefficients in the two terms, which either include or exclude the Plemelj functions, must also be equal at each pole in S_1^+ and S_1^-. Therefore, the coefficients are evaluated as

$$\xi_{1k} = \eta'_{2k} = \zeta_k, \quad a_{1k} = -\overline{A_{1k}}B_k, \quad a_{2k} = -\overline{A_{2k}}B_k,$$

$$\eta_{1k} = \xi'_{2k} = \zeta'_k, \quad b_{1k} = A_{2k}\overline{B}_k\zeta'^2_k, \quad b_{2k} = A_{1k}\overline{B}_k\zeta'^2_k.$$

Thus,

$$\theta_{1A}(t_1) = -\sum_{k=1}^{N}\frac{\overline{A_{1k}}B_k}{\zeta_k - t_1} + \sum_{k=1}^{N}\frac{A_{2k}\overline{B}_k\zeta'^2_k}{\zeta'_k - t_1}. \quad (31)$$

Finally, $\phi_{1A}(t_1)$ is obtained as

$$\phi_{1A}(t_1) = -\frac{1}{1+\delta_1}\sum_{k=1}^{N}\left[1 + \delta_1\frac{\chi_1(t_1)}{\chi_1(\zeta_k)}\right]\frac{\overline{A_{1k}}B_k}{\zeta_k - t_1} + \frac{1}{1+\delta_1}\sum_{k=1}^{N}\times\left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta'_k)}\right]\frac{A_{2k}\overline{B}_k\zeta'^2_k}{\zeta'_k - t_1} - \frac{\delta_1}{2\pi(1+\delta_1)}\left[D_1\left\{\log(t_1 - 1/\overline{t_{01}}) - \log(t_1 - t_{01}) + \chi_1(t_1)\int_{1/\overline{t_{01}}}^{t_{01}}\frac{d\sigma}{\chi_1(\sigma)(\sigma - t_1)}\right\} + \overline{D_1}\frac{\omega(t_{01}) - \omega(1/\overline{t_{01}})}{\overline{\omega'(t_{01})}}\left[1 - \frac{\chi_1(t_1)}{\chi_1(1/\overline{t_{01}})}\right]\frac{(1/\overline{t_{01}})^2}{t_1 - 1/\overline{t_{01}}} + \overline{D_1}\sum_{k=1}^{N}\left(\frac{1}{\overline{\zeta'_k} - 1} - \frac{1}{\overline{\zeta'_k} - \overline{t_{01}}}\right)\left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta_k)}\right]\frac{B_k}{t_1 - \zeta_k} + 2\pi\sum_{k=1}^{N}\frac{\overline{A_{d1k}}B_k}{t_1 - \zeta_k}\left[1 - \frac{\chi_1(t_1)}{\chi_1(\zeta_k)}\right]\right], \quad (32)$$

where $\delta_1 = \mu_2(\kappa_1+1)/(\mu_1(\kappa_2+1)) = (1+\alpha_D)/(1-\alpha_D)$.

The complex potentials in the final form can be written as

$$\phi_1(t_1) = \phi_{1A}(t_1) + \phi_{d1}(t_1), \quad (33)$$

$$\phi_2(t_2) = -\frac{1}{1+\delta_2}\sum_{k=1}^{N}\left[1 + \delta_2\frac{\chi_2(t_2)}{\chi_2(\zeta_k)}\right]\frac{\overline{A_{2k}}B_k}{\zeta_k - t_2} + \frac{1}{1+\delta_2}\sum_{k=1}^{N}\times\left[1 - \frac{\chi_2(t_2)}{\chi_2(\zeta'_k)}\right]\frac{A_{1k}\overline{B}_k\zeta'^2_k}{\zeta'_k - t_2} + \frac{1}{2\pi(1+\delta_2)}\left[\overline{D_1}\left\{\log(t_2 - t_{01}) - \log(t_2 - 1/\overline{t_{01}}) + \chi_2(t_2)\int_{t_{01}}^{1/\overline{t_{01}}}\frac{d\sigma}{\chi_2(\sigma)(\sigma - t_2)}\right\} - D_1\frac{\overline{\omega(t_{01})} - \overline{\omega(t/\bar{t}_{01})}}{\omega'(t_{01})}\left[1 - \frac{\chi_2(t_2)}{\chi_2(t_{01})}\right]\frac{1}{t_2 - t_{01}} - D_1\sum_{k=1}^{N}\left(\frac{1}{\zeta'_k - 1} - \frac{1}{\zeta'_k - t_{01}}\right)\left[1 - \frac{\chi_2(t_2)}{\chi_2(\zeta'_k)}\right]\frac{\overline{B}_k\zeta'^2_k}{t_2 - \zeta'_k} - 2\pi\sum_{k=1}^{N}\frac{A_{d1k}\overline{B}_k\zeta'^2_k}{t_2 - \zeta'_k}\left[1 - \frac{\chi_2(t_2)}{\chi_2(\zeta'_k)}\right]\right], \quad (34)$$

where $\delta_1 = 1/\delta_2$, and $A_{1k} \equiv \phi'_{1A}(\zeta'_k)$, $A_{2k} \equiv \phi'_2(\zeta'_k)$. It should be noted that the first derivatives of $\phi_1(t_1)$ and $\phi_2(t_2)$ do not involve integral terms present in Eqs. (33) and (34). Thus numerical integration is not necessary to calculate A_{1k} and A_{2k} as well as stress components. However, numerical integration of the integral terms in Eqs. (33) and (34) is necessary to compute displacement components which involve both the complex potentials and their derivatives. The complex constants A_{1k} and A_{2k} $(k = 1, 2, \ldots, N)$ are determined by solving the $4N$ simultaneous linear equations corresponding to the real and imaginary parts of A_{1k} and A_{2k} obtained from their definitions. The complex potentials $\psi_j(t_j)$ $(j = 1, 2)$ are given by Eq. (7). It should be noted that $\phi_2(t_2)$ is regular at $t_2 = t_{01}$ since the expression involving logarithmic terms of Eq. (34) cancel out with the term under the integral sign. Further, the solution to the problem of a point dislocation in material 2 can similarly be obtained by replacing the subscripts "1" and "2" in Eqs. (33) and (34).

Conclusions

The solution of a point dislocation interacting with an elliptical hole has been derived analytically by applying complex variable method. The key point to note in the derivation of Green's func-

tion of the bimaterial problem is to utilize the Green's function $\phi_{d1}(\zeta)$ and $\psi_{d1}(\zeta)$ of the half-plane with a semielliptical notch. The mapping function for a half-plane with a semielliptical notch that is mapped onto a unit circle by means of a rational mapping function is used. Therefore, by changing the mapping function any other hole problem can be analyzed, for example, half-plane with a triangular notch solved earlier by Hasebe and Iida [18]. Any debonding length on the interface can be taken care of by changing variables α and β in the Plemlj function. The fundamental solution can be used as Green's function to simulate cracks by means of distributed dislocation technique that forms the second part of this paper. The Green's function for two equal but opposite point forces P_x+iP_y and $-(P_x+iP_y)$ acting at $t_1=t_{01}$ and $t_1=t_{02}$ in the mapped plane can be obtained by superposing the solutions for two equal and opposite point dislocations at $t_1=t_{01}$ and $t_1=t_{02}$, and changing $D=D_x+iD_y$ to $P=(P_x+iP_y)/(1+\kappa)$ and $\overline{D}=D_x-iD_y$ to $-\kappa\overline{P}=-\kappa(P_x-iP_y)/(1+\kappa)$ in the expressions for point dislocations.

References

[1] Williams, M. L., 1959, "The Stress Around a Fault or Crack in Dissimilar Media," Bull. Seismol. Soc. Am., **49**, pp. 199–204.

[2] Rice, J. R., and Sih, G. C., 1965, "Plane Problems in Dissimilar Media," J. Appl. Mech., **32**, pp. 418–423.

[3] England, A. H., 1965, "A Crack Between Dissimilar Media," J. Appl. Mech., **32**, pp. 400–402.

[4] Erdogan, F., 1965, "Stress Distribution in Bonded Dissimilar Materials with Cracks," J. Appl. Mech., **32**, pp. 403–410.

[5] Rice, J. R., 1988, "Elastic Fracture Mechanics Concepts for Interfacial Cracks," J. Appl. Mech., **55**, pp. 98–103.

[6] Comninou, M., 1977, "The Interface Crack," J. Appl. Mech., **44**, pp. 631–636.

[7] Comninou, M., 1990, "An Overview of Interface Cracks," Eng. Fract. Mech., **37**, pp. 197–208.

[8] Erdogan, F., 1970, "Fracture Problems in Composite Materials," Eng. Fract. Mech., **4**, pp. 811–840.

[9] Hutchinson, J. W., Mear, M. E., and Rice, J. R., 1987, "Crack Paralleling an Interface Between Dissimilar Elastic Materials," J. Appl. Mech., **54**, pp. 828–832.

[10] Murakami, Y., et al., 1992, *Stress Intensity Factors Handbook*, Elsevier Science, New York.

[11] Oda, K., Noda, N., and Arita, S., 2003, "Stress Intensity Factors for Interaction between Interface Crack and Internal Crack and for Kinked Interface Crack in Bonded Semi-Infinite Planes," Key Eng. Mater., **243–244**, pp. 375–380.

[12] Hasebe, N., Okumura, M., and Nakamura, T., 1992, "Bonded Bi-material Half-Planes With Semi-elliptical Notch Under Tension Along the Interface," J. Appl. Mech., **59**, pp. 77–83.

[13] Okumura, M., Hasebe, N., and Nakamura, T., 1995, "Bi-material Plane with Elliptic Hole under Uniform Tension Normal to the Interface," Int. J. Fract., **71**, pp. 293–310.

[14] Hasebe, N., and Inohara, S., 1980, "Stress Analysis of a Semi-infinite Plate with an Oblique Edge Crack," Ingenieurs, **49**, pp. 51–62.

[15] Hasebe, N., Qian, J., and Chen, Y. Z., 1996, "Fundamental Solutions for Half Plane with an Oblique Edge Crack," Eng. Anal. Boundary Elem., **17**, pp. 263–267.

[16] Muskhelishvili, N. I., 1963, *Some Basic Problems of the Mathematical Theory of Elasticity*, 4th ed., Noordhoff, The Netherlands.

[17] Hasebe, N., Irikura, H., Nakamura, T., 1991, "A Solution of the Mixed Boundary Value Problem for an Infinite Plate with a Hole under Uniform Heat Flux," J. Appl. Mech., **113**, pp. 996–1000.

[18] Hasebe, N., and Iida, J., 1978, "A Crack Originating from a Triangular Notch on a Rim of a Semi-infinite Plate," Eng. Fract. Mech., **10**, pp. 773–782.

P. B. N. Prasad
Postdoctoral Researcher

Norio Hasebe
Professor

X. F. Wang
Research Associate

Department of Civil Engineering,
Nagoya Institute of Technology,
Gokiso-Cho, Showa-Ku,
Nagoya 466, Japan

Interaction Between Interfacial Cavity/Crack and Internal Crack—Part II: Simulation

This paper discusses the interaction of an interfacial cavity/crack with an internal crack in a bimaterial plane under uniform loading at infinity. The point dislocation solution is used to simulate internal crack by using the distributed dislocation technique. The resulting singular integral equation is solved numerically and the stress intensity factor variations are plotted for some cases of internal crack interacting with interfacial cavity/crack. [DOI: 10.1115/1.1876433]

1 Introduction

Multiple crack interaction studies constitute a major research in understanding the structural integrity of components under loading. Many cracked structures show that multiple cracks are a rule rather than an exception. Several works on crack interactions have been carried out analytically and numerically to estimate the severity of cracks and to assess the overall integrity of components. Works carried out in this direction include cracks of arbitrary lengths in homogeneous media interacting with each other; hole edge cracks interacting with internal cracks; internal cracks interacting with interfacial cracks, etc. The mathematical difficulties while solving crack interaction problems make numerical techniques inevitable. Several problems of practical interest have been solved numerically and compiled in Murakami et al. [1].

In this paper the problem of an internal crack interacting with an interfacial cavity/crack is examined. The point dislocation solution derived earlier [2] is used in conjunction with the distributed dislocation technique to simulate the internal crack. This method has been used successfully to study several crack interaction problems that are not amenable to analytical solutions [1,3]. In the simulations, two internal crack orientations and corresponding far-field loading directions are considered (see Fig. 1) (a) internal crack parallel to and loading normal to the interface (b) internal crack normal to and loading parallel to the interface. Crack interaction effects will be studied by examining the variations in stress intensity factors at the crack tips by varying the distance between the internal crack and the interfacial hole. The numerical procedure to obtain stress intensity factors will be explained in Sec. 2. Stress intensity factors variations for some cases of internal crack and loading orientations will be studied in Sec. 3.

2 Analysis

The solution of the problem involves simulating internal crack by distributing a point dislocation along the putative crack length. Earlier, the solution to the problem of a point dislocation in a bimaterial plane with an interfacial cavity/crack has been obtained. This solution will be used to simulate the internal crack. We assume that the internal crack exists in material 1. The given problem can be solved in two stages. In the first stage, tractions along the putative internal crack length induced by far-field loadings are obtained. In the second stage, tractions along the internal crack length due to unit point dislocations acting in normal and tangential directions to the internal crack are multiplied by unknown dislocation densities and equated to the negative of tractions obtained in the first stage. This sets up the singular equation for the given problem the solution of which gives the unknown dislocation densities. Once the unknown densities are known, stress intensity factors at crack tips can be evaluated.

Dislocation density is defined by [4],

$$h_j(t) = \frac{2\mu_1}{1+\kappa_1}\frac{d}{dt}(u_j^+(t) - u_j^-(t)), \quad |t| < c, \tag{1}$$

where c is the semicrack length, and $j=1,2$ correspond to x and y displacements, respectively. In the above expression μ_1 is shear modulus and κ_1 is given by $(3-4\nu_1)$ under plane strain and $(3-\nu_1)/(1+\nu_1)$ under plane stress conditions in which ν_1 is the Poisson's ratio of material 1. The bracketed term in Eq. (1) denotes the jump in displacement across the crack face.

The condition of single valued displacement on the crack surface is given by

$$\int_{-c}^{c} h_j(t)dt = 0. \tag{2}$$

Singular integral equation can be setup by noting that the stresses induced due to given loading should be zero on the putative crack site as the internal crack is traction free. Thus, negating the tractions induced due to far-field loading at the crack site by distributing the dislocations leads to the following singular integral equation

$$N_C(s) + iT_C(s) = \int_{-c}^{c} \{h_n(t)[\sigma_{nn}(t,s) + i\sigma_{n\tau}(t,s)] + h_\tau(t)[\sigma_{\tau n}(t,s) + i\sigma_{\tau\tau}(t,s)]\}dt \tag{3}$$

where $\sigma_{jk}(t,s)$ represents the traction component in the k direction at point s induced by dislocation density $h_j(t)$ at point t in the j direction. The left-hand side, $N_C(s)$ and $T_C(s)$, denotes negative of the traction induced due to far-field loading at the crack site. The condition of single valued displacement of Eq. (2) can be rewritten as

$$\int_{-c}^{c} \{h_n(t) + ih_\tau(t)\}dt = 0. \tag{4}$$

The explicit forms of Eqs. (3) and (4) are difficult to present and are therefore solved numerically using the Gauss–Chebyshev integration method as explained in [4,5].

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, August 5, 2003; final revision, November 24, 2003. Associate Editor: Z. Suo. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

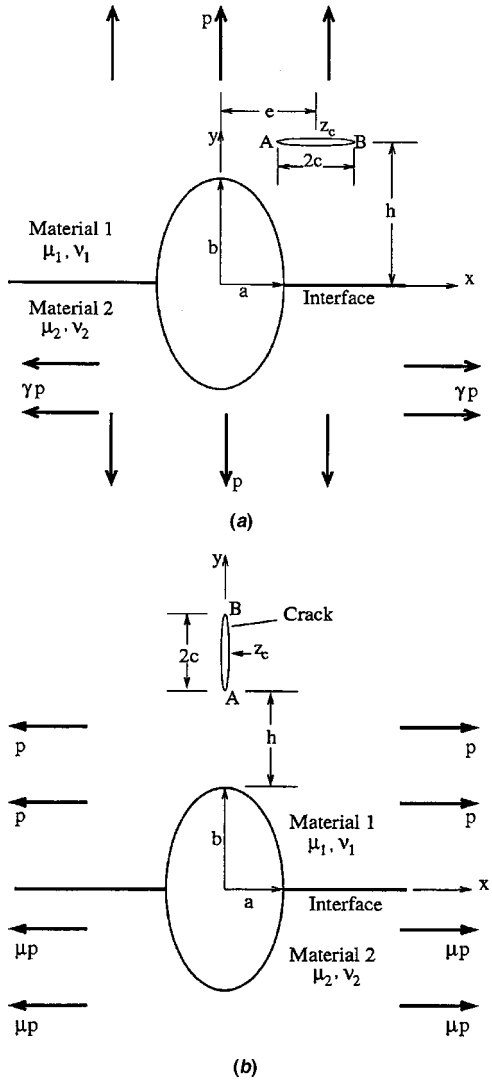


Fig. 1 Internal crack interacting with interfacial cavity

The stress components on the interface near the debonding tip D (see Fig. 10 below) can be expressed as [6]

$$\sigma_{Yj}^{M} = \frac{\cosh(\pi\varepsilon_j)}{\sqrt{2\pi r}}\sqrt{K^{(j)}\overline{K^{(j)}}}\cos(\Theta_0^{(j)} - \varepsilon_j \ln r) + O(r^0),$$

$$\sigma_{Xj}^{M} = \frac{\cosh(\pi\varepsilon_j) + 2\sinh(\pi\varepsilon_j)}{\sqrt{2\pi r}}\sqrt{K^{(j)}\overline{K^{(j)}}}\cos(\Theta_0^{(j)} - \varepsilon_j \ln r) + O(r^0),$$

$$\tau_{XYj}^{M} = \pm\frac{\cosh(\pi\varepsilon_j)}{\sqrt{2\pi r}}\sqrt{K^{(j)}\overline{K^{(j)}}}\sin(\Theta_0^{(j)} - \varepsilon_j \ln r) + O(r^0),$$

where the signs + and − are for material 1 and 2; respectively; $K^{(j)} = K_{\mathrm{I}}^{(j)} + iK_{\mathrm{II}}^{(j)}$ ($j=1,2$ correspond to materials 1 and 2) is the complex stress intensity factor of the interfacial crack, $\Theta_0^{(j)}$ is an argument of $K^{(j)}$, r is the distance from the debonded crack tip on the interface, and ε_j is the imaginary part of m_j (refer to [6] for the definition of m_j). The condition of stress continuity on the interface relates $K^{(1)}$ and $K^{(2)}$ as $\overline{K^{(1)}} = K^{(2)} \equiv K$ [6]. In the above expressions, $\cosh(\pi\varepsilon_j)\sqrt{K^{(j)}\overline{K^{(j)}}}$ denotes stress intensity at the debonded tip similar to stress intensity factor for a homogeneous crack. Thus, $\cosh(\pi\varepsilon_j)\sqrt{K^{(j)}\overline{K^{(j)}}}$ is called stress intensity of debonding (SID) in this paper to distinguish it from that of a homogeneous case. For the homogeneous case, $\varepsilon_j = 0$, and SID reverts to $\sqrt{K_{\mathrm{I}}^2 + K_{\mathrm{II}}^2}$ of a homogeneous crack.

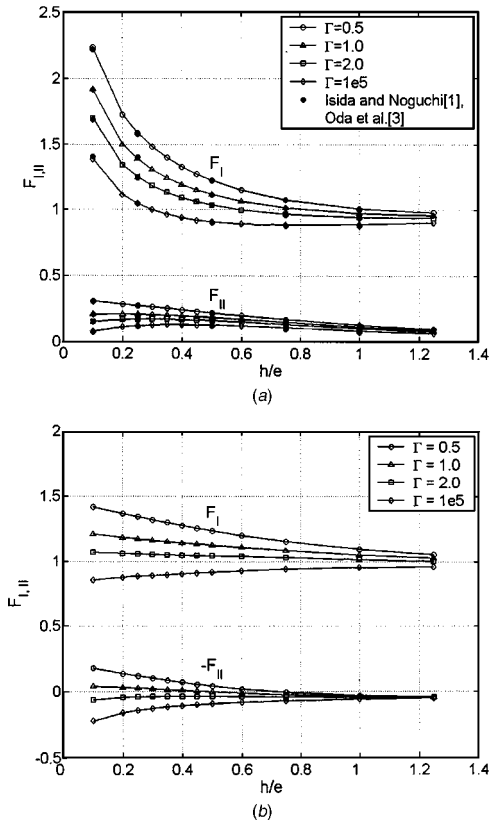


Fig. 2 Normalized SIF of an internal crack parallel to the interface and interacting with an interfacial crack ($a/c=1$, $b/c=0$, $e/c=2$) (loading normal to the interface): (a) crack tip A (b) crack tip B; ($\nu_1=\nu_2=0.3$; $F_{\mathrm{I,II}}=K_{\mathrm{I,II}}/p\sqrt{\pi c}$)

Complex stress intensity factor for the interfacial circular hole edge crack is given by

$$K_D = K_1 + \int_{-c}^{c}[h_n(t)K_2(t) + h_\tau(t)K_3(t)]dt, \tag{5}$$

where $K_2(t)$ and $K_3(t)$ represent complex stress intensity factors of the interfacial hole edge crack induced by the point dislocation at point t in normal and tangential directions, respectively. K_1 denotes complex stress intensity factor of the interfacial crack due to far-field loading. K_1, K_2, and K_3 at the debonding crack tip D (see Fig. 10 below) are calculated by [6]

$$K_j = 2\sqrt{2\pi}\exp(-\pi\varepsilon_1) \times \frac{|\omega'(\beta)(\beta-\alpha)|^{m_1}F_1(\beta)\exp[-im_1(\pi+\delta/2)]}{\omega'(\beta)(\beta-\alpha)}, \quad (j=1,2,3). \tag{6}$$

The definitions of various terms in Eq. (6) and the procedure to evaluate them are detailed in [6].

3 Numerical Results

The above procedure will now be applied to solve a few crack interaction problems. Assuming *plane stress* condition for all the cases considered in this paper, two distinct internal crack orientations are considered for numerical simulations: (a) crack parallel to the interface (b) and crack perpendicular to the interface. Stress intensity factors (SIF) at the crack tips of an internal crack are evaluated and their normalized values are plotted. Normalization

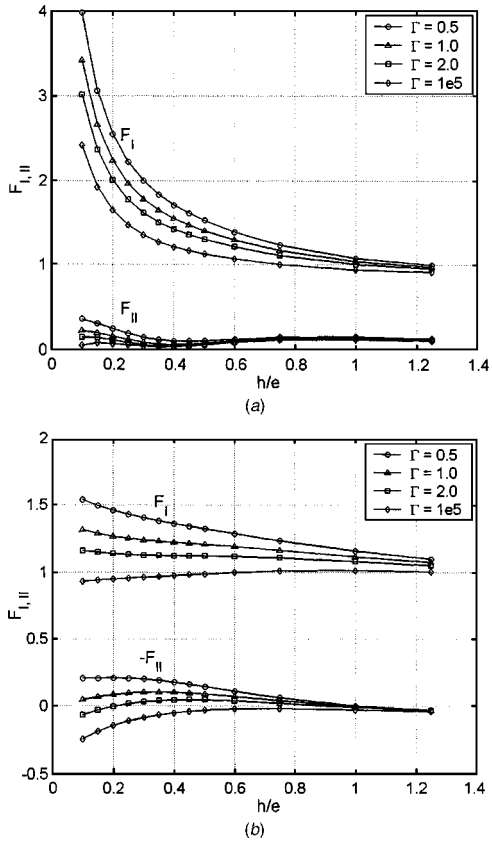


Fig. 3 Normalized SIF of an internal crack parallel to the interface and interacting with an interfacial circular hole ($a/c=1$, $b/c=1$, $e/c=2$) (loading normal to the interface): (*a*) crack tip *A* (*b*) crack tip *B*; ($\nu_1=\nu_2=0.3$; $F_{I,II}=K_{I,II}/p\sqrt{\pi c}$)

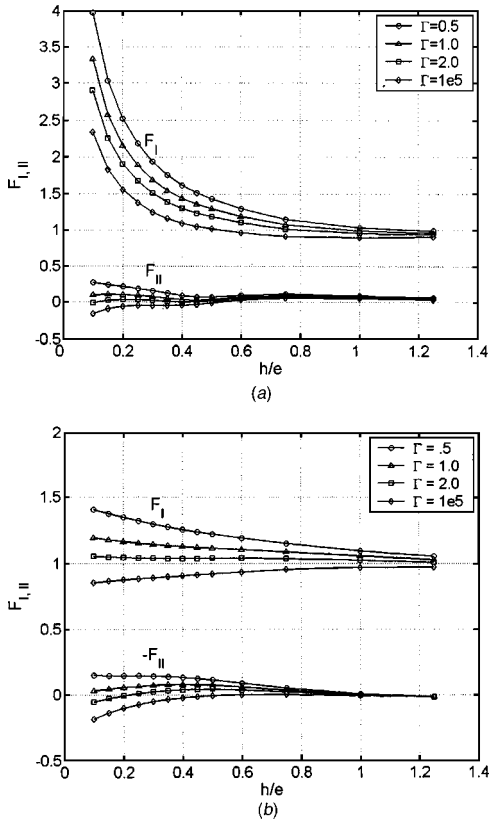


Fig. 4 Normalized SIF of an internal crack parallel to the interface and interacting with an interfacial elliptical hole ($a/c=0.5$, $b/c=1$, $e/c=1.5$) (loading normal to the interface): (*a*) crack tip *A* (*b*) crack tip *B*; ($\nu_1=\nu_2=0.3$; $F_{I,II}=K_{I,II}/p\sqrt{\pi c}$)

is done with respect to $p\sqrt{\pi c}$, where p is the remote stress normal/parallel to the interface and c is half-length of the internal crack. The value of c is taken as unity in all the simulations. Shear modulus and Poisson's ratio for materials 1 and 2 are given by μ_j and ν_j $(j=1,2)$, respectively. Poisson's ratio of $\nu_1=\nu_2=0.3$ is taken for all the cases considered. $\Gamma=\mu_2/\mu_1$ denotes the elastic mismatch between the two bonded dissimilar half-planes. All the numerical results obtained from the numerical procedure outlined in Sec. 2 are accurate to within 0.1%. This is accomplished by evaluating the difference between the results obtained from two successive values of M (Gauss points) and calculating the error percentage. In the present paper three values of M (8, 16, 32) are taken for numerical purposes.

The internal crack orientation parallel to the bi-material interface will be considered first. Figure 1(*a*) shows an elliptical hole on a bimaterial interface interacting with an internal crack *AB* located at a vertical distance h from the interface in material 1 $(y>0)$. The center of the internal crack (z_c) is offset by $e=(c+a)$ from the y-axis for all the cases with this crack orientation. Remote loadings p are applied normal to and γp on material 2 parallel to the interface as shown in Fig. 1(*a*). $\gamma=2(2\beta_D-\alpha_D)/(1-\alpha_D)$ where α_D and β_D are the Dundurs parameters. The solution for the problem of an elliptical hole at the interface of two bonded half planes subjected to far-field loading normal to the interface [6] will be used to find the left-hand side of Eq. (3).

The efficiency of the above procedure is verified by solving the problem of an interfacial crack interacting with a parallel internal crack. This problem can be modeled by taking the value b of the elliptical hole zero (see Fig. 1(*a*)). SIF for this problem have been compiled in Murakami et al. [1] and reworked by Oda et al. [3] using a different procedure. Figure 2 shows normalized stress intensity factors at crack tips *A* and *B*. It can be seen that there is a good agreement between the results of Fig. 2 and [1,3]. Figure 3 shows the normalized SIF of an internal crack interacting with a circular hole on the interface. The values of SIF at crack tip *A* are higher in this case compared to the previous case. This can be attributed to the hoop stress on the circular hole that tends to increase the SIF as the crack moves closer to the hole (i.e., as h becomes small). However, SIF at crack tip *B* show little variation when compared to the previous case. Figure 4 shows normalized SIF of an internal crack interacting with an elliptical hole whose minor axis is along the interface. Results show similar trend as in the case of a circular hole. The hoop stress on the elliptical hole surface tends to increase SIF of crack tip *A*. Figure 5 shows the normalized SIF of an internal crack interacting with an elliptical hole whose major axis is along the interface. The magnitude of mode-I SIF is less when compared to previous two cases (Figs. 3(*a*) and 4(*a*)). However, the SIF at crack tip *B* show little variation for all the interfacial hole shapes considered implying that the shape of the cavity has little effect on the SIF at crack tip *B*.

The problem of an internal crack perpendicular to the interface will now be considered as shown in Fig. 1(*b*). Remote loadings of p on material 1 and μp on material 2 is applied parallel to the interface where $\mu=(1+\alpha_D)/(1-\alpha_D)$. The center of the internal crack (z_c) is at a distance of $(h+c+b)$ from the interface. The SIF at both the crack tips of the internal crack will be evaluated and their normalized values will be plotted against the normalized distance h/c. The solution for the problem of an elliptical hole at the interface of two bonded half-planes subjected to far-field loading parallel to the interface [7] will be used to find the left-hand side of Eq. (3).

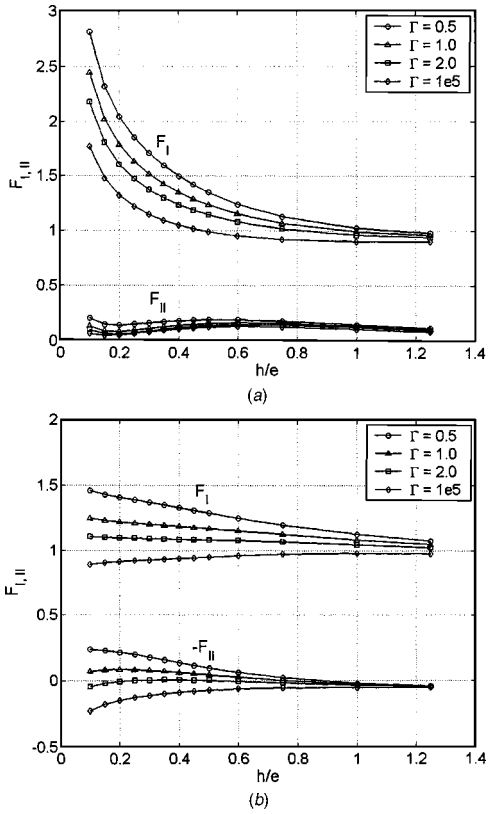


Fig. 5 Normalized SIF of an internal crack parallel to the interface and interacting with an interfacial elliptical hole ($a/c=1$, $b/c=0.5$, $e/c=2$) (loading normal to the interface): (*a*) crack tip *A* (*b*) crack tip *B*; ($\nu_1=\nu_2=0.3$; $F_{I,II}=K_{I,II}/p\sqrt{\pi c}$)

Figure 6 shows the normalized SIF of an internal crack interacting with a circular hole. It can be seen from the plot that the SIF at crack tip *A* nearer to the circular hole increases indefinitely as the distance between them becomes small. Figure 7 shows the normalized SIF of an internal crack interacting with an elliptical hole whose major axis is perpendicular to the interface. Figure 8 shows the normalized SIF of an internal crack interacting with an elliptical hole whose major axis is parallel to the interface. Figure 9 shows the normalized SIF of an internal crack interacting with a crack bisecting the bimaterial interface ($a=0$).

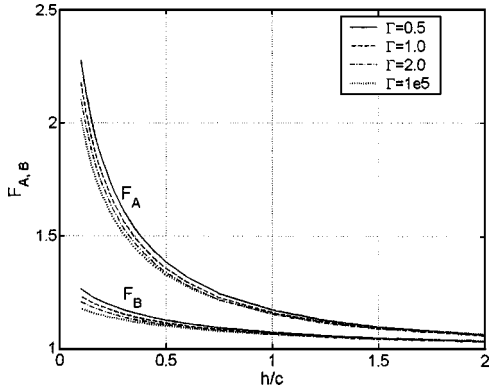


Fig. 6 Normalized SIF of an internal crack perpendicular to the interface and interacting with an interfacial circular hole ($a/c=1$, $b/c=1$) (loading parallel to the interface); ($\nu_1=\nu_2=0.3$; $F_{A,B}=K_{A,B}/p\sqrt{\pi c}$)

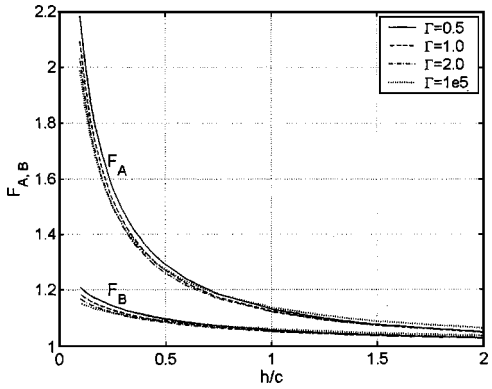


Fig. 7 Normalized SIF of an internal crack perpendicular to the interface and interacting with an interfacial elliptical hole ($a/c=0.5$, $b/c=1$) (loading parallel to the interface); ($\nu_1=\nu_2=0.3$; $F_{A,B}=K_{A,B}/p\sqrt{\pi c}$)

The cases considered above involve varying the vertical distance *h* of the internal crack from the interface and examining the variation in SIF. Cracks may also evolve from cavities under external loading. Cavities are a major source of stress concentrations from which cracks are most likely to emanate. Such cracks may interact with internal cracks leading to their growth and coalescence. Crack growth due to void growth and coalescence consti-

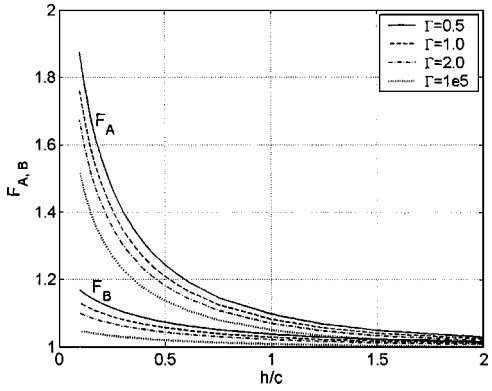


Fig. 8 Normalized SIF of an internal crack perpendicular to the interface and interacting with an interfacial elliptical hole ($a/c=1$, $b/c=0.5$) (loading parallel to the interface); ($\nu_1=\nu_2=0.3$; $F_{A,B}=K_{A,B}/p\sqrt{\pi c}$)

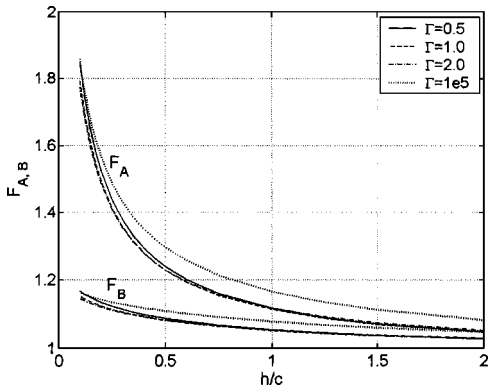


Fig. 9 Normalized SIF of an internal crack perpendicular to the interface and interacting with an interfacial crack ($a/c=0$, $b/c=1$) (loading parallel to the interface); ($\nu_1=\nu_2=0.3$; $F_{A,B}=K_{A,B}/p\sqrt{\pi c}$)

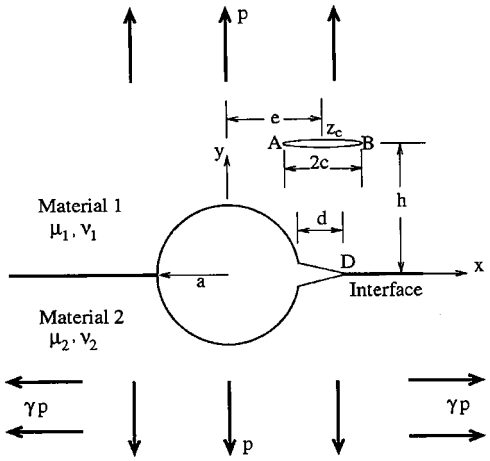


Fig. 10 Interfacial hole edge crack interacting with an internal crack

tutes a major study in the field of ductile fracture. A number of studies have been carried out to understand the role of cracks emanating from holes of various shapes. A compilation of studies on hole edge crack problems can be found in Murakami et al. [1].

Hasebe and Chen [8] have studied the problem of a circular hole edge crack interacting with an internal crack in a homogeneous medium by applying the rational mapping technique. Okumura et al. [6] and Hasebe et al. [7] have examined the problems of an elliptical hole on the interface subjected to far-field loads acting normal to and parallel to the interface, respectively. Their work will now be extended to study internal crack interacting with interfacial cavity. The problem of a circular hole edge crack on a bimaterial interface interacting with an internal crack will now be examined (Fig. 10). The center of the internal crack (z_c) is offset from the y-axis by $e=a+c$. SIF at crack tips A and B of the internal crack and SID of the debonded tip D of the interfacial hole edge crack are evaluated to study crack interaction effects.

Figure 11 shows the normalized SID $(\cosh(\pi\varepsilon_1)\sqrt{K_D\overline{K_D}}/p\sqrt{\pi c})$ of a hole edge interface crack with increasing debond length. It can be seen that SID increases with debond length. It is interesting to note the variation of SID, which shows an initial increase followed by a flat portion ($1.4<d/c<1.7$) and a final increase as Γ becomes large. The flat portion of the curve indicates that the tendency of increasing SID due to increasing debond length is hindered by crack shielding due to internal crack. For small values of Γ the initial rapid increase in SID is slowed down due to internal crack shielding. Figure 12 shows the variation of a normalized

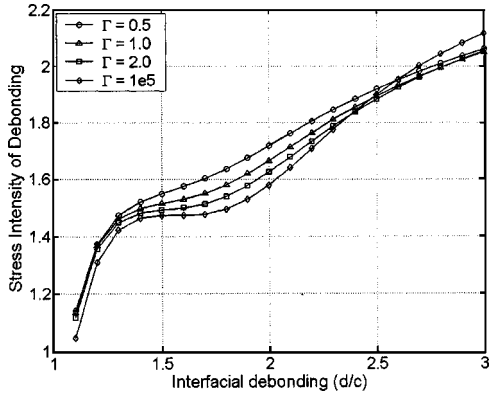


Fig. 11 Normalized stress intensity of debonding (SID) of an interfacial hole edge crack ($a/c=2$, $e/c=3$, $h/c=2$; $\nu_1=\nu_2=0.3$; $\mathrm{SID}=\cosh(\pi\varepsilon_1)\sqrt{K_D\overline{K_D}}/p\sqrt{\pi c}$)

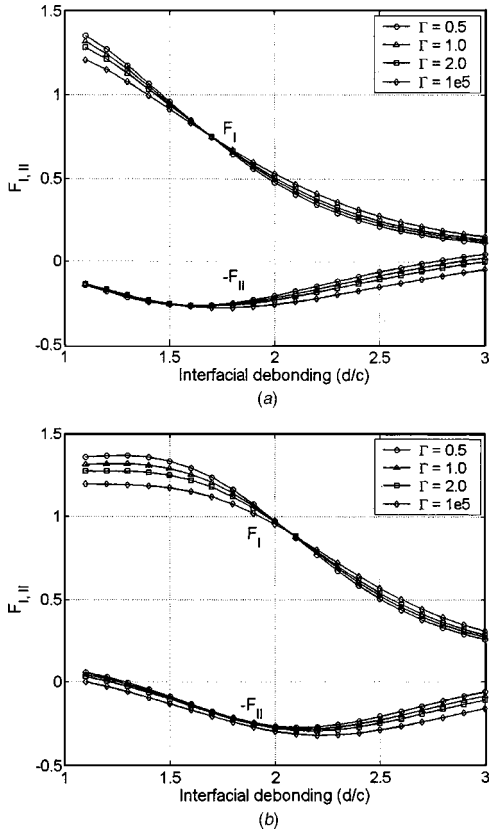


Fig. 12 Normalized SIF of an internal crack interacting with an interfacial circular hole edge crack (loading normal to the interface): (*a*) crack tip *A* (*b*) crack tip *B* ($a/c=2$, $e/c=3$, $h/c=2$; $\nu_1=\nu_2=0.3$; $F_{I,II}=K_{I,II}/p\sqrt{\pi c}$)

SIF of an internal crack with respect to the interfacial debond length. Mode-I SIF at both the tips of internal crack decrease with increasing debond length indicating crack shielding by the hole edge crack. Further, the variation in mode-I SIF (F_I) at crack tip B is negligible for $d/c<1.5$. This shows that interfacial debonding has little effect on tip B of the internal crack until $d/c\approx1.5$. As the debond length increases more and more of the internal crack is shielded further decreasing SIF at tips A and B of the internal crack.

4 Conclusions

This paper addresses crack interaction problems between an internal crack and interfacial cavity. The solution of a point dislocation interacting with an elliptical cavity has been obtained earlier [2]. By applying the distributed dislocation technique this fundamental solution is used to simulate internal crack. Since the fundamental solution satisfies boundary condition on the interfacial hole, the stress distribution and stress concentration can be calculated easily and exactly on the hole surface. Two orientations of internal crack with respect to the interface have been considered. When the internal crack is parallel to (loading normal to) the bimaterial interface, it is seen that the SIF of the internal crack increases with decreasing distance (h/c) from the interface. In particular, SIF at crack tip A is larger for the case of a circular hole compared to other hole shapes considered. This can be attributed to the hoop stress on the hole surface which tends to increase the SIF. The SIF at crack tip A decreases as the dimension of the elliptical hole perpendicular to the interface decreases. When the internal crack is perpendicular to the interface, it is seen that the SIF of crack tip A of the internal crack increases indefinitely as the interfacial hole is approached. An interesting case of an edge

crack from a circular interfacial hole interacting with an internal crack is also considered. As the debond length increases it is observed that more and more of internal crack is shielded by the hole edge crack. The SID at the debonding tip D is seen to increase initially and then remains unchanged for large values of Γ before increasing finally. This indicates crack shielding by the internal crack on the interfacial hole edge crack for debonding length in the range $1.4 < d/c < 1.7$. The method presented in this paper can also be used to solve multiple crack problems.

References

[1] Murakami, Y., et al., 1992, *Stress Intensity Factors Handbook*, Elsevier Science, New York.

[2] Prasad, P. B. N., Hasebe, N., Wang, X. F., and Shirai, Y., 2003, "Green's Function of a Bimaterial Problem With a Cavity on the Interface. Part I: Theory," **72**, pp. 389–393.

[3] Oda, K., Noda, N., and Arita, S., 2003, "Stress Intensity Factors for Interaction between Interface Crack and Internal Crack and for Kinked Interface Crack in Bonded Semi-Infinite Planes," Key Eng. Mater., **243–244**, pp. 375–380.

[4] Chen, Y. Z., and Hasebe, N., 1992, "An Alternative Fredholm Integral Equation Approach for Multiple Crack Problem and Multiple Rigid Line Problem in Plane Elasticity," Eng. Fract. Mech., **43**, pp. 257–268.

[5] Erdogan, F., Gupta, G. D., and Cook, T. S., 1973, "Numerical Solution of Singular Integral Equations," in *Methods of Analysis and Solutions of Crack Problems*, edited by G. C. Sih, Noordhoff, Leyden, Chap. 7, pp. 369–425.

[6] Okumura, M., Hasebe, N., and Nakamura, T., 1995, "Bi-material Plane with Elliptic Hole Under Uniform Tension Normal to the Interface," Int. J. Fract., **71**, pp. 293–310.

[7] Hasebe, N., Okumura, M., and Nakamura, T., 1992, "Bonded Bi-material Half-Planes With Semi-elliptical Notch Under Tension Along the Interface," J. Appl. Mech., **59**, pp. 77–83.

[8] Hasebe, N., and Chen, Y. Z., 1996, "Stress Intensity Factors for the Interaction Between a Hole Edge Crack and a Line Crack," Int. J. Fract., **77**, pp. 351–366.

P. Liu
Department of Mechanical Engineering,
University of Wyoming,
Box 3295,
Laramie, WY 82071

R. V. Kukta
Department of Mechanical Engineering,
State University of New York,
Stony Brook, NY 11794-2300

D. Kouris
Department of Mechanical Engineering,
University of Wyoming,
Box 3295,
Laramie, WY 82071

Strain-Modulated Adatom and Surface Vacancy Pair Interactions

Adsorbed atoms (adatoms) and vacancies have a significant role to play in the physics of surfaces and the mechanisms of film growth on a substrate. This paper investigates the effect of applied or residual strain on the energetic interaction between pairs of adatoms and vacancies. The analysis is based on a continuum-level point-defect model, where adatoms and vacancies have strain-dependent properties. Atomistic calculations are used to evaluate the defect properties for Si(111). The result is used as input for the defect model in order to investigate the strength and character of the interaction versus strain, separation distance, and relative orientation of the defects. It is found that strain may cause the defects to align in certain direction and modulate their interaction between repulsion and attraction, providing a mechanism for controlled building of nanostructures. [DOI: 10.1115/1.1875392]

1 Introduction

Surface processes such as material deposition onto a substrate, surface reconstruction, and self-assembly are subjects of intense investigation due to their application in the development of nanostructured materials and devices. Understanding the mechanisms that control morphology during film growth might yield new fabrication techniques and methods for improving the reliability of micro- and nanoscale devices. Film growth occurs by material aggregation on various length scales, from clustering of adatoms to the coalescence of large islands. Adatom clustering sets the stage for subsequent growth and can greatly impact morphological development. Hence, the mechanisms that govern clustering are of fundamental interest for controlling growth.

Various models have been used to investigate material aggregation on a surface. Continuum analyses, molecular dynamics (MD), and kinetic Monte Carlo (KMC) are three widely used examples. MD models atomic processes by tracking atom motion as governed by interatomic potentials and classical equations of motion. They have been extensively used to study many processes including diffusion of adatoms (e.g., see Kallinteris et al. [1] and Evangelakis and Papanicolaou [2]). In KMC, atomic kinetics is modeled through a series of statistical events. It has been used to study adatom diffusion and growth of patterned nanostructures by Sabiryanov et al. [3], growth of adatom clusters and islands by Larsson [4], and self-assembly processes on surfaces by Larsson et al. [5]. In continuum analysis, surface defects are treated as point sources of strain. Lau and Kohn [6] treat an adatom as a point dilatation (force dipole), and Marchenko and Parshin [7] model an atomic surface step as a point dilatation and a moment on a half-plane. These models capture the induced displacement field at points sufficiently removed from the defect. The strength of the point source is usually obtained by matching displacements fields or defect-defect interaction energies to atomistic predictions (e.g., see Shilkrot and Srolovitz [8], Kouris et al. [9], and Peralta et al. [10]). When evaluating the source strength, care must be taken to ensure that the atomistic regime is sufficiently large, so that a finite region can be identified where the atomistic and continuum models converge without substantial boundaries effects. This is especially important in the case of surface steps where fields, unaccounted for by the Marchenko–Parshin model, may be significant at points quite far from the defect, as shown by Kukta and his co-workers [11,12]. Kukta et al. [13,14] discuss an alternative method for calculating the source strength from the energy of an isolated defect, as a function of applied strain. This allows one to estimate the point source from an atomistic model without relaxing atomic positions to their minimum energy configuration.

Defect point sources are typically treated as fixed quantities, independent of applied fields or fields induced by other defects. This is the assumption made by Lau and Kohn [6]. According to their model, an adatom on a half-space induces a strain field that decays as the inverse cube of distance and the interaction energy between adatoms decays similarly as the inverse cube of the separation. Neither the induced field nor the interaction energy depend on applied strain. Kukta et al. [13] report that adatom interactions should generally depend on strain. Otherwise a linear relationship is obtained between surface stress and adatom coverage, which contradicts experimental measurements by Ibach and his co-workers [15,16].

In a previous article [14], the authors developed a general three-dimensional framework for adatoms with strain-dependent source fields and discussed the nature of far-field interactions between adatoms. The term strain-dependent source field implies that the elastic field induced by a defect depends on the local value of strain. The present paper extends that work to include surface vacancies and investigates some other issues not previously addressed, including the near-field behavior. Like adatoms, surface vacancies will generally have strain dependent properties. Many phenomena that occur on surfaces involve mutual interactions between adatoms and vacancies. It is known, for example, that strain may cause a flat surface to roughen. This phenomenon has been studied extensively in a macroscopic framework where the surface evolves as a smooth continuous function (e.g., see Asaro and Tiller [17], Grinfeld [18], Freund and Jonsdittir [19], Kukta and Freund [20], Shilkrot and Srolovitz [21]). On the atomic scale, roughening of a flat surface requires the formation of adatom-vacancy pairs. For the surface to evolve towards a rough morphology, atoms must move from the surface layer—leaving vacancies—to sites atop the surface—where they become adatoms. Subsequently, adatoms must combine to form mounds and vacancies must combine to form troughs. This process occurs through their mutual interactions. The macroscopic evolution of the system is determined by complex interactions among large ensembles of defects. It is not the purpose of this paper to draw conclusions about evolution on the macroscale, but rather to provide a basic ingredient necessary for linking atomic scale proper-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, October 27, 2003; final revision; October 6, 2004. Editor: Robert McMeeking. Discussion on the paper should be addressed to the Editor, Professor Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

ties with the macroscale evolution, namely the strain dependence of defect interactions. It is noted that while the focus is on adatoms and vacancies, the analysis also applies to adatom and vacancy clusters and to larger defects like nano- and micrometer scale islands. This paper addresses only mechanical interactions. It is noted that entropic interactions may be large, particularly for small defects as adatoms and vacancies. Unless temperature is low, one must also account for entropic effects.

The next section reviews the model. Constitutive constants for Si adatoms and surface vacancies on Si(111) are estimated and their far-field interactions are discussed. It is found that depending on strain, defect pairs may repel or attract each other and orient themselves in certain directions. It is also observed that the model accounts for the distinct near- and far-field behavior. In Sec. 3 the near-field interaction is evaluated and the nature of the interaction is compared with the far-field results.

2 Mechanics of an Elastic Point Phase

Many structural elements in materials systems can be treated as point phases. Some examples are adatoms, surface and bulk vacancies, interstitial and substitutional atoms, surface clusters, islands, and bulk inclusions. The term *elastic point phase* refers to cases where the elastic field induced by a point phase depends on strain. This section reviews such a model that was developed elsewhere [13,14]. It is applied to adatoms and vacancies in the following section.

Consider a half-space with energy per unit volume $W(\varepsilon)$ where ε denotes strain. A point phase is placed at $\mathbf{x}=\mathbf{y}(t)$ on the surface. It has an energy $\Psi(\overline{\varepsilon}^T-\overline{\varepsilon}^*)$ that depends on the total surface strain $\overline{\varepsilon}^T$ evaluated at the point phase, minus the strain $\overline{\varepsilon}^*$ that is produced by the point phase itself. The self-strain $\overline{\varepsilon}^*$ must be omitted because it is singular at the point phase. Strains ε and ε^* are considered to be infinitesimal. Bulk stress is defined as

$$\sigma_{ij}(\varepsilon)=\frac{\partial W(\varepsilon)}{\partial \varepsilon_{ij}} \tag{1}$$

The field quantity $\sigma_{ij}(\varepsilon)$ is the total stress, which includes the contribution of externally applied loads and the self-stress (residual stress) induced by the point phase. The self-stress of the point phase is evaluated as $\sigma^*_{ij}=\sigma_{ij}(\varepsilon^*)$. The total free energy of the system is written as the sum of the bulk and point phase contributions plus the work done by applied loads, and the elastic field is found by minimizing energy with respect to compatible strain fields, ε and ε^*. The following relations are found:

$$\frac{\partial \sigma_{ij}}{\partial x_j}-\frac{\partial \sigma^*_{ij}}{\partial x_j}=0 \tag{2}$$

and

$$\sigma^*_{ij,j}+\delta_{i\alpha}D_{\alpha\beta}\frac{\partial \delta(\mathbf{x})}{\partial x_\beta}=0 \tag{3}$$

in the half-space volume and

$$\sigma_{ij}n_j=0 \tag{4}$$

in addition to

$$\sigma^*_{ij}n_j=0 \tag{5}$$

on the surface of the half-space. Repeated indices are summed, $\delta_{i\alpha}$ is Kronecker's delta symbol, $\delta(\mathbf{x})$ is the Dirac delta function, and n_j denotes the outward unit normal at the half-space surface. Latin indices denote components of a three-dimensional tensor, while greek indices denote components of a two-dimensional surface tensor

$$D_{\alpha\beta}(\overline{\varepsilon}^T-\overline{\varepsilon}^*)=\frac{\partial \Psi(\overline{\varepsilon}^T-\overline{\varepsilon}^*)}{\partial \varepsilon_{\alpha\beta}} \tag{6}$$

is the stress of the point phase. Equations (1)–(6) along with remote boundary conditions determine the elastic field. It follows from Eq. (3) that the elastic field induced by the point phase is that of a force dipole $D_{\alpha\beta}$ applied on the surface. According to (6) and due to symmetry of the strain, $D_{\alpha\beta}=D_{\beta\alpha}$, which implies that the dipole has no net moment. If the point phase has an energy Ψ that is independent of strain, its dipole vanishes and it induces no elastic field. If the energy is linear in strain, the dipole is a constant and the induced elastic field is independent of strain. However, in general the dipole will depend on strain and the point phase is said to be elastic. The displacement field produced by the point phase is evaluated from the surface elastic Greens function $G_{ij}(\mathbf{x}-\mathbf{y})$ as

$$u^*_i(\mathbf{x})=D_{\alpha\beta}\frac{\partial G_{i\alpha}(\mathbf{x}-\mathbf{y})}{\partial x_\beta}. \tag{7}$$

Energy as a function of strain characterizes the elastic field of a point defect and how it changes in the presence of an applied strain and strains associated with other defects, surface heterogeneities, etc. Typically the elastic field of a point defect is found by comparing atomic displacements of a fully relaxed atomistic simulation with the continuum field produced by a point dipole on an elastic half-space [e.g., Refs. [9,11,12]]. Comparing displacements can be very time consuming, particularly for anisotropic materials where analytical forms for the displacement field are not generally available. Equation (6) provides an alternative. Defect energy $\Psi(\overline{\varepsilon}^0)$ is calculated as the increase in energy when the defect is introduced on a surface that is uniformly strained by the amount $\boldsymbol{\varepsilon}^0$ and it is readily evaluated from an atomistic [13] or ab initio model [14]. With (6) it is possible to estimate the elastic field by calculating the energy from unrelaxed atomic positions. The estimate could provide an initial guess for finding the relaxed atomic positions.

Surface defects, such as adatoms and vacancies, are mobile and their motion is biased in the presence of a nonuniform strain field. They tend to move such that the free energy of the system decreases, which is characterized by an energetic driving force **d**. The driving force is evaluated from the variation in total free energy E with defect position $\mathbf{y}(t)$ as in [14]:

$$d_\alpha=-\frac{\delta E}{\delta y_\alpha}=-D_{\beta\gamma}\left(\frac{\partial \varepsilon_{\beta\gamma}}{\partial x_\alpha}-\frac{\partial \varepsilon^*_{\beta\gamma}}{\partial x_\alpha}\right)\Bigg|_{\mathbf{x}=\mathbf{y}(t)} \tag{8}$$

The free energy decreases as a defect moves in the direction of its driving force. If the self-strain of the defect is the only source of strain then $\varepsilon_{ij}=\varepsilon^*_{ij}$ and the driving force vanishes. This must be true because the energy of a surface point defect on an unbounded half-space is invariant with position. Likewise, the driving force vanishes in the case of a uniform applied strain. Nonuniform strains, other than the defect's self-strain, give rise to a nonzero driving force. One such example corresponds to the motion of a defect biased by the nonuniform strain field induced by another. This methodology is utilized in the study of interactions between adatoms and vacancies on Si(111) that follows.

3 Constitutive Relations for Adatoms and Vacancies

The Si(111) surface is chosen to illustrate the effect of strain on adatoms and vacancies. For simplicity the surface is considered to be unreconstructed. The Si(111) surface is elastically isotropic and it is therefore possible to approximate the elastic field of a defect using the isotropic elastic Green's function, further simplifying the analysis. Figure 1(*a*) illustrates the diamond cubic structure viewed on a cross-section normal to the (111) surface. The two shades of circles represent atoms on different planes parallel to the page. The structure is composed of an alternating sequence of these two layers. Figure 1(*b*) depicts a surface vacancy, wherein a

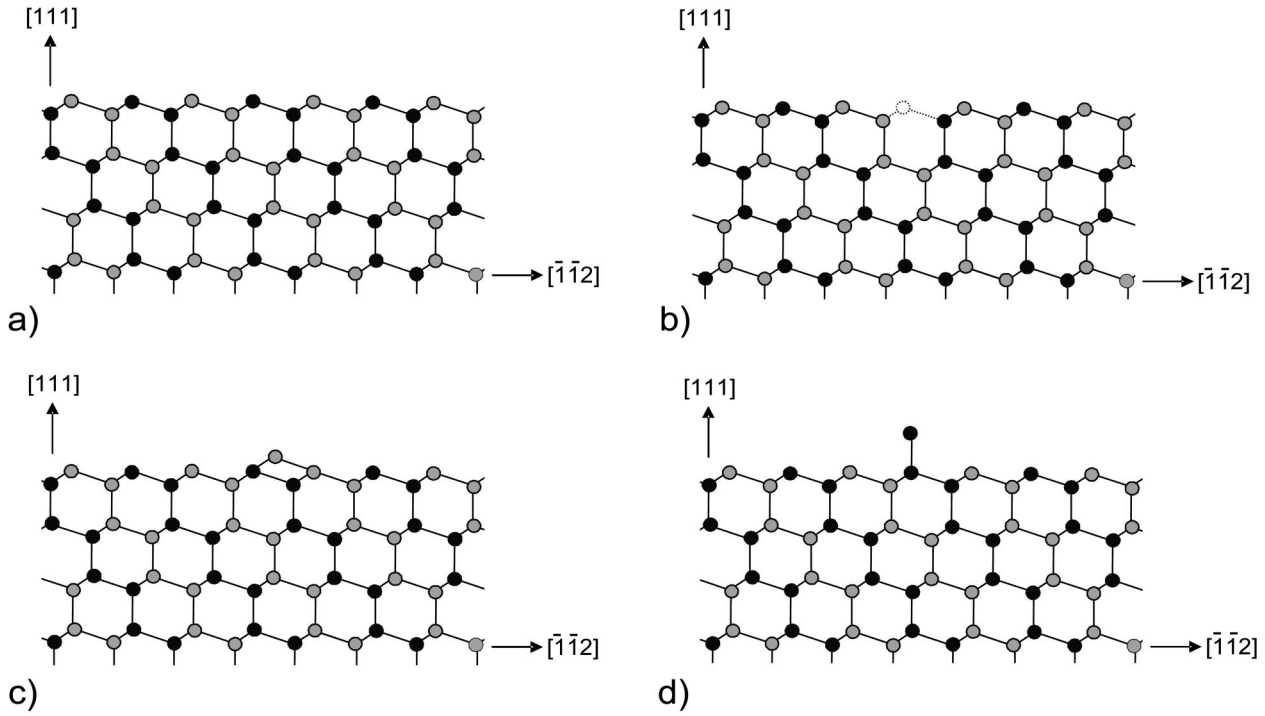


Fig. 1 Schematic view of the diamond cubic structure on a cross section normal to the (111) surface. Shown is (*a*) the unreconstructed defect-free surface, (*b*) a surface vacancy, and two different types of adatoms (*c*) and (*d*).

single atom is removed from the surface, and Figs. 1(*c*) and 1(*d*) depict adatoms on the surface. The adatom of Fig. 1(*d*) would continue the bulk crystal structure.

Each defect is characterized by energy as a function of strain. Defect energy is calculated using Tersoff's potential [22] with unrelaxed atomic positions. Use of the unrelaxed structures should not qualitatively affect the results. The energy of each system in Fig. 1 is calculated as a function of a uniform applied strain $\boldsymbol{\varepsilon}^0$. Let $\mathbf{x}^I$ be the position of adatom I in the unstrained configuration. Strain $\boldsymbol{\varepsilon}^0$ is imposed by transforming the structure as $(\mathbf{I}+\boldsymbol{\varepsilon}^0)\mathbf{x}^I$, where $\mathbf{I}$ is the identity tensor, energy is evaluated and the result is expanded as a Taylor series about zero strain. As the surface is considered to be traction-free, strain $\boldsymbol{\varepsilon}^0$ is constrained to enforce the boundary condition. To find the constraint on $\boldsymbol{\varepsilon}^0$, energy W of the bulk structure is determined for an arbitrary strain and the stress-strain relation is evaluated from (1). Considering the $\mathbf{e}_1$ direction to be normal to the surface and $\mathbf{e}_2$ and $\mathbf{e}_3$ directions to be in the plane of the surface, strains ε^0_{11}, ε^0_{12}, and ε^0_{13} found in terms of the surface in-plane strains ε^0_{22}, ε^0_{33}, and ε^0_{23} such that the surface traction vanishes. Defect energies depend only of the surface strain because of this constraint. The energy of each defect is determined by subtracting the energy of Fig. 1(*a*) from those of Figs. 1(*b*)–1(*d*). For defects on the (111) surface, energy is of the form

$$\Psi(\overline{\boldsymbol{\varepsilon}}) = \Psi_0 + D\varepsilon_{\alpha\alpha} + F\varepsilon_{\alpha\beta}\varepsilon_{\alpha\beta} + \frac{1}{2}H\varepsilon_{\alpha\alpha}\varepsilon_{\beta\beta} \tag{9}$$

and using (6) the defect dipole is

$$D_{\alpha\beta}(\overline{\boldsymbol{\varepsilon}}) = D\delta_{\alpha\beta} + 2F\varepsilon_{\alpha\beta} + H\varepsilon_{\gamma\gamma}\delta_{\alpha\beta} \tag{10}$$

where $\alpha, \beta = 2, 3$. For the structures of Figs. 1(*b*)–1(*d*) respectively

$$\Psi_0^b = 6.04 \text{ eV}, \quad D^b = 0.37 \text{ eV}, \quad F^b = -12.47 \text{ eV},$$
$$H^b = -15.47 \text{ eV} \tag{11}$$

$$\Psi_0^c = -5.36 \text{ eV}, \quad D^c = -2.47 \text{ eV}, \quad F^c = -14.25 \text{ eV},$$
$$H^c = 40.87 \text{ eV} \tag{12}$$

and

$$\Psi_0^d = -2.02 \text{ eV}, \quad D^d = -0.68 \text{ eV}, \quad F^d = 3.13 \text{ eV},$$
$$H^d = -1.02 \text{ eV} \tag{13}$$

According to these estimates, the adatom of Fig. 1(*c*) typically has lower energy than the one of Fig. 1(*d*). It is only in very extreme cases that the adatom of Fig. 1(*d*) has the lower energy, as in the case where biaxial compressions are larger than about 23%, which is too large to be meaningful here. Since the adatom of Fig. 1(*d*) has such a large energy, only the one of Fig. 1(*c*) is considered further.

4 The Interaction Between Adatoms and Vacancies

Surface defects, like adatoms and vacancies, interact through the coupling of their induced elastic fields. Analysis of such interactions between surface defects is important in an effort to understand their relative motion on a surface. In this section, the defect induced strain field is determined, based on the model presented earlier. In addition, the interaction energy and driving force between two surface defects is analyzed.

It is worth noting that the terms "surface defect," "dipole," "adatom," "vacancy," and "point phase" can be used interchangeably in this derivation since the necessary mathematical treatment is identical.

With the displacement field given in Eq. (7), the strain field caused by a defect can be written as

$$\varepsilon^*_{\alpha\beta} = \tfrac{1}{2}D_{\eta\chi}(G_{\alpha\eta,\chi\beta} + G_{\beta\eta,\chi\alpha}) \tag{14}$$

where the defect dipole $D_{\eta\chi}$ depends not only on the applied strain but also on the strain induced by other defects. To avoid the singularity caused by the defect at its location, the self-induced

strain is excluded in the calculation of the dipole [14]. As a result, when two surface defects are considered and the total surface strain and the applied strain denoted as ε_{ij}^{T} and ε_{ij}^{R}, respectively, the strain due to the defect dipole is represented by

$$\varepsilon_{\alpha\beta} = \varepsilon_{\alpha\beta}^{T} - \varepsilon_{\alpha\beta}^{*I(I)} \tag{15}$$

or

$$\varepsilon_{\alpha\beta} = \varepsilon_{\alpha\beta}^{R} + \varepsilon_{\alpha\beta}^{*I(II)} \tag{16}$$

where the superscript *A(B)* denotes the effect of defect "B" at location "A." Therefore, $\varepsilon_{\alpha\beta}^{*I(I)}$ and $\varepsilon_{\alpha\beta}^{*I(II)}$ represent the induced strain at defect I due to itself and due to defect II.

In this study, only mechanical interactions between defects are considered without the inclusion of coupling effects from other surface processes. It is noted that surface reconstructions are not considered but are readily amenable to the analysis. In the context of the present model, differently reconstructed surfaces result in different constitutive constants for an adatom, which can be evaluated from an atomistic model as discussed earlier. Adatom properties depend on its species and also on the species and structure of its substrate.

Following Eq. (14), the induced strain by defect II at defect I is expressed as

$$\varepsilon_{\alpha\beta}^{*I(II)} = \tfrac{1}{2}\{D^{(II)}\delta_{\eta\chi} + 2F^{(II)}[\varepsilon_{\eta\chi}^{R} + \varepsilon_{\eta\chi}^{*II(I)}] + H^{(II)}[\varepsilon_{\gamma\gamma}^{R} + \varepsilon_{\gamma\gamma}^{*II(I)}]\delta_{\eta\chi}\} \times (G_{\alpha\eta,\chi\beta} + G_{\beta\eta,\chi\alpha})_{x=x^{I}(t)} \tag{17}$$

where the dipole coefficients $D^{(II)}$, $F^{(II)}$, and $H^{(II)}$ are atomic properties associated with defect II. The equations for the induced strain at defect II due to defect I have the same format as Eq. (17) and can be obtained just by switching the superscript I and II. A detailed discussion on solving Eq. (17) for two different types of defects can be found in the Appendix.

For the simple case when the two defects are identical i.e., $D^{(I)}=D^{(II)}=D$, $F^{(I)}=F^{(II)}=F$, $H^{(I)}=H^{(II)}=H$, reciprocity simplifies (17) to

$$\varepsilon_{\alpha\beta}^{*} = \tfrac{1}{2}[D\delta_{\eta\chi} + 2F(\varepsilon_{\eta\chi}^{R} + \varepsilon_{\eta\chi}^{*}) + H(\varepsilon_{\gamma\gamma}^{R} + \varepsilon_{\gamma\gamma}^{*})\delta_{\eta\chi}]\,(G_{\alpha\eta,\chi\beta} + G_{\beta\eta,\chi\alpha})\big|_{x=x^{I}(t)} \tag{18}$$

Superscripts for the induced strain are omitted for simplicity. Furthermore, if the two defects are far apart, the induced strains is very small compared to the applied strain, since the induced strain decays with the inverse cubic of the distance. As a result, Eq. (18) yields a simple long-range solution for the strain

$$\varepsilon_{\alpha\beta}^{*(0)} = \tfrac{1}{2}[D\delta_{\eta\chi} + 2F\varepsilon_{\eta\chi}^{R} + H\varepsilon_{\gamma\gamma}^{R}\delta_{\eta\chi}]\,(G_{\alpha\eta,\chi\beta} + G_{\beta\eta,\chi\alpha})\big|_{x=x^{I}(t)} \tag{19}$$

This long-range solution can be used for a far-field analysis of the defect interaction.

With no simplifying assumptions associated with the distance between the two defects, the general solution of the induced strain can be obtained in terms of the long-range solution as

$$\varepsilon_{\alpha\beta}^{*} = \frac{1}{\omega}[\varepsilon_{\alpha\beta}^{*(0)} + Q_{\alpha\beta\eta\chi}\varepsilon_{\eta\chi}^{*(0)}] \tag{20}$$

where parameters ω and $Q_{\alpha\beta\eta\chi}$ depend on the dipole properties and Green function and are provided in the Appendix. In terms of the long-range solution, the dipole can be expressed as

$$D_{\alpha\beta} = D\delta_{\alpha\beta} + 2F\left\{\varepsilon_{\alpha\beta}^{R} + \frac{1}{\omega}[\varepsilon_{\alpha\beta}^{*(0)} + Q_{\alpha\beta\eta\chi}\varepsilon_{\eta\chi}^{*(0)}]\right\} + H\left\{\varepsilon_{\gamma\gamma}^{R} + \frac{1}{\omega}[\varepsilon_{\gamma\gamma}^{*(0)} + Q_{\gamma\gamma\eta\chi}\varepsilon_{\eta\chi}^{*(0)}]\right\}\delta_{\alpha\beta} \tag{21}$$

The energy change of a defect with respect to its position provides the driving force associated with its motion on the surface. Under a homogeneously applied external strain, the driving force at defect I is given by

$$d_{\alpha}^{(I)} = -\ \tfrac{1}{2}D_{\beta\gamma}^{(I)}D_{\eta\chi}^{(II)}(G_{\beta\eta,\chi\gamma\alpha} + G_{\gamma\eta,\chi\beta\alpha})\big|_{\mathbf{x}=\mathbf{x}^{I}(t)} \tag{22}$$

where dipoles $D_{\beta\gamma}^{(I)}$ and $D_{\eta\chi}^{(II)}$ of defect I and II can be calculated from (21) when the two defects are of the same type. It is important to point out that Eq. (22) is valid for general cases where the physical and atomic properties of the two defects differ, i.e., $D^{(I)} \neq D^{(II)}$. The Appendix includes a discussion on how to obtain the general solution for two defects in such cases.

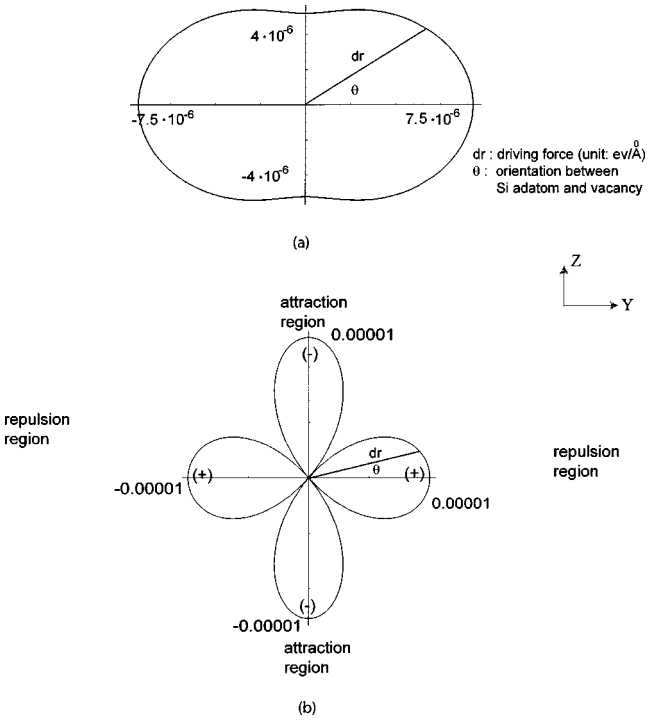


Fig. 2 The strain effect on the driving force: (*a*) Si on Si, uniaxial strain $\varepsilon_{zz}=0.005$, $r=3a_0$, always repulsion, and (*b*) Si on Si, uniaxial strain $\varepsilon_{zz}=0.02$, $r=3a_0$

5 Results and Discussion

Many surface processes occur due to the motion of defects and their continuing relocation. The interaction among defects plays a key role in determining their relative movement and final equilibrium positions. In order to investigate how defect interaction affects motion, we have utilized the driving force derived in previous sections for the cases of Si adatoms and vacancies. Figure 2 illustrates the magnitude of the driving force at the location of the vacancy, due to its interaction with an adatom and under far-field uniaxial strain. It is a polar plot that presents the driving force dependence on the relative angle between the line connecting adatom and vacancy and the coordinate axes. The distance between the two defects is constant and equal to three lattice spaces. If the applied strain is small, there is only repulsion between the Si adatom and vacancy [Fig. 2(*a*)]. As the applied strain increases, attraction can develop. Figure 2(*b*) illustrates clearly the division in attractive and repulsive regions for uniaxial strain $\varepsilon_{zz}=0.02$. Depending on their initial angular position, the two defects can either attract or repel each other.

A schematic of the predicted motion between the adatom and the vacancy is shown in Fig. 3. Initially the adatom and the vacancy are oriented so that they repel each other. The angular alignment will drive the adatom and vacancy into the attraction region where they will start to approach each other to shorten their separation. The stable equilibrium orientation is $\theta=0$ along the *Y* axis.

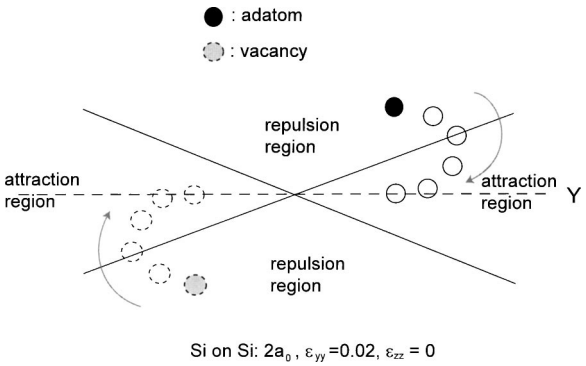


Fig. 3 The interaction between Si adatom and vacancy under large uniaxial strain

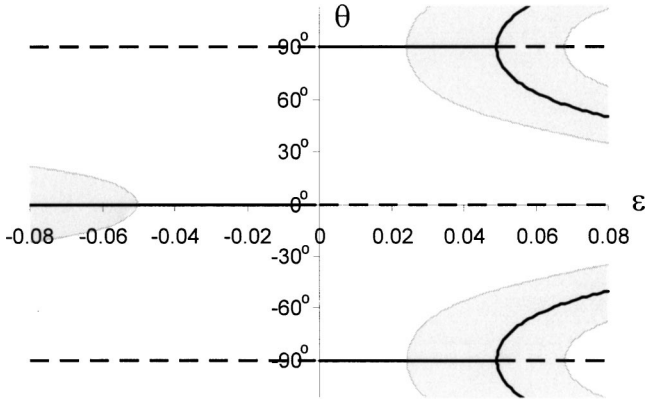


Fig. 6 Adatom-adatom interaction

To observe how the magnitude of strain influences defect interaction, a detailed investigation has been undertaken for the case of uniaxial strain. The notation and geometry are illustrated in Fig. 4. The orientation driving force and radial driving force have been used to determine stability and to identify attraction or repulsion regions. Figure 5 illustrates the interaction between two vacancies. Solid lines denote stable orientations, dashed lines unstable equilibrium orientations, while shaded regions denote attraction and unshaded regions denote repulsion. It was found that if compression is applied, the two vacancies are stable along the direction perpendicular to the applied strain. This result is independent of the magnitude of the applied compression. If a tensile strain is applied, a region of stable orientation exists that depends on the magnitude of the applied strain. It was also observed that the attraction region increases with increasing strain.

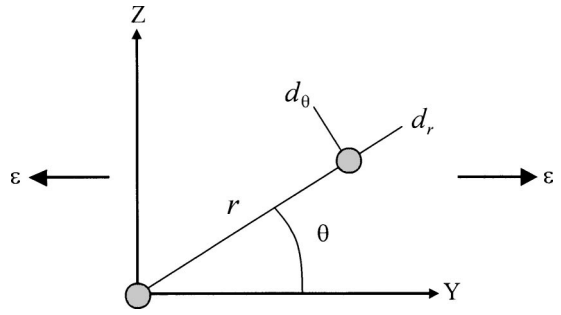


Fig. 4 Configuration to illustrate defect orientation and driving forces

Figure 6 illustrates the interaction between adatoms. The stable orientation for Si adatom interaction is quite different from that obtained from the interaction of vacancies. The direction perpendicular to the applied strain corresponds to stable orientation under small tensile loading.

The results of the adatom-vacancy interaction are presented in Fig. 7. They indicate that the stable orientation between adatom and vacancy is quite similar to that between two adatoms. It is important to note that adatoms and vacancies always attract each other under compression strain. Even when a small tensile strain is applied, their attraction region still dominates. This suggests that adatoms have the tendency to fill nearby vacancies in order to complete a perfect lattice structure. Even though the illustrations in Figs. 5–7 are based on far-field analysis under uniaxial loading, the numerical calculations of the near-filed under various of applied strains have produced very similar results.

For large separation distances, the effect of direct defect interaction is small compared to the effect of the applied strain. By ignoring this interaction when deriving the dipole properties, a far-field analysis can be easily obtained. For purposes of comparison, a near-field analysis has been performed. The calculation included defect-induced fields and defect interaction, for a variety of applied strains and separation distances. Results of the comparison between near and far-field are illustrated in Fig. 8, for different applied strains with a separation of two lattice spaces. It is clear that even for very small separations the differences are small. When the separation goes beyond five lattice spaces, even the quantitative differences between far and near-field become negligible. This confirms the fact that, in these problems, the far-field approximation can capture the nature of defect interaction qualitatively as well as quantitatively, as long as the defects are not located essentially next to each other. Nevertheless, one needs

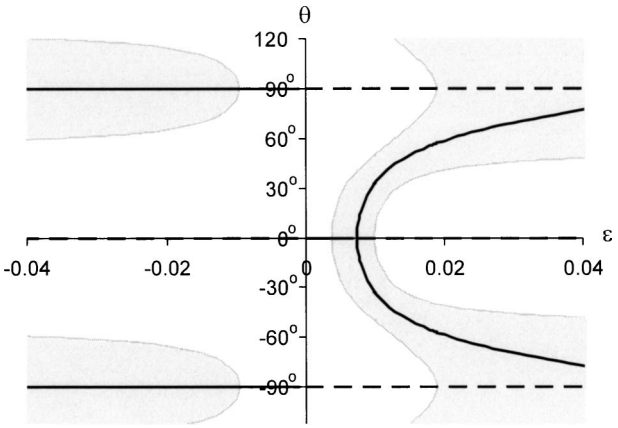


Fig. 5 Vacancy-vacancy interaction

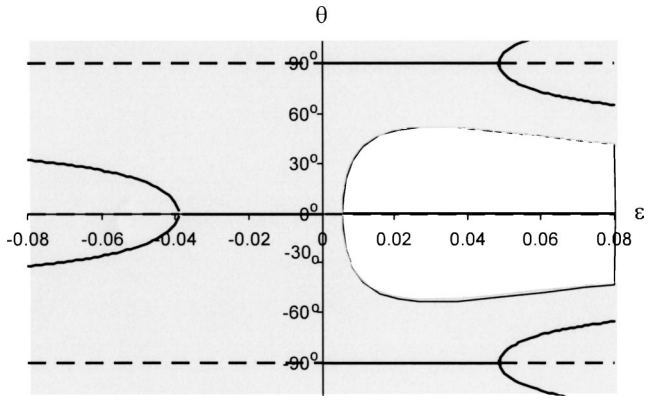


Fig. 7 Adatom-vacancy interaction

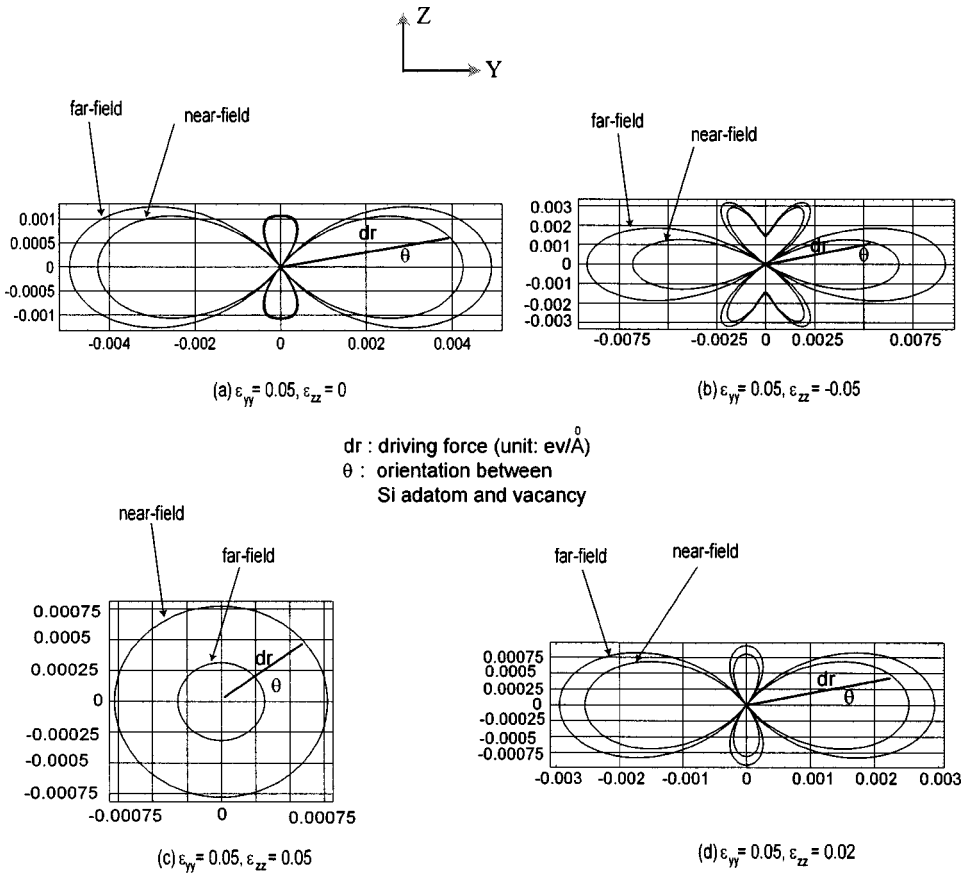


Fig. 8 Comparison of far- and near-field under separation $2a_0$

to keep in mind that other effects like electronic interaction can have a substantial influence on the driving force of the interaction.

In addition to the model described earlier, atomic simulations were performed in order to investigate how surface defects behave under the effect of applied strain. Figure 9 illustrates the motion of two Cu adatoms on a free (111) surface of a Cu substrate. In this simulation, an EAM molecular dynamics methodology has been implemented, utilizing the Johnson atomic potential [23,24]. All the modeling parameters used for the simulation originated from Johnson's semi-empirical results [23]. To implement the strain boundary condition, the Parrinello–Rahman [25,26] algorithm was used, with the volume of the modeled cell allowed to change with time. It can be found that the motion of the two adatoms permits them to align, in accordance with the influence of the applied strain. The simulation had been run long enough to confirm that the alignment corresponds to a stable orientation. This result was in agreement with the prediction by the continuum analysis.

Figure 10 illustrates how the two clusters of Cu adatoms behave on a (001) surface under applied strain. The simulation has shown that the applied strain did "help" the two clusters congregate to form a bigger cluster. This result further clarifies the possibility of manipulating the assembly process of nanostructure through controlling the external effects such as the applied strain. The continuum analysis of the interaction among clusters of surface defects is currently under investigation in our group.

6 Concluding Remarks

The purpose of the study presented in this manuscript is to quantitatively characterize the elastic interaction between surface defects in the presence of a mechanical field. It is a fully three-dimensional analysis that has emerged from the discrete adatom model developed earlier by the authors. The interaction between adatoms and vacancies has been examined for Si(111), as a func-

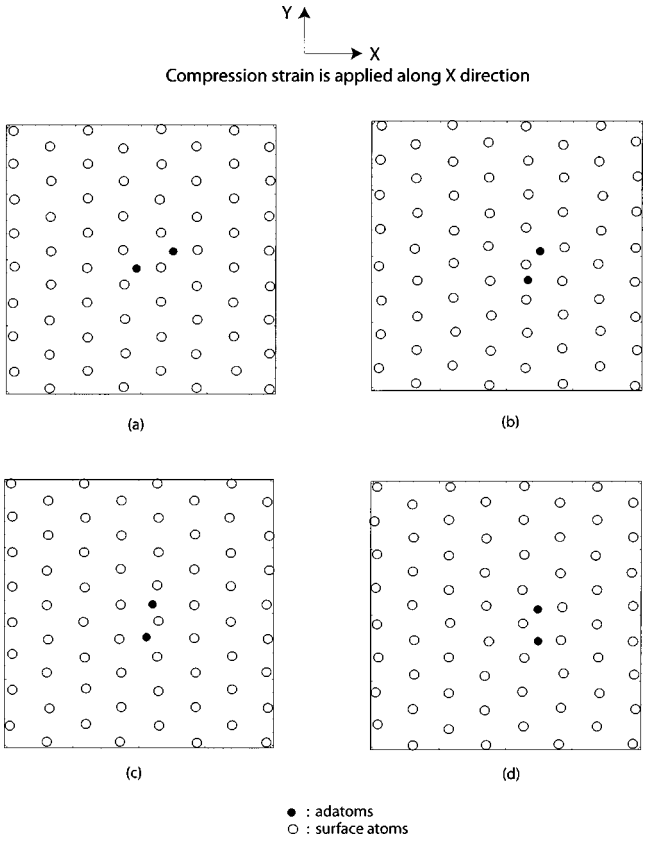


Fig. 9 Molecular dynamics simulation of adatoms' alignment under applied strain

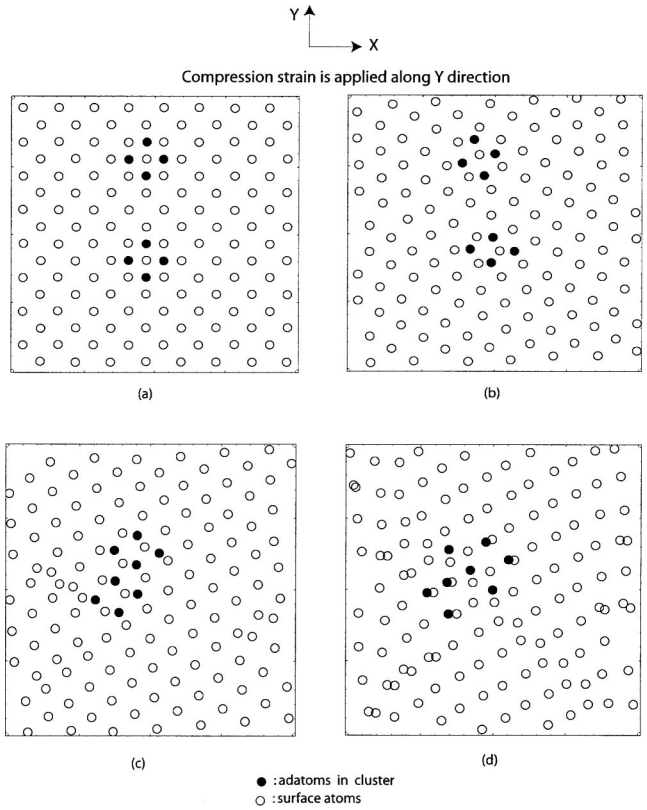


Fig. 10 Molecular dynamics simulation illustrating the growth of a cluster

tion of applied strain, angular orientation, and separation distance. The far- and near-field behaviors have been compared successfully. Results indicate that in most cases, the differences are not significant, at least when only elastic effects are considered.

According to the near-field evaluation, defect interaction is influenced by a number of factors including the applied strain, separation, relative orientation between defects and defect properties. Increasing applied strain can either strengthen or weaken this interaction; it depends on how the initial orientation of the two defects with respect to the applied strain. The study also suggests that the interaction decays rapidly with the separation distance between the defects. An important conclusion emanating from the analysis is that the nature of the interaction can be changed by increasing strain, e.g., increased strain can switch the character of adatom-vacancy interaction from repulsion to attraction for the examined Si(111). Even though the level of strain for this change is high for Si, it is clear that such a general result will quite possibly apply to other systems, not yet examined, for far lower levels of strain. The phenomenon is not only due to the adatom or vacancy properties but also depends on the elastic properties of the substrate material.

The strong strain dependence of the defect interaction shown in this investigation is important in that it could provide an inexpensive mechanism for controlling surface growth. Consequently, the application of appropriate strain levels can lead to the controlled formation of atomic clusters of a larger scale such as quantum wires and dots as well as other desired surface structures. The possible applications, particularly in the electronics industry, are of major significance, given the typical methods currently used like the expensive lithography and the typically uncontrolled self-organization of certain nanostructures.

This work is being extended in order to examine the behavior of surface defects of larger size. In principle, the analysis of larger defect clusters will involve additional mathematical complexities but should not dramatically alter the physics of the problem. So far, the results provided by atomic simulations agree well with the continuum analysis. We expect that the study of larger scale surface defects using both approaches will provide additional light to the question of how the applied strain affects the assembly process of growing nanostructures on surfaces.

Acknowledgments

R.K. acknowledges the support of NSF Grant No. CMS-0134123. D.K. and P.L. are grateful for the support provided by the NSF Grant Nos. CMS-0196437 and DMR-0090079.

Appendix

Equation (17) provides the strain evaluated at the site of one defect, induced by the other. For the most general case where the two defects are different, as in an adatom versus a vacancy, six unknowns need to be determined for a surface since each strain has three components. The equation for solving the induced strain takes the following general form:

$$\begin{bmatrix} I_3 & U^{(I)} \\ U^{(II)} & I_3 \end{bmatrix} \begin{bmatrix} \varepsilon^{*II(I)} \\ \varepsilon^{*I(II)} \end{bmatrix} = \begin{bmatrix} c^{II} \\ c^{I} \end{bmatrix} \tag{A1}$$

where I_3 is the 3×3 identity matrix, $U^{(I)}$ and $U^{(II)}$ are matrices related to the dipole properties of defect I and defect II and the Green functions. $\varepsilon^{*II(I)}$ is the induced strain at the site of defect II due to defect I and $\varepsilon^{*I(II)}$ is the induced strain at the site of defect I due to defect II; each contains three surface components. C^{II} and C^{I} are constant vectors determined by the dipole properties and Green functions evaluated at the defect sites.

With the surface represented by the *Y-Z* plane, the matrices $U^{(I)}$ and $U^{(II)}$ can be expressed through the isotropic dipole properties H and F as follows:

$$U^{(k_1)} = H^{(k_1)} \left[\begin{matrix} G_{yy,yy}+G_{yz,yz} & G_{yy,yy}+G_{yz,yz} & 0 \\ G_{zz,zz}+G_{yz,yz} & G_{zz,zz}+G_{yz,yz} & 0 \\ \frac{1}{2}(G_{yy,yz}+G_{zz,yz}+G_{yz,yy}+G_{yz,zz}) & \frac{1}{2}(G_{yy,yz}+G_{zz,yz}+G_{yz,yy}+G_{yz,zz}) & 0 \end{matrix} \right]_{x=x^{k_2}(t)} + F^{(k_1)}$$
$$\times \left[\begin{matrix} 2G_{yy,yy} & 2G_{yz,yz} & 2(G_{yy,yz}+G_{yz,yy}) \\ 2G_{yz,yz} & 2G_{zz,zz} & 2(G_{zz,yz}+G_{yz,zz}) \\ G_{yy,yz}+G_{yz,yy} & G_{zz,yz}+G_{yz,zz} & G_{yy,zz}+2G_{yz,yz}+G_{zz,yy} \end{matrix} \right]_{x=x^{k_2}(t)} \tag{A2}$$

The superscript k_1 and k_2 vary from I to II in order to identify the different properties of the two defects. All the derivatives of the Green function are evaluated at the defect site where induced

strain is calculated. The three components of the constant vector are given by

$$c_1^{(k_1)} = [D^{(k_1)} + H^{(k_1)}\varepsilon_{\alpha\alpha}^R]\,(G_{yy,yy} + G_{yz,yz})|_{x=x^{k_2(t)}} + 2F^{(k_1)} \times[\varepsilon_{yy}^R\, G_{yy,yy}|_{x=x^{k_2(t)}} + \varepsilon_{zz}^R\, G_{yz,yz}|_{x=x^{k_2(t)}} + \varepsilon_{yz}^R(G_{yy,yz} + G_{yz,yy})|_{x=x^{k_2(t)}} \quad \text{(A3a)}$$

$$c_2^{(k_1)} = [D^{(k_1)} + H^{(k_1)}\varepsilon_{\alpha\alpha}^R]\,(G_{zz,zz} + G_{yz,yz})|_{x=x^{k_2(t)}} + 2F^{(k_1)} \times[\varepsilon_{yy}^R\, G_{yz,yz}|_{x=x^{k_2(t)}} + \varepsilon_{zz}^R\, G_{zz,zz}|_{x=x^{k_2(t)}} + \varepsilon_{yz}^R\,(G_{zz,yz} + G_{yz,zz})|_{x=x^{k_2(t)}}] \quad \text{(A3b)}$$

$$c_3^{(k_1)} = \tfrac{1}{2}[D^{(k_1)} + H^{(k_1)}\varepsilon_{\alpha\alpha}^R](G_{yy,yz} + G_{yz,yy} + G_{yz,zz} + G_{zz,yz})_{x=x^{k_2(t)}} + F^{(k_1)}[\varepsilon_{yy}^R(G_{yz,yz} + G_{yz,yy})|_{x=x^{k_2(t)}} + \varepsilon_{zz}^R(G_{yz,zz} + G_{zz,yz})_{x=x^{k_2(t)}} + \varepsilon_{yz}^R(G_{yy,zz} + 2G_{yz,yz} + G_{zz,yy})_{x=x^{k_2(t)}} \quad \text{(A3c)}$$

The numerical results following the solution of Eq. (A1) for two different types of defects were discussed in Sec. 3. If the two defects are of the same type and have exactly the same dipole properties, it can be assumed that they will induce the same strain on each other. The reciprocity will simplify the 6×6 Eq. (A1) to the following 3×3 equation

$$[I_3 + U][\varepsilon] = [c] \quad \text{(A4)}$$

where the matrix U has the same format as in (A2) and the constant vector $[c]$ as in (A3). The solution of this equation has been obtained in terms of the long-range solution $\varepsilon_{\alpha\beta}^{*(0)}$, namely

$$\varepsilon_{\alpha\beta}^{*} = \frac{1}{\omega}[\varepsilon_{\alpha\beta}^{*(0)} + Q_{\alpha\beta\delta\gamma}\varepsilon_{\delta\gamma}^{*(0)}] \quad \text{(A5)}$$

The parameters ω and $Q_{\alpha\beta\delta\gamma}$ are determined by atomic properties and the Green function. Using the following notation:

$$q_{11} = H\,(G_{yy,yy} + G_{yz,yz})|_{x=x^{k_2(t)}} + 2FG_{yy,yy}|_{x=x^{k_2(t)}} \quad \text{(A6a)}$$

$$q_{12} = H(G_{yy,yy} + G_{yz,zy})_{x=x^{k_2(t)}} + 2FG_{yz,yz}|_{x=x^{k_2(t)}} \quad \text{(A6b)}$$

$$q_{13} = 2F\,(G_{yy,zy} + G_{yz,yy})|_{x=x^{k_2(t)}} \quad \text{(A6c)}$$

$$q_{21} = H\,(G_{zz,zz} + G_{yz,yz})|_{x=x^{k_2(t)}} + 2FG_{yz,yz}|_{x=x^{k_2(t)}} \quad \text{(A6d)}$$

$$q_{22} = H\,(G_{yz,yz} + G_{zz,zz})|_{x=x^{k_2(t)}} + 2FG_{zz,zz}|_{x=x^{k_2(t)}} \quad \text{(A6e)}$$

$$q_{23} = 2F\,(G_{yz,zz} + G_{zz,yz})|_{x=x^{k_2(t)}} \quad \text{(A6f)}$$

$$q_{31} = \frac{1}{2}[H(G_{yy,yz} + G_{yz,yy} + G_{yz,zz} + G_{zz,yz})|_{x=x^{k_2(t)}} + 2F\,(G_{yy,yz} + G_{yz,yy})|_{x=x^{k_2(t)}} \quad \text{(A6g)}$$

$$q_{32} = \frac{1}{2}[H\,(G_{yy,yz} + G_{yz,yy} + G_{yz,zz} + G_{zz,yz})|_{x=x^{k_2(t)}} + 2F\,(G_{yz,zz} + G_{zz,yz})|_{x=x^{k_2(t)}}] \quad \text{(A6h)}$$

$$q_{33} = F\,(G_{yy,zz} + 2G_{yz,yz} + G_{zz,yy})|_{x=x^{k_2(t)}} \quad \text{(A6i)}$$

the parameters ω and $Q_{\alpha\beta\delta\gamma}$ are given by

$$\omega = 1 - q_{11} - q_{22} - q_{33} - q_{12}q_{21} + q_{11}q_{22} - q_{13}q_{31} - q_{23}q_{32} + q_{11}q_{33} + q_{22}q_{33} + q_{13}q_{22}q_{31} - q_{12}q_{23}q_{31} - q_{13}q_{21}q_{32} + q_{11}q_{23}q_{32} + q_{12}q_{21}q_{33} - q_{11}q_{22}q_{33} \quad \text{(A7)}$$

$$Q_{yyyy} = -\,(q_{22} + q_{33} + q_{23}q_{32} - q_{22}q_{33}) \quad \text{(A8a)}$$

$$Q_{yyzz} = q_{12} + q_{13}q_{32} - q_{12}q_{33} \quad \text{(A8b)}$$

$$Q_{yyyz} = q_{13} + q_{12}q_{23} - q_{13}q_{22} \quad \text{(A8c)}$$

$$Q_{zzyy} = q_{21} + q_{23}q_{31} - q_{21}q_{33} \quad \text{(A8d)}$$

$$Q_{zzyz} = q_{23} + q_{13}q_{21} - q_{11}q_{23} \quad \text{(A8e)}$$

$$Q_{yzyz} = -\,(q_{11} + q_{22} + q_{12}q_{21} - q_{11}q_{22}) \quad \text{(A8f)}$$

References

[1] Kallinteris, G. C., Evangelakis, G. A., and Papanicolaou, N. I., 1996, "Molecular Dynamics Study of the Vibrational and Transport Properties of Copper adatoms on the (111) Copper Surface; Comparison with the (001) Face," Surf. Sci., **369**, pp. 185–198.

[2] Evangelakis, G. A., and Papanicolaou, N. I., 1996, "Adatom Self-Diffusion Processes on (001) Copper Surface by Molecular Dynamics," Surf. Sci., **347**, pp. 376–386.

[3] Sabiryanov, R. F., et al., 2003, "Surface Diffusion and Growth of Patterned Nanostructures on Strained Surfaces," Phys. Rev. B, **67**, p. 125412.

[4] Larsson, M. I., 2001, "Kinetic Monte Carlo Simulations of Adatom Island Decay on Cu(111)," Phys. Rev. B, **64**, p. 115428.

[5] Larsson, M. I., Sabiryanov R. F., Cho, K., and Clemens, B. M., 2003, "Surface Strain Effects on Adatom Kinetics and Self-Assembly," Surf. Sci., **536**, pp. L389–L395.

[6] Lau, K. H., and Kohn, W., 1977, "Elastic Interaction of Two Atoms Adsorbed on A Solid Surface," Surf. Sci., **65**, pp. 607–618.

[7] Marchenko, V. I., and Parchin, A. Y., 1980, "Elastic Properties of Crystal Surface," Sov. Phys. JETP, **52**, pp. 129–131.

[8] Shilkrot, L. E., and Srolovitz, D. J., 1996, "Elastic Field of a Surface Step: Atomistic Simulations and Anisotropic Elastic Theory," Phys. Rev. B, **53**, p. 11120.

[9] Kouris, D., Peralta, A., and Sieradzki, K., 2000, "Surface Islands and Their Elastic Interaction with Adatoms," Surf. Sci., **445**, pp. 420–429.

[10] Peralta, A., Kouris, D., Knap, J., and Sieradzki, K., 1998, "Eigenstrains and the Elastic Field of an Adatom," J. Mech. Phys. Solids, **46**, pp. 1557–1579.

[11] Kukta, R. V., and Bhattacharya, K., 2002. "A Micromechanical Model of Surface Steps," J. Mech. Phys. Solids, **50**, pp. 615–649.

[12] Kukta, R. V., Peralta, A., and Kouris, D., 2002, "Elastic Interaction of Surface Steps: Effect of Atomic-Scale Roughness," Phys. Rev. Lett., **88**, p. 186102.

[13] Kukta, R. V., Kouris, D., and Sieradzki, K., 2003, "Adatoms and Their Relation to Surface Stress," J. Mech. Phys. Solids, **51**(7), pp. 1243–1266.

[14] Kukta, R. V., Liu, P., and Kouris, D., 2003, "On the Dependence of Adatom Interactions on Strain," J. Mech. Phys. Solids, **51**(12), pp. 2149–2167.

[15] Sander, D., Linke, U., and Ibach, H., 1992, "Adsorbate-Induced Surface Stress: Surface, Oxygen and Carbon on Ni(100)," Surf. Sci., **272**, pp. 318–325.

[16] Ibach, H., 1994, "Adsorbate-Induced Surface Stress," J. Vac. Sci. Technol. A **12**(4), pp. 2240–2245.

[17] Asaro, R. J., and Tiller, W. A., 1972, "Interface Morphology Development during Stress Corrosion Cracking—Part I. Via Surface Diffusion." Metall. Trans. **3**, pp. 1789–1796.

[18] Grinfeld, M. A., 1993, "The Stress Driven Instability in Elastic Crystals: Mathematical Models and Physical Manifestations," J. Nonlinear Sci. **3**, pp. 35–83.

[19] Freund, L. B., and Jonsdittir, F., 1993, "Instability of a Biaxially Stressed Thin Film on a Substrate Due to Material Diffusion Over Its Free Surface," J. Mech. Phys. Solids, **41**, pp. 1245–1263.

[20] Kukta, R. V., and Freund, L. B., 1997, "Minimum Energy Configuration of Epitaxial Material Clusters on a Lattice-Mismatched Substrate," J. Mech. Phys. Solids, **45**, pp. 1835–1860.

[21] Shilkrot, L. E., and Srolovitz, D. J., 1997, "Anisotropic Elastic Analysis and Atomistic Simulation of Adatom-Adatom Interactions on Solid Surfaces," J. Mech. Phys. Solids, **45**, pp. 1861–1873.

[22] Tersoff, J., 1988, "New Empirical Approach for the Structure and Energy of Covalent Systems," Phys. Rev. B, **37**, pp. 6991–7000.

[23] Johnson, R. A., 1989, "Alloy Models with the Embedded-Atom Method," Phys. Rev. B, **39**, pp. 12554–12559.

[24] Johnson, R. A., 1988, "Analytic Nearest-Neighbor Model for fcc Metals," Phys. Rev. B, **37**, pp. 3924–3931.

[25] Parrinello, M., and Rahman, A., 1981, "Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method," J. Appl. Phys., **52**, pp. 7182–7190.

[26] Parrinello, M., and Rahman, A., 1980, "Crystal Structure and Pair Potentials: A Molecular-Dynamics Study," Phys. Rev. Lett., **45**, pp. 1196–1199.

A Comparison of the Structural Response of Clamped and Simply Supported Sandwich Beams With Aluminium Faces and a Metal Foam Core

V. L. Tagarielli

N. A. Fleck
e-mail: naf1@eng.cam.ac.uk

Engineering Department,
University of Cambridge,
Trumpington Street,
Cambridge, CB2 1PZ, UK

Plastic collapse modes for clamped sandwich beams have been investigated experimentally and theoretically for the case of aluminium face sheets and a metal foam core. Three initial collapse mechanisms have been identified and explored with the aid of a collapse mechanism map. It is shown that the effect of clamped boundary conditions is to drive the deformation mechanism towards plastic stretching of the face sheets. Consequently, the ultimate strength and level of energy absorption of the sandwich beam are set by the face sheet ductility. Limit load analyses have been performed and simple analytical models have been developed in order to predict the postyield response of the sandwich beams; these predictions are validated by both experiments and finite elements simulations. It is shown experimentally that the ductility of aluminium face sheets is enhanced when the faces are bonded to a metal foam core. Finally, minimum weight configurations for clamped aluminium sandwich beams are obtained using the analytical formulas for sandwich strength, and the optimal designs are compared with those for sandwich beams with composite faces and a polymer foam core. [DOI: 10.1115/1.1875432]

1 Introduction

A large amount of research has been conducted recently on the mechanical performance of sandwich structures, stimulated by the development of stiff and strong, lightweight core materials [1–3]. For example, Chen et al. [4] and Bart-Smith et al. [5] have explored the quasi-static behavior of simply supported aluminium sandwich beams in three-point bending. The competing collapse modes of core shear, face yield, and indentation were observed, and the sensitivity of the collapse strength to geometry and to material properties was determined. However, there has been little prior attention paid to the effect of the support condition upon the collapse mechanism. Sandwich panels are often clamped to a stiff and strong support framework (e.g., a ship hull), and this can be represented in the laboratory by a fully clamped end condition.

In the present study, the response of sandwich beams comprising aluminium face sheets and an aluminium alloy foam core is explored for both simply supported and fully clamped boundary conditions. Potential modes of initial collapse are identified, and simple analytical models are stated. A mechanism map for initial collapse is generated from these formulas in order to relate the governing collapse mechanism of clamped beams to their geometry and material properties. Three sandwich geometries are selected from the collapse map, with each one lying in a different regime. Sandwich specimens with these geometries are manufactured and tested with both simply supported and clamped end conditions. The operative collapse mechanisms and measured load versus deflection curves are compared with both analytical predictions and finite element simulations. The analytical formulas for initial collapse are then used to determine minimum weight designs for clamped sandwich beams as a function of an appropriate structural load index. These minimum weight configurations are compared with minimum weight designs for clamped sandwich beams with composite face sheets and polymer foam cores. The study concludes with a short experimental study on the degree to which the foam core stabilizes the faces against necking.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, March 5, 2004; final revision, September 18, 2004. Editor: R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Professor Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Analytical Models for the Collapse Response

We begin by summarizing analytical formulas for the elastic stiffness, initial collapse load, and postyield behavior of sandwich beams, assuming that both face sheets and core can be considered as elastic—perfectly plastic materials, and the beams are either simply supported or fully clamped. The analytical formulas are used to construct collapse mechanism maps, and to enable the design of specimen geometries so that a variety of failure modes are activated.

Consider a sandwich beam of length ℓ and uniform width b, comprising two identical face-sheets of thickness t, bonded to a metal foam core of thickness c, as shown in Fig. 1. A flat-bottomed punch of width a is used to load the beam transversely at midspan by a force F and corresponding deflection u. The outer supports react with two vertical forces $F/2$ in the simply supported case plus bending moments M and in-plane horizontal forces P in the clamped case. When the beam is simply supported its length exceeds the span ℓ by an overhang H at each end.

A suffix f-denotes the face sheet, while the suffix c denotes the core; we introduce the symbols E_f, σ_f, ν_f, ρ_f, and E_c, σ_c, ν_c, ρ_c to denote the Young's modulus, yield strength, Poisson ratio, and density of the faces and core, respectively. It is useful to nondimensionalize the geometrical and material parameters according to the following definitions:

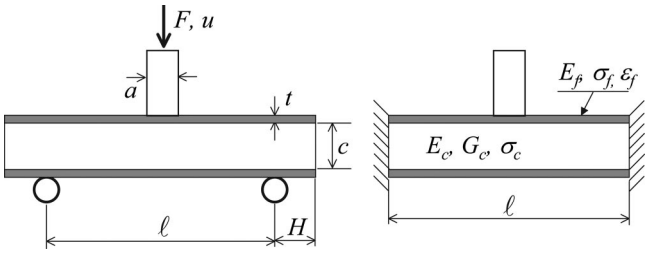


Fig. 1 Geometries of simply supported and clamped sandwich beams transversely loaded by a flat punch

$$\bar{c} = \frac{c}{\ell};\ \bar{t} = \frac{t}{c};\ \bar{a} = \frac{a}{\ell};\ \bar{u} = \frac{u}{\ell};\ \bar{\sigma} = \frac{\sigma_c}{\sigma_f};\ \bar{\rho} = \frac{\rho_c}{\rho_f} \tag{1}$$

Furthermore, we define the following nondimensional indices for the load F, energy absorption W, and mass M as:

$$\bar{F} = \frac{F}{b\ell\sigma_f};\ \bar{W} = \frac{W}{b\ell^2\sigma_f};\ \bar{M} = \frac{M}{b\ell^2\rho_f} = (2\bar{t} + \bar{\rho})\bar{c} \tag{2}$$

2.1 Elastic Regime. Elastic theory for sandwich beams is well established [6], and the transverse deflection u at midspan of the beam is

$$u = \frac{F\ell^3}{48EI_{\text{eq}}} + \frac{F\ell}{4AG_{\text{eq}}} \tag{3}$$

in the simply supported case, and

$$u = \frac{F\ell^3}{384EI_{\text{eq}}} + \frac{F\ell}{4AG_{\text{eq}}} \tag{4}$$

in the fully clamped case. The equivalent flexural and shear rigidities are given by

$$EI_{\text{eq}} = \frac{E_f btd^2}{2} + \frac{E_f bt^3}{6} + \frac{E_c bc^3}{12} \approx \frac{E_f btd^2}{2}$$
$$AG_{\text{eq}} = \frac{bd^2}{c}G_c \approx bcG_c \tag{5}$$

where G_c is the shear modulus of the core and $d = c + t$. The flexural and shear terms have comparable magnitudes for the sandwich beams considered later, and so it is necessary to include both.

2.2 Mechanisms of Initial Collapse. Consider the response of an elastic-ideally plastic sandwich beam, with an end condition of either fully clamped or simply supported. As the applied load is increased a limit load is attained, corresponding to initial plastic collapse. For the case of a clamped beam, membrane effects become significant with continued deformation beyond initial collapse, and a subsequent hardening behavior is observed.

The initial limit load for initial plastic collapse is calculated for a number of trial collapse mechanisms using the upper bound theory of plasticity. The face sheets and core are taken to be rigid, ideally plastic with uniaxial strength σ_f for the faces and σ_c for the core. Ashby et al. [1] have identified the competing collapse modes for sandwich beams with metallic face sheets and cores as *face yield*, *core shear*, and *indentation*. We calculate collapse loads for each of these mechanisms, for both simply supported and clamped boundary conditions, and since the transverse deflections are small, we neglect membrane effects.

In the current study only plastic collapse mechanisms are considered. Alternative failure modes are expected when the face sheets or core are made from elastic-brittle solids such as ceramics or ceramic-fiber composites.

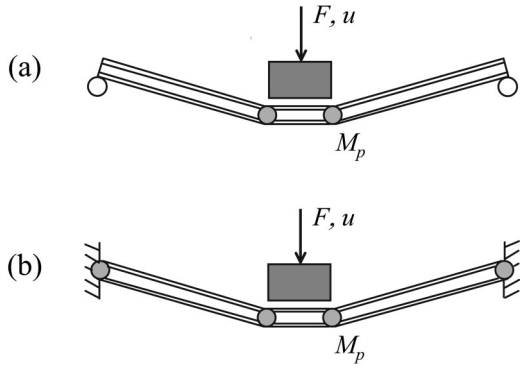


Fig. 2 Initial collapse by face yielding of sandwich beams (*a*) simply supported case and (*b*) built-in case

Face yield. Consider the plastic collapse of a simply supported sandwich beam, with the collapse mechanism given by rotation about plastic hinges adjacent to the central punch, as sketched in Fig. 2(*a*). The plastic bending moment for the beam is given by

$$M_p = dtb\sigma_f + \frac{c^2}{4}b\sigma_c \tag{6}$$

A straightforward work calculation gives the plastic limit load F_{FYS} for face yield of the simply supported beam as

$$F_{\text{FYS}} = \frac{4bt(c+t)}{\ell - a}\sigma_f + \frac{bc^2}{\ell - a}\sigma_c \tag{7}$$

which can be re-expressed in nondimensional form as

$$\bar{F}_{\text{FYS}} = \frac{F_{\text{FYS}}}{b\ell\sigma_f} = \frac{\bar{c}^2}{1 - \bar{a}}[4\bar{t}(1 + \bar{t}) + \bar{\sigma}] \tag{8}$$

The same result can be obtained by considering equilibrium and yield, via the lower bound theorem, but this is not detailed here. Consequently, this formula is exact within the context of rigid, ideally plastic beam theory.

A closely related result follows for the clamped sandwich beam. Now, however, four plastic hinges exist, two at the punch and one at each support. The collapse load is twice that for the simply supported beam, and is given in nondimensional form as

$$\bar{F}_{\text{FYC}} = \frac{F_{\text{FYC}}}{b\ell\sigma_f} = \frac{2\bar{c}^2}{1 - \bar{a}}[4\bar{t}(1 + \bar{t}) + \bar{\sigma}] \tag{9}$$

for face yield of the clamped beam.

Core shear. The transverse shear force on a sandwich beam is carried mainly by the core, and plastic collapse by core shear can result. Consider first the case of a simply supported sandwich beam with an overhang H beyond the outer rollers, as shown in Fig. 1. Two competing collapse mechanisms can be identified. Mode A entails plastic shear of the core and rotation about plastic hinges in the face sheets at the central punch, see Fig. 3(*a*), note that the sandwich beam shears beyond the outer supports. Alternatively, in mode B, the sandwich beam does not shear beyond the outer supports but this necessitates the formation of additional plastic hinges in the face sheets at the outer supports, see Fig. 3(*b*). Simple work calculations give the collapse loads for modes A and B, respectively, as

$$F_A = 2\frac{bt^2}{\ell - a}\sigma_f + 2bc\tau_c\left(1 + \frac{H}{\ell - a}\right) \tag{10}$$

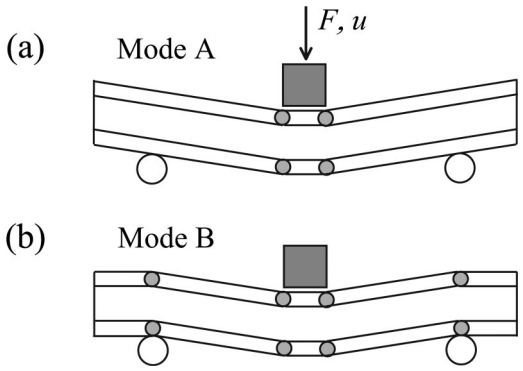


Fig. 3 Two alternative modes of initial collapse by core shear

$$F_B = 4\frac{bt^2}{\ell - a}\sigma_f + 2bc\tau_c \tag{11}$$

A comparison of these formulas confirms that mode A is more likely to occur at short overhangs; Chen et al. [4] have found the characteristic overhang H_t associated with transition from mode A to mode B

$$H_t = \frac{t^2\sigma_f}{2c\tau_c} \tag{12}$$

in which the shear strength of the core τ_c can be taken as $\tau_c \approx 2\sigma_c/3$.

For the case of clamped beams the only possible collapse mechanism is mode B, with the associated collapse load given by Eq. (11). In the present study we consider simply supported beams with an overhang length H exceeding the transition value H_t, so that the collapse mechanism is again mode B. The initial collapse load is insensitive to the boundary condition, and is given by the nondimensional form of (11), as

$$\overline{F}_{\rm CS} = \frac{F_B}{b\ell\sigma_f} = 4\overline{c}\left(\frac{\overline{ct}^2}{1-\overline{a}} + \frac{\overline{\sigma}}{3}\right) \tag{13}$$

Indentation. An alternative collapse mode is plastic indentation of the upper face sheet beneath the central punch, as sketched in Fig. 4. Again, a simple analytical formula can be obtained for the plastic collapse load using an upper bound approach, see Ashby et al. [1] and Bart-Smith et al. [5]. The mode involves plastic crushing of the core over a length of $(2\lambda + a)$ and the formation of four plastic hinges in the upper face sheet. The spacing λ between the hinges is obtained by minimizing the upper bound collapse load. For both the simply supported and clamped beams, the nondimensional indentation load is

$$\overline{F}_{\rm IN} = \frac{F_{\rm IN}}{b\ell\sigma_f} = 2\overline{tc}\sqrt{\overline{\sigma}} + \overline{a}\overline{\sigma}; \quad \lambda = t\sqrt{\frac{\sigma_f}{\sigma_c}} \tag{14}$$

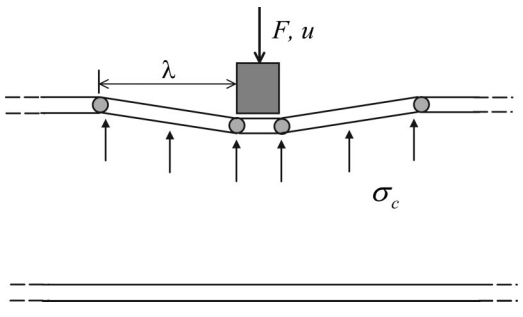


Fig. 4 Initial collapse of sandwich beams by indentation of the upper face sheet

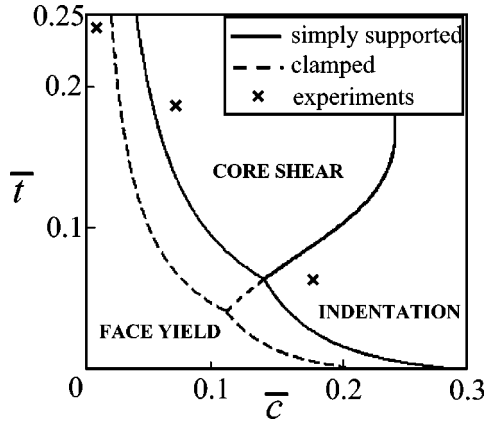


Fig. 5 Initial collapse mechanism map for simply supported and clamped sandwich beams in three-point bending. $\overline{\sigma}=0.034$ and $\overline{a}=0.1$. Test geometries are marked on the map.

2.3 Mechanism Maps for Initial Collapse. The observed initial collapse mechanism for a sandwich beam is the one associated with the lowest collapse load for a given geometry and material properties. The active modes can be shown graphically by plotting a nondimensional measure of the upper bound collapse load $\overline{F} = F/(b\ell\sigma_f)$ on a diagram with the nondimensional axes $\overline{c}$ and $\overline{t}$, for selected values of $\overline{\sigma}$ and $\overline{a}$. This method follows that pioneered by Gibson and Ashby [7] for polymeric foam cores and aluminium alloy face sheets.

A collapse mechanism map, for both simply supported and clamped beams, is given in Fig. 5, for the choice $\overline{\sigma}=0.034$ and $\overline{a}=0.1$, and the map is representative of the materials used in this study. It is assumed that the overhang H for the simply supported case exceeds the transition value H_t so that core shear mechanism is mode B. The regimes of dominance for each collapse mechanism are marked, and the three data points marked on the figure give the three structural geometries tested and analyzed later.

Note that the maps for simply supported and fully clamped coincide along the indentation—core shear boundary, since only the face yield collapse load changes when we switch from the simply supported to the clamped boundary condition. The regime of face yielding is significantly larger for the simply supported beam than for the fully clamped beam.

2.4 Finite Deflection of Clamped Sandwich Beams. It is shown experimentally and theoretically later that simply supported beams undergo continued plastic collapse at nearly constant load; eventually, the transverse deflection becomes sufficiently large that the structure fails by fracture of the face sheets or core. In contrast, clamped beams undergo membrane stretching of the face sheets beyond initial yield, and this gives rise to a hardening macroscopic response. We now analyze the postyield response of clamped sandwich beams.

Initial plastic collapse of clamped sandwich beams occurs by face yield, core shear, or indentation at small transverse deflections. Subsequent transverse deflection, however, involves tensile stretching of the faces and core. The stress distribution within the beam evolves from that associated with the initial collapse load to that of pure membrane action, with the membrane solution achieved when the deflection is about equal to the thickness of the beam $H_S = C + 2t$. Thereafter, the beam deforms in a membrane mode, and yields axially until the face sheets tear when the axial plastic strain attains the material ductility. Equilibrium considerations give an expression for the load versus deflection response in the membrane phase as

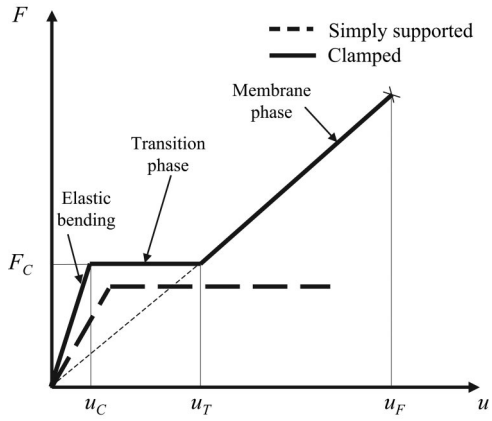


Fig. 6 Stages of collapse of simply supported and clamped sandwich beams

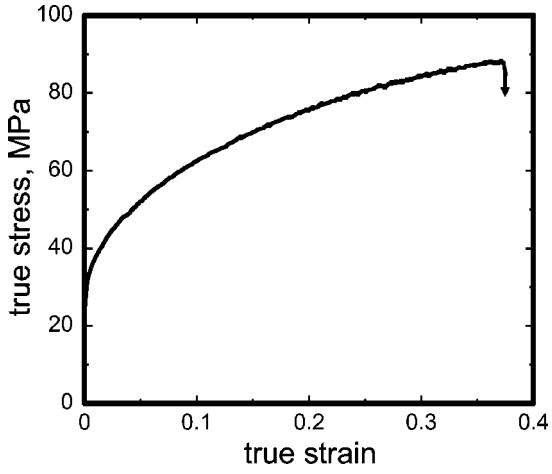


Fig. 7 Tensile response of the annealed aluminium face sheets

$$F(u) = \frac{8tb\sigma_f}{\ell - a}u \tag{15}$$

assuming that the deflection u is small compared with the span ℓ, and that the net axial force in the faces is much greater than that in the core.

It is difficult to obtain a general failure criterion for the beam, since the plastic strain distribution within the sandwich structure depends upon both the initial collapse mechanism and the membrane stretching phase of deformation. Here, we state a simple failure criterion based on an estimate of the strain in the face sheets due to stretching of the beam, and neglect the plastic strains due to bending. For an assumed ductility ε_F of the face sheet material, the deflection u_F at failure is given by

$$u_F = \ell\sqrt{\tfrac{1}{2}(1-\bar{a})\varepsilon_F} \tag{16}$$

2.5 Summary of Clamped Beam Response. The load versus deflection response of clamped beams may be subdivided into three phases, as sketched in Fig. 6

(1) *Elastic bending*. The beam deflects elastically until the applied load attains the initial collapse load F_C associated with the operative collapse mechanism. The load F_C is reached at an elastic deflection u_C as dictated by Eq. (4).
(2) *Plateau phase*. Once initial collapse has been attained, it is assumed that the load remains constant under increasing transverse deflection up to a transverse deflection u_T, at which the load predicted by (15) equals the initial collapse load.
(3) *Membrane phase*. The beam stretches in the manner of a plastic string and the load versus deflection response is given by Eq. (15). The sandwich beam deflects until there is a sudden loss of load carrying capacity due to face sheet tearing when the deflection attains the value u_F.

The energy absorption W is the area under the load versus deflection curve of the sandwich beam. Upon neglecting the elastic contribution to energy absorption, the nondimensional measure $\bar{W} = W/b\ell^2\sigma_f$ for a clamped beam, is taken as

$$\bar{W} = \bar{F}_c\bar{u}_T + \frac{4\bar{t}\bar{c}}{1-\bar{a}}(\bar{u}_F^2 - \bar{u}_T^2) \tag{17}$$

where

$$\bar{u}_T = \frac{u_T}{\ell}; \quad \bar{u}_F = \frac{u_F}{\ell} \tag{18}$$

3 Materials Characterization and Test Technique

Three structural geometries have been selected within the failure map of Fig. 5, with each geometry lying in a different region of the map. The sandwich beams (of width about 50 mm) were manufactured by bonding aluminium face sheets to aluminium alloy foam cores, and were subsequently tested in three-point bending. A commercially pure fully annealed aluminium sheet was used to manufacture the faces, whereas the foam core was a closed-cell aluminium-alloy foam, with trade-name Alporas[1]; its relative density (density of the foam divided by the density of the cell wall material) was $\hat{\rho}=11\%$, and the average cell size was 3 mm. Annealed aluminium was used to ensure that the clamped specimens did not fail in the transition phase, in order to observe the membrane regime.

The aluminium face sheets were degreased and abraded, and were then adhered to the foam core using Redux 322 epoxy adhesive on a nylon carrier mesh. The sandwich beams were air-cured at 180 °C for 1 h, and bonding was facilitated by imposing a dead load with a nominal contact pressure of 0.01 MPa. The shear strength of the cured Redux 322 adhesive was taken to be 20 MPa, from Hexcel's data sheets: this strength is about one order of magnitude higher than that of the Alporas foam, and so no adhesive failure was observed.

3.1 Face Sheet Material. The mechanical properties of the annealed aluminium face sheets material were measured as follows. Tensile specimens of dog-bone geometry were cut from the aluminium face sheets. The tensile tests were performed in a servo hydraulic test machine at a strain rate of 10-4/s; the axial strain was measured using both strain gauges and a laser extensometer, while the transverse strain was measured with a strain gauge.

The measured true stress versus true strain response is given in Fig. 7. The Young's modulus is $E_f=70$ GPa, and the Poisson ratio is $\nu_f=0.33$. The annealed aluminium has a 0.2% offset yield strength of 30 MPa, an ultimate tensile strength of 85 MPa and an elongation to failure of about 40%.

3.2 Core Material. The tensile, compressive, and shear stress versus strain response has been already reported by Chen et al. [4]. In brief, the Young's modulus of the Alporas foam is $E_c=1.06$ GPa, and the compressive and tensile yield strength is $\sigma_c=2.1$ MPa, with a tensile ductility of 1.1%.

3.3 Test Method for Sandwich Beams. The sandwich beams

[1]European supplier, Karl Bula, Innovation Services, Ch-5200 Brugg, Herrenmatt 7F, Switzerland.

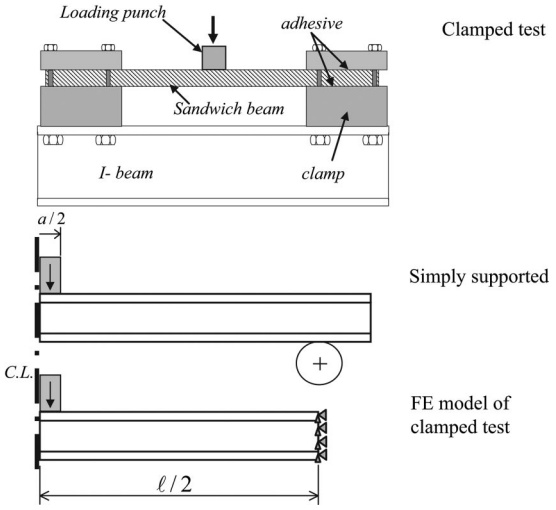


Fig. 8 The loading configurations, with boundary conditions used in the finite element calculations

were loaded in three point bending using a fully clamped rig and a simply supported rig, as sketched in Fig. 8. Selected specimens were instrumented in order to confirm the mechanism of collapse. Laser extensometers were used to measure the deflection and the change in height of beam directly under the indenter, and 120Ω resistance strain gauges of length 2 mm were placed at midspan on the bottom face sheet. A clip gauge was used to measure the relative sliding displacement of the face sheets, and thereby the average shear strain in the core.

The sandwich beams were loaded at a constant speed of 0.3 mm/s by flat indenters of width 0 (roller) to 18 mm. Fixed rollers of diameter 19 mm were used in the simply supported tests, while a stiff steel rig, bolted to an underlying I-beam, was used in the fully clamped tests to restrain the specimens against end displacement and rotation.

4 Effect of Boundary Conditions on Collapse Response

In order to investigate the effect of boundary conditions on the response of sandwich beams, three geometries of specimen have been manufactured and tested in the simply supported and clamped conditions. The geometries are summarised in Table 1. For each geometry, we compare the measured load versus deflection response of the clamped and simply supported beams.

4.1 Face Yield Specimens. Consider first the measured collapse response of beams undergoing face yield, see Fig. 9(*a*). The two beams initially collapse at different load levels; as predicted by Eqs. (8) and (9), the collapse load for the clamped beam is about twice that for the simply supported beam. After initial collapse, the simply supported beam deflects at almost constant load; it fails by tearing of the bottom face at midspan when the tensile plastic strain has attained the material ductility. The clamped beam first undergoes face yield; then, at deflections exceeding the thickness of the sandwich beam, the deformation mode switches to plastic stretching of the faces and core. This stretching phase is characterized by a steeply rising linear load versus deflection response. Both the simply supported and clamped tests were arrested prior to tensile tearing of the face sheets. Figure 9(*a*) includes photographs of two duplicate specimens tested under different boundary conditions. The extent of deflection of these duplicate specimens is labeled on the collapse responses. The two different modes of collapse at large deflections are evident.

Table 1 Geometry of sandwich beam specimens

No.	t(mm)	c(mm)	l(mm)	a(mm)
1 (FY)	0.8	3	200	0 (roller)
2 (CS)	0.8	4	70	18
3 (IN)	0.8	15	100	3.5

4.2 Core Shear Specimens. Figure 9(*b*) gives results for the sandwich beams initially collapsing by core shear. Again, the tests were not taken to final failure and again photographs are shown of two duplicate specimens. The degree of deflection of these duplicate specimens is labeled on the load versus deflection curves to aid their interpretation.

The simply supported beam was given a very large overhang in order to inhibit collapse by core shear mode A. With this choice, the initial collapse mechanism (and therefore the initial collapse load) is identical for the clamped and simply supported cases. The measured responses confirm this prediction, see Fig. 9(*b*). Now consider the collapse responses beyond initial yield. The load carried by the simply supported beam increases slightly to a peak value at a large transverse deflection of 8 mm. The peak in the load versus deflection curve is due to shear fracture of the foam core.

In contrast, the clamped beam undergoes axial stretching of the faces beyond initial collapse and the load steeply rises above the initial collapse strength, as suggested by Eq. (15). After a transition phase, of up to $u \approx H_S$, the load rises almost linearly with deflection; this supports the assertion of the analytical model that the specimen is in a pure membrane state.

Visual observations during the tests on the clamped and simply supported beams revealed that inclined shear cracks developed within the core once the core had sheared by a few percent. This is consistent with the fact that the Alporas foam has a shear ductility of 2%, see Chen et al. [4].

4.3 Indentation Specimens. The load versus deflection responses of the indentation geometry are given in Fig. 9(*c*), together with photographs of the as-tested specimens. It is clear from the images that the specimens are squat in shape and collapsed by indentation. Visual observations during each test confirmed that initial collapse was by indentation beneath the central punch. The initial collapse load of the clamped beam is approximately 20% greater than that of the simply supported beam, while the analytical predictions for the rigid, ideally plastic case give an identical yield load for both grip conditions. A possible explanation is that the bending moment at midspan for the clamped case is only half that for the simply supported case, at any given load. Therefore, the higher bending moments in the simply supported beam give rise to higher compressive stresses within the upper face sheet, and this facilitates the indentation mechanism.

Now consider the finite collapse response of the beams subsequent to the initial collapse. For the simply supported specimen, the separation of the faces diminishes with increasing transverse deflection, and so the plastic collapse moment (and consequently the applied load) drops. Finally, the bottom face tears at midspan.

In the clamped beam test the continued activation of the indentation mechanism is inhibited by the development of membrane tension within the faces. At sufficiently large transverse deflections the stress state again approaches the pure membrane state.

Figure 9(*c*) includes photographs of the as-tested specimens. Although the total transverse deflection is very similar in the two specimens, the degree of core crushing in the clamped beam is much less than that observed in the simply supported beam. This is consistent with the fact that tensile membrane stresses within the indented face of the fully clamped specimen have stabilized it against indentation.

5 Numerical Simulation of Beams Response

The three-point bending response of simply supported and clamped sandwich beams has been modelled with the commercial

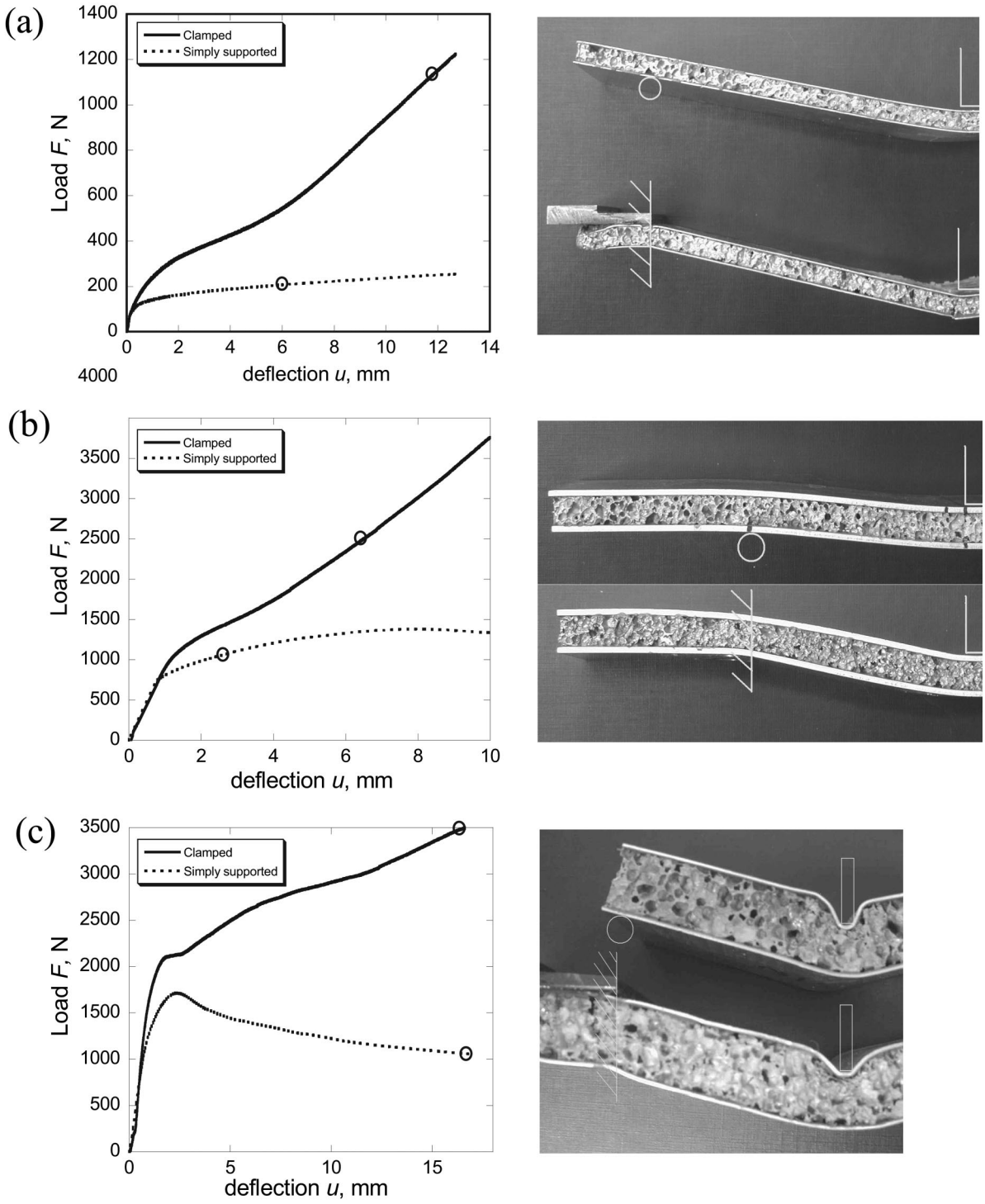


Fig. 9 Measured load vs deflection response and photographs of simply supported and clamped sandwich beams. Initial collapse is by (*a*) face yield, (*b*) core shear, and (*c*) indentation

finite elements code ABAQUS in order to compare it with analytical predictions and experiments. Due to symmetry, only half the length of the sandwich structure has been modeled. Eight-noded two-dimensional rectangular elements, with full integration, have been used to discretize the sandwich core and the aluminium skins. Typically, each face sheet has three elements in the thickness direction and 200 elements along the semi-span, while the core is twenty elements deep by 200 elements along the semi-span.

Loading by the frictionless flat punch is modeled by prescribing a uniform vertical displacement to the appropriate boundary nodes of the upper face sheet, as sketched in Fig. 8. In the simply supported case, contact between the beam and the rollers is modeled by the contact surfaces provided by ABAQUS. In the clamped case, both the vertical and horizontal displacements of nodes along the ends of the beam are constrained to vanish. This boundary condition is somewhat stiffer than the actual clamped condition used in the experimental investigation, see Fig. 8. A preliminary mesh sensitivity study has been performed to ensure an accurate representation of the sandwich specimen.

In the finite element model, the aluminium skins are modeled by the J_2 flow theory of plasticity, and the foam is described by the metal foam constitutive model of Deshpande and Fleck [8], as implemented in ABAQUS by Chen [9]. In this model the yield function Φ is assumed to be

$$\Phi = \hat{\sigma} - Y = 0 \tag{19}$$

where Y is the uniaxial yield strength and $\hat{\sigma}$ is the effective stress, defined by

$$\hat{\sigma}^2 = \frac{1}{1 + (\alpha/3)^2}(\sigma_e^2 + \alpha^2 \sigma_m^2) \tag{20}$$

where α defines the aspect ratio of the elliptical yield surface in the Mises stress σ_e and mean stress σ_m space. For the case $\alpha = 0$, the effective stress $\hat{\sigma}$ reduces to σ_e and the J_2 flow theory is recovered. For simplicity, isotropic hardening is assumed, i.e., the yield surface grows in a geometrically self-similar manner with strain. To model the postyield behavior, an effective plastic strain rate $\dot{\hat{\varepsilon}}$, the work rate conjugate to $\hat{\sigma}$, is introduced as

$$\dot{\hat{\varepsilon}}^2 = [1 + (\alpha/3)^2](\dot{\varepsilon}_e^2 + \dot{\varepsilon}_m^2/\alpha^2)$$

$$\dot{\varepsilon}_e^2 = (2/3)\dot{\varepsilon}_{ij}^p \dot{\varepsilon}_{ij}^p, \quad \dot{\varepsilon}_m = \dot{\varepsilon}_{ii}^p \tag{21}$$

where $\dot{\varepsilon}_{ij}^p$ is the plastic strain rate, $i, j = 1,2,3$, and the convention of summation over repeated indices applies. With the assumption of normality, the plastic strain rate is given by

$$\dot{\varepsilon}_{ij}^p = \dot{\hat{\varepsilon}} \frac{\partial \Phi}{\partial \sigma_{ij}} = \frac{\dot{\hat{\varepsilon}}}{1 + (\alpha/3)^2}\left(\frac{3}{2}\frac{s_{ij}}{\hat{\sigma}} + \frac{\alpha^2}{3}\delta_{ij}\frac{\sigma_m}{\hat{\sigma}}\right) \tag{22}$$

where $s_{ij} = \sigma_{ij} - \sigma_m \delta_{ij}$ is the deviatoric stress, δ_{ij} is the Kronecker delta, and the effective strain rate is connected to the effective stress rate by

$$\dot{\hat{\varepsilon}} = \frac{\dot{\hat{\sigma}}}{H(\hat{\sigma})} \tag{23}$$

Here, $H(\hat{\sigma})$ is the tangent of the uniaxial true stress versus logarithmic plastic strain curve at stress level $\sigma = \hat{\sigma}$. The constitutive models for both the aluminium faces and the foam core were calibrated against measured uniaxial data.

6 Comparison of Experiments and Predictions

It is instructive to compare the analytical predictions of the elastic-plastic collapse response with detailed finite element analysis for the three clamped beam geometries as detailed in Table 1 and shown in Fig. 5. A similar comparison has already been presented by Chen et al. [4] for simply supported aluminium sandwich beams, where excellent agreement is demonstrated.

Figure 10 shows the measured and predicted load versus deflection response for a specimen initially collapsing by face yield, core shear, and indentation, respectively. Each plot includes the analytical predictions of the elastic stiffness, the initial collapse load and the large-deflection membrane solution. The predicted transition point between the end of initial plastic collapse and the start of the membrane phase occurs at a deflection equal to the height of the beam, and this transition point is marked in the figures.

It is clear from Figs. 10(*a*) and 10(*b*) that, for the cases of face yield and core shear, there is a good agreement between the analytical predictions, the numerical model and the measured response. In particular, the prediction of the membrane phase accurately captures the measured response at $u > H_S$. In contrast, both the finite element predictions and analytical formulas underestimate the measured initial collapse load for the specimen collapsing by indentation, see Fig. 10(*c*) It is argued that this is due to the fact that the predictions neglect the presence of a strengthened boundary layer within the metal foam. This phenomenon has been

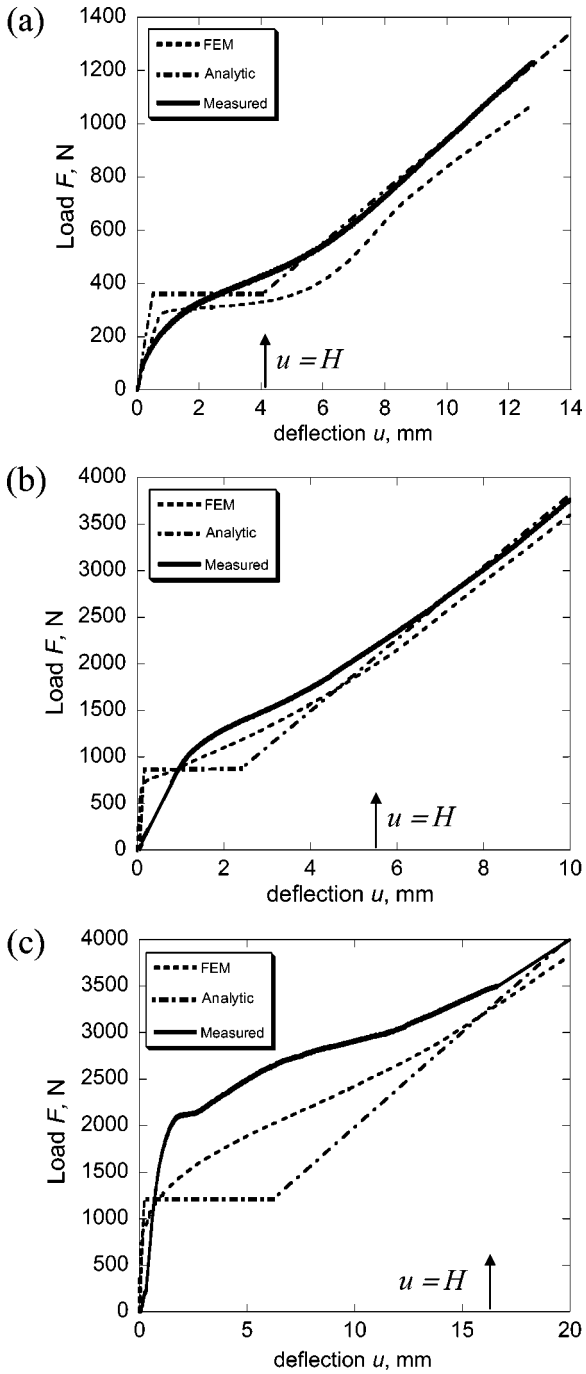


Fig. 10 Comparison of measured and predicted collapse responses for sandwich beams collapsing by (*a*) face yield, (*b*) core shear, and (*c*) indentation

observed previously for simply supported beams by Chen et al. [4], and has been analyzed in detail by Chen and Fleck [10]. They have discussed boundary layers for sandwich layers subjected to simple shear and shown experimentally and theoretically that the strength is enhanced when the thickness of the core is comparable to the cell size. A similar elevation is expected when the width of the indenter is comparable to the cell size, as in the present study. The source of the boundary layer is the fact that the foam cell walls are adhered to the face sheets and behave as encaster beams. For the indentation geometry the membrane solution is recovered when the transverse deflection u is comparable to the height H_S of the sandwich beam; the predicted large deflection solutions are again in reasonable agreement with the measured response.

Table 2 Geometry, face sheet strength, and observed mechanism of initial collapse for an additional set of experiments (key: SS = simply supported, CL = clamped, FY = face yield, CS = core shear, IN = indentation)

Spec. No.	t(mm)	c(mm)	l(mm)	a(mm)	b(mm)	Face sheet alloy	σ_Y(MPa)	Support conditions	Observed collapse mode
1	0.5	7	240	19	56	1	110	SS	FY
2	0.5	7	240	19	56	1	110	CL	FY
3	2	10	160	12.6	49	2	287	SS	CS
4	2	10	160	12.6	49	2	287	CL	CS
5	0.5	40	160	12.6	50	3	90	SS	IN
6	0.5	40	160	12.6	50	3	90	CL	IN
7	3	19	220	19	57	1	120	CL	CS
8	2	10	160	12.6	49	2	287	CL	CS
9	0.5	7	100	8	50	3	90	CL	CS
10	0.5	40	160	12.6	50	3	90	SS	IN
11	0.5	42	220	19	57	4	70	SS	FY
12	3	19	220	19	57	1	120	SS	CS
13	0.5	42	220	19	57	4	70	CL	FY

Additional Tests. Additional tests have been performed on clamped and simply supported specimens, using Alporas foam core and four different grades of aluminium alloy for the face sheets (the alloys are labeled in Table 2 as alloy 1 =BS HH/S1C, alloy 2=BS HE30TF, alloy 3=BS HH/S1C, and alloy 4=commercially pure, fully annealed aluminium). The geometry and strength of the faces have been varied over a wide range in order to explore the accuracy of the analytical predictions of initial collapse strength. A summary of the specimen geometries and the associated face sheet properties is presented in Table 2. The predicted mode of collapse is in agreement with the observed mode. In Fig. 11 the predicted initial collapse loads are compared with the corresponding measured values. It is evident that the analytical predictions are adequate for design purposes.

7 Minimum Mass Design of Clamped Sandwich Structures

A common requirement is to optimize the design to achieve a minimum mass for a given structural stiffness, strength, or level of energy absorption. Here we make use of the formulas developed in Sec. 2 in order to design clamped sandwich beams of minimum mass for a given initial collapse strength in three point bending. A complementary optimisation task has already been performed for simply supported aluminium sandwich beams by Chen et al. [4].

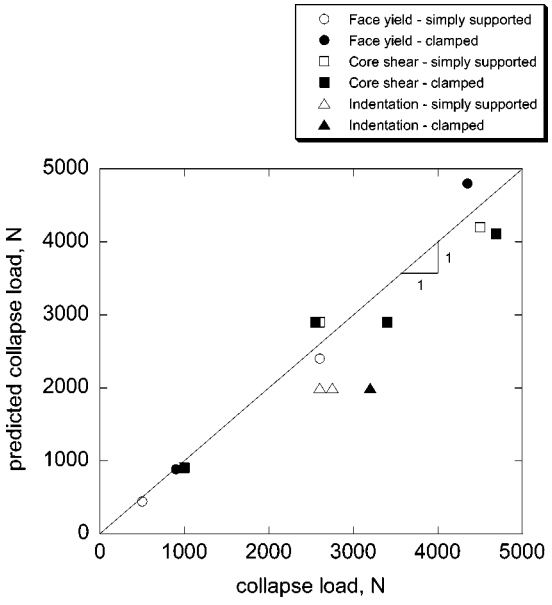


Fig. 11 Comparison of measure and analytical prediction of initial collapse strength for the specimens listed in Table 2

The first step is to construct a collapse mechanism map in terms of the nondimensional geometrical parameters $\bar{c}=c/\ell$ and $\bar{t}=t/c$, for a given a set of material properties of face sheets and core. A typical map is given in Fig. 12 for a clamped sandwich beam with aluminium alloy faces and an Alporas foam core, with the choice $\bar{\sigma}=\sigma_c/\sigma_f=0.034$, $\bar{a}=a/\ell=0.1$, $\bar{\rho}=0.11$. The dominant collapse modes are shown, as in Fig. 5, along with contours of nondimensional collapse load $\bar{F}=F/b\ell\sigma_f$ and mass $\bar{M}=M/b\ell^2\rho_f$. The geometry which minimises $\bar{M}$ at any fixed $\bar{F}$ is obtained by scanning along the contour of $\bar{F}$ to locate the point where the gradient $\nabla\bar{M}$ is locally parallel to $\nabla\bar{F}$. Upon repeating this procedure for increasing values of $\bar{F}$ a minimum mass trajectory is located, as shown in Fig. 12. Algebraic calculations, not reported here for the sake of brevity, give explicit analytical expressions for the dependence on the minimum mass index $\bar{M}_{\min}$ as a function of the required structural strength $\bar{F}$.

The definitions (2) for $\bar{F}$ and $\bar{M}$ involve the strength σ_f and density ρ_f of the face sheets. To allow for a direct comparison of the performance of various material combinations, the normalized values $\bar{F}^N$ of $\bar{F}$ and $\bar{M}^N$ of $\bar{M}$ are introduced, by using the strength σ_s and density ρ_s of a medium strength steel, taken as 400 MPa and 8000 kg/m^3, respectively;

$$\bar{F}^N=\frac{\sigma_f}{\sigma_s}\bar{F};\quad \bar{M}^N=\frac{\rho_f}{\rho_s}\bar{M} \tag{24}$$

The normalized minimum mass design $\bar{M}^N_{\min}$ is plotted as a function of the structural load index $\bar{F}^N$ in Fig. 13. The figure includes

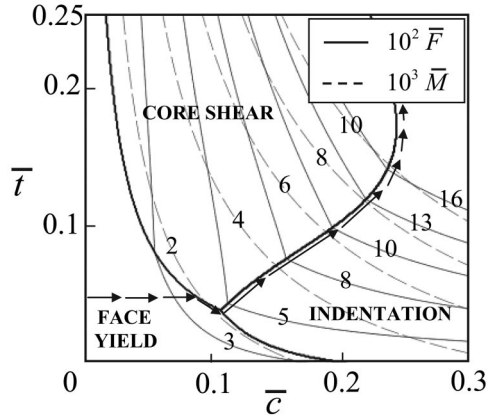


Fig. 12 Collapse mechanism map with contours of the nondimensional strength and mass index ($\bar{\sigma}=0.034, \bar{a}=0.1, \bar{\rho}=0.11$). The minimum mass trajectory is included.

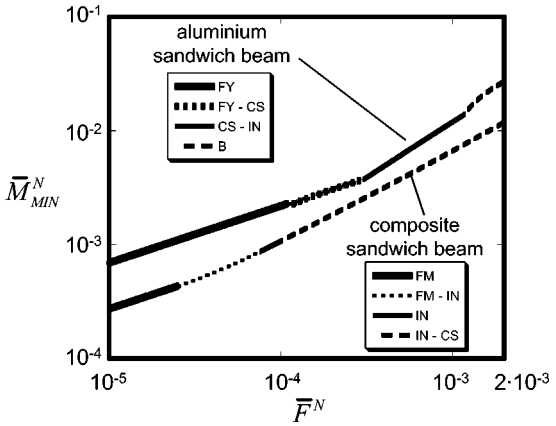


Fig. 13 Normalized minimum mass vs structural load index for a clamped sandwich beam of metallic construction and of composite construction (key: FM = face microbuckling, FY = face yield, CS = core shear, IN = indentation)

the minimum weight design plot for a clamped beam with glass-vinylester composite faces and H100 Divinycell foam core, taken from a parallel study [11]. The metallic sandwich performance is similar to that of the composite construction, and additional benefit would accrue from the use of heat-treated aluminium alloy face sheets.

8 Effect of Foam Core Upon Plane Strain Necking of Face Sheets

The present experimental study made use of annealed aluminium alloy face sheets. These possessed adequate strain hardening capacity to maintain stability and not undergo necking during the beam bending tests. Preliminary experiments (not reported here) using high strength aluminium alloy revealed that the peak load of clamped beams is set by sheet necking of the faces.

It is anticipated that the presence of a foam core delays the onset of tensile necking of the face sheets in the membrane phase of the response. Sheet metal necking involves a local reduction in thickness of the sheet, and a foam core provides resistance to this instability. This phenomenon has been explored experimentally as follows. Dog-bone shaped tensile specimens were made from a sandwich plate with faces comprising a BS HH/S1C grade of commercially pure, cold rolled aluminium of thickness t =0.9 mm, and Alporas form core of relative density 11% and thickness in the range 3–25 mm. The dog-bone specimens had a gauge length of 70 mm and a width of 25 mm; testing of the sandwich specimens was performed both along the rolling direction of the faces and transverse to the rolling direction.

The choice of material for the face sheets of the sandwich specimens was dictated by the requirement for the faces to undergo tensile necking at a low ductility (of the order of 1%) prior to tensile rupture of the foam core. The measured tensile ductility of the faces was ε_F=0.82% in the rolling direction and ε_F =1.12% in the transverse direction; for the two orientations the 0.2% offset yield strength equals 100 and 120 MPa, respectively.

Longitudinal sections of the necked face sheet are shown in Fig. 14(*a*) (no foam core present) and in Fig. 14(*b*) (foam core present). A typical load versus nominal strain curve for the sandwich specimen (core thickness c=25 mm) is given in Fig. 14(*c*), for the case of loading transverse to the roll direction of the faces. The figure includes a simple rule-of-mixtures estimate for the tensile response of the sandwich plate, based on the assumption that the axial strain is uniform across the section. It is evident that the prediction is accurate up to an axial strain of about 0.8%; beyond this strain, unsupported face sheets undergo tensile necking while

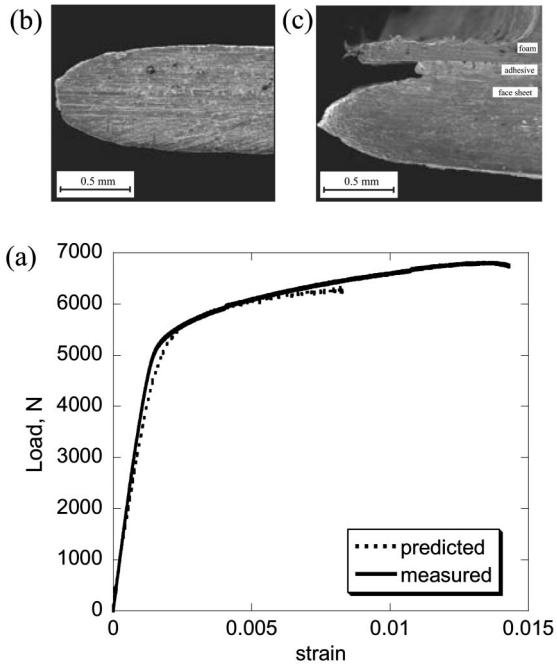


Fig. 14 Scanning electron micrographs of the tensile necks in (*a*) aluminium alloy face sheet with no foam support, and (*b*) aluminium alloy face sheet as part of a sandwich plate. (*c*) Measured tensile load vs strain response for a sandwich dog-bone specimen. The predicted response by an upper bound, rule-of-mixtures calculation is included.

the sandwich specimen remains stable up to a strain of 1.45%. This supports the hypothesis that the foam core stabilises the faces against tensile necking.

The magnitude of the delay in necking is dependent upon ratio of face sheet thickness to core thickness t/c, as shown by the ratio of ductility of sandwich ε_{sw} to that of the faces ε_{fs}, see Fig. 15. It is evident from the figure that this ratio increases with decreasing face sheet thickness and with decreasing ductility of the face sheet. The effect can be large: the measured ductility of the sandwich specimen can be almost doubled by the presence of the foam.

9 Concluding Remarks

This study has focused on the effect of boundary conditions on the flexural response of sandwich beams comprizing aluminium faces and an aluminium foam core. For both simply supported and

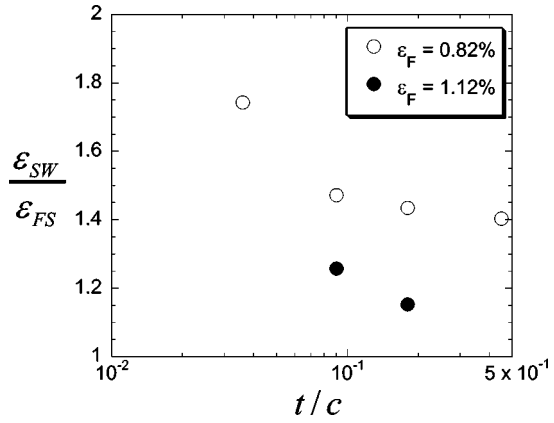


Fig. 15 Sensitivity of tensile ductility of dog-bone sandwich specimens to the ratio of face sheet to core thickness

clamped beams, initial collapse is by core shear, face yield, or by face indentation. Simple limit load calculations and more detailed finite element calculations capture the collapse response except for the case of face sheet indentation; the measured indentation strength is significantly higher than that predicted and it is argued that this is due to the presence of a strengthened boundary layer within the foam adjacent to the face sheets, along the lines discussed by Chen and Fleck [10]. In all clamped beam tests, initial collapse was followed by a stable regime of increasing load with transverse deflection. This regime of membrane stretching begins when the transverse deflection is comparable with the depth of the beam, and ends with tearing of the face sheets. The tensile ductility of the faces is found to be increased by the presence of the foam core—this beneficial effect is due to the stabilization offered by the core to the onset of sheet necking of the faces.

The dominant modes of initial collapse are summarized in a collapse mechanism map, with axes given in terms of the geometry of the beam. The map is useful in the optimizing beam geometry for minimum mass, for any given value of structural load index.

Acknowledgments

This work was supported by the U.S. Office of Naval Research, Contract No. 0014-91-J-1916. The authors are grateful to Dr. V. S. Deshpande and Dr. M. P. Zupan for helpful discussions.

References

[1] Ashby, M. F., Evans, A. G., Fleck, N. A., Gibson L. J., Hutchinson J. W., and Wadley H. N. C., 2000, *Metal Foam: A Design Guide*, Butterworth, Washington, DC.

[2] Evans, A. G., Hutchinson, J. W., Fleck, N. A., Ashby, M. F., and Wadley, H. N. G., 2001, "The Topological Design of Multifunctional Cellular Material," Prog. Mater. Sci., **46**(3–4), pp. 309–327.

[3] Wadley, H. N. G., Fleck, N. A., and Evans, A. G., 2003, "Fabrication and Structural Performance of Periodic Cellular Metal Sandwich Structures," Compos. Sci. Technol. **63**, pp. 2331–2343.

[4] Chen, C., Harte, A.-M., and Fleck, N. A., 2000, "The Plastic Collapse of Sandwich Beams With a Metallic Foam Core," Int. J. Mech. Sci., **43**, pp. 1483–1506.

[5] Bart-Smith, H., Hutchinson, J. W., and Evans, A. G., 2001, "Measurement and Analysis of the Structural Performance of Cellular Metal Sandwich Construction," Int. J. Mech. Sci., **43**, pp. 1945–1963.

[6] Allen, H. G., 1969, *Analysis and Design of Structural Sandwich Panels*, Pergamon Press, NewYork.

[7] Gibson, L. J., and Ashby, M. F., 1997, *Cellular Solids, Structure and Properties*, 2nd. ed., Cambridge University Press, Cambridge, UK.

[8] Deshpande, V. S., and Fleck, N. A., 2000, "Isotropic Constitutive Models for Metallic Foams," J. Mech. Phys. Solids **48**(6–7), pp. 1253–1283.

[9] Chen, C., 1998, "Manual for a UMAT User Subroutine," Cambridge University Engineering Department Report, CUED/C-MICROMECHANICS/TR.4.

[10] Chen, C., and Fleck, N. A., 2002, "Size Effects in the Constrained Deformation of Metallic Foams," J. Mech. Phys. Solids, **50**, pp. 955–977.

[11] Tagarielli, V. L., Fleck, N. A., and Deshpande, V. S., 2003, "Three-Points Bending of Clamped Composite Sandwich Beams," Composites, Part B, **35**, pp. 523–534.

H. J. Chu
Department of Mechanics and Engineering Science,
Peking University,
Beijing 100971, P. R. China
and
College of Hydraulic Science and Engineering,
Yangzhou University,
Yangzhou 225009, P. R. China

M. Z. Wang
Department of Mechanics and Engineering Science,
Peking University,
Beijing 100871, P. R. China
e-mail: wangminz@pku.edu.cn

A Note on the Limit Definition of Concentrated Loads

Based on Sternberg and Eubanks' limit definition of concentrated loads, an improved definition is proposed. The polar-symmetry example is given to show that the new definition is more reasonable. Sternberg and Eubanks' attractive counter-example is reconsidered. [DOI: 10.1115/1.1875472]

1 Introduction

The classical uniqueness theorem does not hold at concentrated-load points (singularities). To assure the unique solution to a concentrated-load problem, each concentrated load is often replaced with a sequence of distributed loadings over the load region surrounding the load point, and the solution to the original problem is reached by considering the limit solution to the modified problem. Sternberg and Eubanks [1] gave their celebrated limit definition which was analogous to Kelvin's definition through a limit process of the solution associated with a concentrated force applied at an internal point of a medium occupying the entire space. Sternberg and Rosenthai [2] gave an infinite aggregate of distinct "solutions" corresponding to the half-space and sphere under normal concentrated loads, each of which possesses Kelvin's three conditions, but not their new forth condition. The limit definition of concentrated surface loads was obtained by Turteltaub and Sternberg [3]. Their work shows that the traditional concept of concentrated loads is not sufficient, and if the additional property is neglected, the solution to such a problem will not be unique.

In contrast to the Sternberg and Eubanks limit definition, an improved version with their additional property being relaxed is proposed in this paper, and the amended expression is very simple. We also give a counter-example that satisfies the relaxed property but fails to agree with the foregoing limit definition to show that the new definition is more reasonable.

By the way, we argue about the polar symmetry counter-example which was used by Sternberg and Eubanks to prove that the fourth condition should not be neglected [1], and get the exact solution of the displacement which is different from theirs. It should be noted that this counter-example itself supports their final conclusion.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, April 18, 2004; final revision; September 27, 2004. Associate Editor: Z. Suo. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

2 Sternberg and Eubanks' Limit Definition and the Improved One

2.1 Definition 1 (Sternberg and Eubanks's Limit Definition [1,3–5]). A concentrated load $\boldsymbol{F}$ applied at the point $\boldsymbol{r}_0$ in the entire space E^3 can be replaced with a sequence of body-force fields $\boldsymbol{f}_n(\boldsymbol{r}-\boldsymbol{r}_0)$ with the properties:

(a) $\quad \boldsymbol{f}_n(\boldsymbol{r}-\boldsymbol{r}_0) \in C^2(E^3)$,

(b) $\quad \boldsymbol{f}_n(\boldsymbol{r}-\boldsymbol{r}_0) = 0$ for $\boldsymbol{r} \notin \Sigma_{1/n}(\boldsymbol{r}_0)$,

$$\int\int\int_{E^3} \boldsymbol{f}_n(\boldsymbol{r}-\boldsymbol{r}_0)\mathrm{d}\tau_r \to \boldsymbol{F} \text{ as } n \to \infty$$

(d) $$\int\int\int_{E^3} |\boldsymbol{f}_n(\boldsymbol{r}-\boldsymbol{r}_0)|\mathrm{d}\tau_r \text{ remains bounded as } n \to \infty$$

where $\boldsymbol{r}$ is the position vector of an arbitrary point from the origin, $C^2(E^3)$ denotes an aggregate of twice continuous differentiable functions in three-dimensional Euclidean space, and $\Sigma_{1/n}(\boldsymbol{r}_0)$ denotes a sphere region with its center $\boldsymbol{r}_0$ and radius $1/n$.

2.2 Definition 2 (the Improved limit definition). The concentrated load $\boldsymbol{F}$ applied at the point $\boldsymbol{r}_0$ in the entire space E^3 can be replaced with a sequence of body-force fields $\boldsymbol{f}_n(\boldsymbol{r}-\boldsymbol{r}_0)$ with properties (a), (b), (c), and amended property (d')

$$(\mathrm{d}') \qquad \frac{1}{n}\int\int\int_{E^3} |\boldsymbol{f}_n(\boldsymbol{r}-\boldsymbol{r}_0)|\mathrm{d}\tau_r \to 0 \text{ as } n \to \infty$$

where symbols have the same meaning as before.

Compared with the first definition, the requirements of the second is relaxed obviously.

Theorem 1: The displacement under the action of the sequence of body-force fields $\boldsymbol{f}_n(\boldsymbol{r}-\boldsymbol{r}_0)$ with properties defined by Definition 2 tends to the Kelvin solution as $n \to \infty$.

Proof: The displacement under the action of the sequence of body-force fields $\boldsymbol{f}_n(\boldsymbol{r}-\boldsymbol{r}_0)$ is

$$\boldsymbol{u}_n(\boldsymbol{r}) = \boldsymbol{P}_n(\boldsymbol{r}) - \frac{1}{4(1-\nu)}\nabla\left[P_{0n}(\boldsymbol{r}) + \boldsymbol{r}\cdot\boldsymbol{P}_n(\boldsymbol{r})\right] \tag{1}$$

and the Kelvin solution is

$$\boldsymbol{u}^F(\boldsymbol{r}) = \boldsymbol{P}(\boldsymbol{r}) - \frac{1}{4(1-\nu)}\nabla\left[P_0(\boldsymbol{r}) + \boldsymbol{r}\cdot\boldsymbol{P}(\boldsymbol{r})\right] \tag{2}$$

where

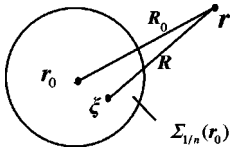


Fig. 1 Geometry of the loading region

$$\boldsymbol{P}_n(\boldsymbol{r}) = \frac{1}{4\pi\mu}\int\int\int_{\Omega}\frac{\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)}{R}d\tau_\xi \tag{3a}$$

$$P_{0n}(\boldsymbol{r}) = -\frac{1}{4\pi\mu}\int\int\int_{\Omega}\frac{\boldsymbol{\xi}\cdot\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)}{R}d\tau_\xi \tag{3b}$$

$$\boldsymbol{P}(\boldsymbol{r}) = \frac{1}{4\pi\mu}\frac{\boldsymbol{F}}{R_0}, \quad P_0(\boldsymbol{r}) = -\frac{1}{4\pi\mu}\frac{\boldsymbol{r}_0\cdot\boldsymbol{F}}{R_0} \tag{3c}$$

with

$$\boldsymbol{R} = \boldsymbol{r}-\boldsymbol{\xi}, \quad \boldsymbol{R}_0 = \boldsymbol{r}-\boldsymbol{r}_0$$

where μ and ν are the shear modulus and Poisson's ratio, respectively.

From the geometry relation of variables in Eq. (3) (Fig. 1), we get

$$|R-R_0| \leqslant |\boldsymbol{\xi}-\boldsymbol{r}_0| \leqslant \frac{1}{n}, \quad R \geqslant \left|R_0-\frac{1}{n}\right|$$

where $\boldsymbol{r}$ is in the entire space E^3, but not in the sphere region $\Sigma_{1/n}(\boldsymbol{r}_0)$.

From [1,5] and with some manipulation, we get

$$\begin{aligned}|\boldsymbol{P}_n(\boldsymbol{r})-\boldsymbol{P}(\boldsymbol{r})| &\leqslant \frac{1}{4\pi\mu R}\left|\int\int\int_{\Sigma_{1/n}}\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)\mathrm{d}\tau_\xi-\boldsymbol{F}\right| \\ &\quad+\frac{1}{4\pi\mu}\left|\int\int\int_{\Sigma_{1/n}}\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)\left(\frac{1}{R}-\frac{1}{R_0}\right)\mathrm{d}\tau_\xi\right| \\ &\leqslant \frac{1}{4\pi\mu R_0}\left|\int\int\int_{\Sigma_{1/n}}\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)\mathrm{d}\tau_\xi-\boldsymbol{F}\right| \\ &\quad+\frac{1}{4\pi\mu}\int\int\int_{\Sigma_{1/n}}|\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)|\frac{|R-R_0|}{RR_0}\mathrm{d}\tau_\xi \\ &\leqslant \frac{1}{4\pi\mu R_0}\left|\int\int\int_{\Sigma_{1/n}}\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)\mathrm{d}\tau_\xi-\boldsymbol{F}\right| \\ &\quad+\frac{1}{4\pi\mu}\int\int\int_{\Sigma_{1/n}}|\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)|\mathrm{d}\tau_\xi\frac{1}{nR_0(R_0-1/n)}\end{aligned}$$

Before the final inequality is obtained, the geometry relation has been used. When $n>2/R_0$, we have

$$\begin{aligned}|\boldsymbol{P}_n(\boldsymbol{r})-\boldsymbol{P}(\boldsymbol{r})| &\leqslant \frac{1}{4\pi\mu R_0}\left|\int\int\int_{\Sigma_{1/n}}\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)\mathrm{d}\tau_\xi-\boldsymbol{F}\right| \\ &\quad+\frac{1}{4\pi\mu}\int\int\int_{\Sigma_{1/n}}\frac{2|\boldsymbol{f}_n(\boldsymbol{\xi}-\boldsymbol{r}_0)|}{nR_0^2}\mathrm{d}\tau_\xi\end{aligned}$$

By the hypothesis (c), the first term at the right side of the above inequality tends to zero; by the hypothesis (d′), the second term tends to zero too, as $n\to\infty$.

With the same method, it is easy to prove $P_{0n}(\boldsymbol{r})\to P_0(\boldsymbol{r})$, as $n\to\infty$. We can also prove that the first-order derivatives of $\boldsymbol{P}_n(\boldsymbol{r})$ and $P_{0n}(\boldsymbol{r})$ are convergent to their corresponding derivatives of $\boldsymbol{P}(\boldsymbol{r})$ and $P_0(\boldsymbol{r})$, respectively. Then, Theorem 1 is proved.

3 Counter-Example and Analysis

In this part, we give an example that satisfies the relaxed property but fails to agree with Definition 1 to show that the relaxed result is effective. This example also supported that Theorem 1 is false if the hypothesis (d′) is omitted.

Let the sequence of body-force field $\boldsymbol{f}_n(\boldsymbol{r})$ be defined by

$$\boldsymbol{f}_n(\boldsymbol{r}) = \boldsymbol{r}^0 f_n(r), \quad \boldsymbol{r}\in\Sigma_{1/n}(O)$$

$$\boldsymbol{f}_n(\boldsymbol{r}) = 0, \quad \boldsymbol{r}\notin\Sigma_{1/n}(O) \tag{4}$$

where,

$$f_n(r) = -\frac{336}{5}(1-\nu)n^{9-\alpha}\left(r-\frac{1}{n}\right)^3\left(6r^2+\frac{3}{n}r+\frac{1}{n^2}\right)$$

where $\boldsymbol{r}^0$ is the unit vector of the position vector $\boldsymbol{r}$, point O is the origin of coordinates, and $0\leqslant\alpha<1$. By direct computing the first and second derivative of $\boldsymbol{f}_n(\boldsymbol{r})$, we get

$$\boldsymbol{f}_n\left(\frac{1}{n}\right) = \boldsymbol{f}_n'\left(\frac{1}{n}\right) = \boldsymbol{f}_n''\left(\frac{1}{n}\right) = 0 \tag{5}$$

whence $\boldsymbol{f}_n(\boldsymbol{r})$ satisfies the first property (a) and (b). For the polar symmetry of $\boldsymbol{f}_n(\boldsymbol{r})$ about O, the integral of the load sequence is

$$\int\int\int_{E^3}\boldsymbol{f}_n(\boldsymbol{r})\mathrm{d}\tau_r = 0 \tag{6}$$

so that $\boldsymbol{F}=0$. From Eq. (2), the Kelvin solution is

$$\boldsymbol{u}^{\boldsymbol{F}}(\boldsymbol{r}) = 0 \tag{7}$$

In order to calculate $\boldsymbol{u}_n(\boldsymbol{r})$, denoting P to be an arbitrary point out of the sphere region $\Sigma_{1/n}(O)$, and its position vector is $\boldsymbol{r}=\boldsymbol{r}(P)$ with the unit vector denoted by $\boldsymbol{r}^0$; let $\boldsymbol{i}$ be an arbitrary unit vector perpendicular to $\boldsymbol{r}^0$ and $\boldsymbol{j}$ equal to $\boldsymbol{r}^0\times\boldsymbol{i}$ be another basic vector. So, the orthogonal coordinate system $\{\boldsymbol{i},\boldsymbol{j},\boldsymbol{r}^0\}$ with the load-point origin is established. Q is an arbitrary point in the sphere region $\Sigma_{1/n}(O)$, and its position vector is $\boldsymbol{\xi}=\boldsymbol{\xi}(Q)$ with the unit vector denoted by $\boldsymbol{\xi}^0$. Then

$$R = |\boldsymbol{r}-\boldsymbol{\xi}| = \sqrt{r^2+\xi^2-2r\xi\cos\theta}$$

$$\boldsymbol{\xi}^0 = \boldsymbol{i}\sin\theta\cos\varphi+\boldsymbol{j}\sin\theta\sin\varphi+\boldsymbol{r}^0\cos\theta \tag{8}$$

where θ and φ are sphere coordinates of point Q. Substituting of Eq. (8) into Eq. (3a) we have

$$\begin{aligned}\boldsymbol{P}_n(\boldsymbol{r}) &= \frac{1}{4\pi\mu}\int\int\int_{E^3}\frac{\boldsymbol{f}_n(\boldsymbol{\xi})}{R}\mathrm{d}\tau_\xi = \frac{-336(1-\nu)n^{9-\alpha}}{10\mu} \\ &\quad\times\int_0^{\frac{1}{n}}\xi^2\left(\xi-\frac{1}{n}\right)^3\left(6\xi^2+\frac{3}{n}\xi+\frac{1}{n^2}\right) \\ &\quad\times\int_0^{\pi}\frac{\boldsymbol{r}^0\sin\theta\cos\theta}{\sqrt{r^2+\xi^2-2r\xi\cos\theta}}\mathrm{d}\theta\mathrm{d}\xi \\ &= -\frac{2(1-\nu)}{3\mu n^{\alpha}}\nabla\left(\frac{1}{r}\right)\end{aligned} \tag{9}$$

According to the potential theory [6], Eq. (3b) can be simplified, and we get

$$p_{0n}(\boldsymbol{r}) = -\frac{1}{4\pi\mu}\int\int\int_{E^3}\frac{\boldsymbol{\xi}\cdot\boldsymbol{f}_n(\boldsymbol{\xi})}{R}\mathrm{d}\tau_\xi$$

$$= \frac{1}{4\pi\mu r}\int\int\int_{\Sigma_{1/n}}\xi f_n(\xi)\mathrm{d}\tau_\xi$$

$$= -\frac{2(1-\nu)}{\mu}\frac{1}{r}\frac{1}{n^\alpha} \tag{10}$$

Substituting Eqs. (9) and (10) into Eq. (1), the displacement field is

$$\boldsymbol{u}_n(\boldsymbol{r}) = -\frac{1-2\nu}{3\mu}\frac{1}{n^\alpha}\nabla\left(\frac{1}{r}\right) \tag{11}$$

On the other hand, the absolute integral of $\boldsymbol{f}_n(\boldsymbol{r})$ can be calculated directly, and yield

$$\int\int\int_{E^3}|\boldsymbol{f}_n(\boldsymbol{r})|\mathrm{d}\tau_r = 16\pi(1-\nu)n^{1-\alpha} \tag{12}$$

By Eqs. (11) and (12), the displacement field and the absolute integral of the load sequence depend on the value of α, as $n\to\infty$. So two cases are discussed in the following part.

Case I: $0<\alpha<1$

By Eq. (11), we have

$$\boldsymbol{u}_n(\boldsymbol{r}) = 0, \quad \text{as } n\to\infty \tag{13}$$

This result shows that the displacement field $\boldsymbol{u}_n(\boldsymbol{r})$ tends to the Kelvin solution as $n\to\infty$. It means that the body-force sequence processes full four properties what the amended limit definition requires. Indeed,

$$\lim_{n\to\infty}\frac{1}{n}\int\int\int_{E^3}|\boldsymbol{f}_n(\boldsymbol{r})|\mathrm{d}\tau_r = 0 \tag{14}$$

However, this conclusion that the displacement field $\boldsymbol{u}_n(\boldsymbol{r})$ tends to the Kelvin solution, as $n\to\infty$, cannot be deduced by Definition 1, because the body-force sequence violates their fourth hypothesis (d). Indeed, as $0<\alpha<1$, we get

$$\int\int\int_{E^3}|\boldsymbol{f}_n(\boldsymbol{r})|\mathrm{d}\tau_r\to\infty, \quad n\to\infty \tag{15}$$

Comparison with the distinct different results derived by two different definitions in solving the same problem, we can conclude that Sternberg and Eubanks' limit definition is too strong, and the amended property (d′) instead of the property (d) is effective.

Case II: $\alpha=0$

By Eq. (11), we have

$$\boldsymbol{u}_n(\boldsymbol{r}) = -\frac{1-2\nu}{3\mu}\nabla\left(\frac{1}{r}\right) \tag{16}$$

According to Eqs. (7) and (16), this result of the displacement field does not tend to the Kelvin solution, as $n\to\infty$. In fact, the sequence of the body-force field $\boldsymbol{f}_n(\boldsymbol{r})$ processes the first three properties, but violates the fourth one (d′). Indeed,

$$\lim_{n\to\infty}\frac{1}{n}\int\int\int_{E^3}|\boldsymbol{f}_n(\boldsymbol{r})|\mathrm{d}\tau_r = 16\pi(1-\nu) \tag{17}$$

So, if the property (d′) is neglected, Theorem 1 is false. In his book [5], M. Z. Wang gives another example with a different form to support this view. The example is defined by

$$\boldsymbol{f}_n(\boldsymbol{r}) = \boldsymbol{i}f_n(r), \quad \boldsymbol{r}\in\Sigma_{1/n}(O)$$

$$\boldsymbol{f}_n(\boldsymbol{r}) = 0, \quad \boldsymbol{r}\notin\Sigma_{1/n}(O) \tag{18}$$

where

$$f_n(r) = Cn^{11}\left(r-\frac{1}{n}\right)^3\left(30r^3-\frac{6}{n}r^2-\frac{3}{n^2}r-\frac{1}{n^3}\right)$$

with

$$C = -1980(1-\nu)\mu$$

where $\boldsymbol{i}$ is unit vector with constant direction. So, this sequence of the body-force field has a constant direction that is different with Sternberg and Eubanks' counter-example [1]. And this example supports that property (d′) should not be neglected, too.

Next, we will reconsider Sternberg and Eubanks' counter-example. When $\alpha=0$, the foregoing counter-example defined by Eq. (4) will be simplified to Sternberg and Eubanks' counter-example defined by Eqs. (4.8), (4.9), and (4.10) in the article [1]. Substitute $\alpha=0$ into Eqs. (9) and (11), we get

$$\frac{-1}{8\pi(1-\nu)}\int\int\int_{E^3}\frac{\boldsymbol{f}_n(\boldsymbol{\xi})}{R}\mathrm{d}\tau_\xi = \frac{1}{3}\nabla\left(\frac{1}{r}\right)$$

$$\boldsymbol{u}(\boldsymbol{r}) = -\frac{1-2\nu}{3\mu}\nabla\left(\frac{1}{r}\right) \tag{19}$$

However, the corresponding expressions in the article [1] are

$$\boldsymbol{V}_n(P) \equiv \frac{-1}{8\pi(1-\nu)}\int\int\int_{E^3}\frac{\boldsymbol{f}_n(\boldsymbol{\xi})}{R}\mathrm{d}\tau_\xi$$

$$= \frac{-1}{8\pi(1-\nu)r}\int\int\int_{\Sigma_{1/n}}\boldsymbol{f}_n(\boldsymbol{\xi})\mathrm{d}\tau_\xi = 0 \tag{20a}$$

$$\boldsymbol{u}(\boldsymbol{r}) = \nabla\left(\frac{1}{r}\right) \tag{20b}$$

The above Eq. (20*a*) is obtained from Eqs. (4.4*b*), (4.13*b*), (4.15*b*) in [1] directly; and Eq. (20*b*) is the formula (4.17) in [1].

Comparison with Eqs. (19) and (20), the two solution to the same problem are contrary; and we think the later is wrong. Based on [1], the first equation in Eq. (20) is obtained through potential theory, for $\boldsymbol{V}_n(P)$ being equivalent to a Newtonian potential at a point of free space of mass distributions over the sphere, whose densities have polar symmetry about the center O, and the value of such a potential at P equal to the value of the potential associated with a single particle at O, whose mass is equal to the total mass of the distribution [1]. In fact, the above deduction is right, only when the densities are scalar fields [6] or vector fields with constant direction (tensor fields are not considered here). To our problem, the density of mass distribution $\boldsymbol{f}_n(\boldsymbol{\xi})$ is a vector, and its direction varies with the position of point Q, thus this deduction cannot be used in this integral, i.e., the expression Eq. (20*a*) cannot be derived by potential theory. So, the latter formula of the displacement field Eq. (20*b*) which is related to the former cannot be obtained, too. However, it is worth mentioning that this attractive counter-example itself is very good to support their final conclusion that the fourth requirement should not be omitted.

4 Conclusion and Discussion

Sternberg and Eubanks' limit definition is monumental. They clarified the confusion about the concept of concentrated loads. But their condition (d) is too strong. The new limit definition of concentrated loads with the fourth condition being relaxed can assure the limit solution to the modified problem convergent to the Kelvin solution. Our counter-example shows that the improved version is more reasonable than the previous one. By establishing an appropriate coordinate, and computing the integral directly, we get the exact displacement field of the attractive polar symmetric counter-example. Moreover, the limit definition of concentrated surface loads given by Turteltaub and Sternberg [3] can be relaxed in the same way.

Acknowledgments

Support from the National Natural Foundation of China (Nos. 10172003 and 10372003) are acknowledged.

References

[1] E. Sternberg and R. A. Eubanks, 1955, "On the Concept of Concentrated Loads and an Extension of the Uniqueness Theorem in the Linear Theory of Elasticity," J. Rat. Mech., Anal., **4**, pp. 135–168.

[2] E. Sternberg and F. Rosenthai, 1952, "The Elastic Sphere under Concentrated Loads," ASME J. Appl. Mech., **19**, pp. 413–421.

[3] M. J. Turteltaub and E. Sternberg, 1968, "On Concentrated Loads and Green's Functions in Elastostatics," Arch. Ration. Mech. Anal., **29**, pp.193–240.

[4] C. Truesdell, 1984, *Mechanics of Solids II*, Springer-Verlag, Berlin, pp. 173–190.

[5] M. Z Wang, 2003, *Advanced Theory of Elasticity* (Chinese edition), Peking University Press, Beijing, pp. 203–252.

[6] O. D. Kellogg, 1929, *Foundations of Potential Theory*, Springer-Verlag, Berlin, pp. 48–84.

A Unified Formalism of Two-Dimensional Anisotropic Elasticity, Piezoelectricity and Unsymmetric Laminated Plates

Wan-Lee Yin
Professor,
School of Civil and Environmental Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332-0355
e-mail: yinwl@aol.com

A unified formalism is presented for theoretical analysis of plane anisotropic elasticity and piezoelectricity, unsymmetric anisotropic plates, and other two-dimensional problems of continua with linear constitutive relations. Complex variables are used to reduce the governing differential equations to algebraic equations. The constitutive relation then yields an eigenrelation, which is easily solved explicitly for the material eigenvalues and eigenvectors. The latter have polynomial expressions in terms of the eigenvalues. When the eigenvectors are combined after multiplication by arbitrary analytic functions containing the corresponding eigenvalues, one obtains the two-dimensional general solution. Important results, including the orthogonality of the eigenvectors, the expressions of the pseudometrics and the intrinsic tensors, are established here for nondegenerate materials, including the case of all distinct eigenvalues. Green's functions of the infinite domain, and of the semi-infinite domain with interior or edge singularities, are determined explicitly for the most general types of point loads and discontinuities (dislocations).
[DOI: 10.1115/1.1828060]

1 Introduction

It is shown in this paper that, in the plane equilibrium problems of anisotropic elasticity and piezoelectricity, laminated plate theory with or without bending-stretching coupling, and other two-dimensional theories of continua, the differential equations governing the kinematical and kinetic variables (and additional variables characterizing the electric and other states) often show analogous mathematical forms, implying that the stress, strain, electric field, electric displacement and other physical variables appear as two-vectors and three-vectors of the gradient type or the solenoidal type. The gradient vectors possess a scalar potential whereas the solenoidal vectors have a skew potential. An eigensolution $\boldsymbol{\chi}$ consisting of all potential functions of the physical fields is expressed as an analytic function of a complex variable $x+\mu y$ multiplied by a (constant) eigenvector $\boldsymbol{\xi}$. When the linear constitutive equation of the material is used to relate the derivatives of the potentials, one obtains an algebraic eigenrelation that may be solved easily for the eigenvalue μ and the eigenvector $\boldsymbol{\xi}$. The two-dimensional general solution of the material is obtained by combining the eigensolutions associated with all eigenvalues.

This unified formalism includes, as special cases, the Lekhnitskii and Stroh formalisms of anisotropic elasticity [1–5], the extended Lekhnitskii and Stroh formalisms of piezoelectricity [3,6–8], and anisotropic laminated plate theory in different formalisms including those of Becker [9], Lu and Mahrenholtz [10], Cheng and Reddy [11], Chen and Shen [12], Hwu [13] and Yin [14–17]. In all such particular theories, the complex-variable formalism leads to an algebraic eigenrelation

$$\mathbf{M}(\mu)\,\boldsymbol{\eta}=0, \tag{1.1}$$

where $\boldsymbol{\eta}$ is a *reduced eigenvector* whose elements form a subset of the eigenvector $\boldsymbol{\xi}$, and $\mathbf{M}(\mu)$ is a symmetric matrix whose elements are polynomial functions. For example, in 2-D anisotropic elasticity, $\mathbf{M}(\mu)$ is a 2×2 matrix whose elements are the well known polynomials $l_4(\mu)$, $-l_3(\mu)$ and $l_2(\mu)$ in Lekhnitskii's theory. The material eigenvalues are determined by the characteristic equation

$$Det[\mathbf{M}(\mu)]=0. \tag{1.2}$$

If this equation has only simple roots, then the *reduced* eigenvector $\boldsymbol{\eta}$ for each root is obtained easily from Eq. (1.1), and the eigenvector $\boldsymbol{\xi}$, unique except for an arbitrary scalar factor, follows from a simple relation. Even if the characteristic equation has multiple roots, Eq. (1.1) may still give a full set of independent eigenvectors, provided that such roots μ_0 are double roots and the rank of the matrix $\mathbf{M}(\boldsymbol{\mu}_0)$ is lower than its dimension by two (such cases are called semi-simple). With rare exceptions [18], most of the studies in the vast literature on anisotropic elasticity, and almost all works on piezoelectricity and anisotropic plates, are concerned only with this relatively simple case of *nondegenerate* materials.

There are, however, abundant cases of degenerate materials, including all isotropic materials in elasticity, all symmetric quasi-isotropic laminates, and unsymmetric laminates composed of dissimilar isotropic layers, for which the general solution contains higher-order eigensolutions that cannot be obtained from Eq. (1.1), and must be obtained from more complicated eigenrelations for multiple eigenvalues. For such materials and laminates, the complete results for the general solution and for Green's functions of simple domains are obtained in the author's recent papers on anisotropic elasticity [4,5,19], anisotropic laminated plates [14–17] and piezoelectricity [7,8].

These results of nondegenerate and degenerate materials of the various physical theories follow from the common mathematical features of the governing differential equations, and from the linear constitutive matrix relating the primary and secondary variables. The theories generally involve various types of physical

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, May 18, 2004; final revision, July 18, 2004. Associate Editor: Z. Suo. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California—Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

coupling, for example, electro-mechanical coupling in piezoelectricity and bending-stretching coupling in unsymmetric laminated plates.

In the unified formalism developed in this paper, the governing differential equations are first reduced to algebraic equations containing the eigenvalues, eigenvectors, and the constitutive matrix. The potentials and skew-potentials of the unknown physical variables are separated into conjugate groups of primary and secondary variables. The secondary variables are then eliminated, to result in an eigenrelation governing a reduced set of the primary variables, which has the form of Eq. (1.1). The equation yields all eigenvalues, and all zeroth-order eigenvectors. Many important results, including orthogonality of eigenvectors, projection operators and intrinsic tensors, may be obtained in a context-independent manner, regardless of the specifics of the theory, such as the number of variables and their physical meanings.

The importance of the unified formalism is that it allows the material eigenvalues and eigensolutions, the general solution, the intrinsic tensors, Green's functions and the solutions of the boundary value problems of various domains to be found in a common form for a diverse range of theories, including the three examples mentioned above. Thus, in the nondegenerate case, the general solution $\boldsymbol{\chi}$ of the potentials and skew potentials is expressed in a unified way as

$$\boldsymbol{\chi}=2\ \mathrm{Re}[\mathbf{Z}_{\perp}\langle f_1(x+\mu_1 y),\ldots,f_N(x+\mu_N y)\rangle], \tag{1.3}$$

where $\{\mu_1,\ldots,\mu_N\}$ is the complete set of eigenvalues with positive imaginary parts, $\langle f_1,\ldots,f_N\rangle$ is a diagonal matrix containing the N arbitrary analytic functions $f_1,\ldots,f_N$, and $\mathbf{Z}_{\perp}$ is a $2N\times N$ matrix whose kth column is the eigenvector associated with μ_k. The general form of Green's function in the infinite domain is given for all three theories by the common expression:

$$\boldsymbol{\chi}=(1/\pi)\mathrm{Re}[\mathbf{Z}_{\perp}\langle -i\log[x+\mu_1 y],\ldots, -i\log[x+\mu_N y]\rangle\mathbf{Z}_{\perp}^{-1}]\boldsymbol{\chi}_0, \tag{1.4}$$

where the elements of the constant vector $\boldsymbol{\chi}_0$ are the concentrated point forces, dislocations, or point charges imposed at the origin. Other important results of the different theories are also given by unified expressions in a context-independent manner.

Furthermore, a solution of the boundary value problem for a certain domain in one theory may be converted directly to a corresponding solution for the same domain in another mathematically equivalent theory. Such a correspondence of solutions exists, for example, between 2-D anisotropic elasticity having in-plane and anti-plane coupling and 2-D piezoelectricity having no such coupling. The anti-plane shear stresses and shear strains of the first theory are replaced by the electric displacement vector and the electric field vector, respectively. Another example is given by anisotropic laminates without bending-stretching coupling. For such mid-plane symmetric laminates, stretching solutions may be converted into bending solutions, and vice versa, but the roles of kinematical and kinetic variables must be interchanged [15]. In a more general context, if a singularity solution or an analytical solution of a boundary value problem is found in one theory, it may suggest the corresponding solutions in other theories with a different set and number of variables through the mathematical analogy of the unified formalism. For example, it is found in Sec. 7 of this paper that Green's function of the semi-infinite region of a nondegenerate material with free, fixed or mixed boundary conditions has the general expression of Eq. (7.8) for different theories. Besides the material eigenvalues and eigenvectors, the expression involves a constant matrix $\mathbf{T}$, which is defined by Eq. (7.7*b*) in terms of the boundary condition matrix $\mathbf{K}$ and the eigenvectors. Thus, all expressions of Green's function for the half space previously obtained in the various particular theories of elasticity or anisotropic laminates are mathematically identical, provided that the physical variables and constitutive matrices are chosen strictly in the manner described in this paper (this requires, in particular, that one must use the Lekhnitskii formalsim instead of the Eshelby-Stroh formalism in anisotropic elasticity, and must use the extended Lekhnitskii formalism in other theories). This common mathematical form of Green's function is automatically extended to a general expression valid for a wide range of theories involving more general linear coupling effects of mechanical, electromagnetic, and other physical types.

Only nondegenerate materials are considered in this paper. The many classes of degenerate materials involve more complicated relations and results, and the complexity increases rapidly with the number of variables. Thus, the complete analysis for nondegenerate and degenerate piezoelectricity [7,8] is significantly lengthier than the complete analysis for anisotropic elasticity [4,5] or unsymmetric laminated plates [14,15]. No complete solution can be given for all degenerate cases in a theory unless the complete set of eigenvalues and their multiplicities are explicitly found. However, many important analytical results including the derivative rule, orthogonality and the structure of eigenspaces, pseudometrics, projection operators and intrinsic tensors [8,15,19] may be established in a unified way independent of the particular theories.

2 Reduction of the Governing Equations to a Single Eigenrelation

S-Vectors and G-Vectors. We consider *two-dimensional* equilibrium problems of a continuum, in which all physical variables that directly participate in the linear constitutive relation (stress, strain, electric field, electric displacement, curvatures and bending and twisting moments in thin plates, etc.) are functions of two rectangular coordinates x and y only. These variables may belong to one of the following four categories (in which S stands for ''solenoidal,'' implying that the two-dimensional vector or tensor has a vanishing divergence; a 2-G vector is the gradient of a scalar function of x and y, and a 3-G vector has its components given by the symmetric part of the gradient of a two-dimensional vector):

1. 3-S vector. Examples are
   $\{\sigma_x,\sigma_y,\tau_{xy}\}^T=\{F_{,yy},F_{,xx},-F_{,xy}\}^T$, where $F(x,y)$ is Airy's stress function in plane elasticity,
   $\{\kappa_y,\kappa_x,-\kappa_{xy}\}^T=\{w_{,yy},w_{,xx},-w_{,xy}\}^T$, where $w(x,y)$ is the normal displacement of a thin plate.
2. 2-S vector. Examples are
   $\{\tau_{xz},\tau_{yz}\}^T=\{\psi_{,y},-\psi_{,x}\}^T$, where $\psi(x,y)$ is the skew-potential of anti-plane shearing stresses,
   $\{D_x,D_y\}^T=\{\theta_{,y},-\theta_{,x}\}^T$, where $\theta(x,y)$ is the skew-potential of the electric displacement;
   $\{B_x,B_y\}^T$, two-dimensional magnetic flux density vector.
3. 3-G vector. Examples are
   $\{\varepsilon_x,\varepsilon_y,\gamma_{xy}\}^T=\{u_{,x},v_{,y},u_{,y}+v_{,x}\}^T$, where $u(x,y)$ and $v(x,y)$ are the displacement components,
   $\{M_y,M_x,-2M_{xy}\}^T=\{\Psi_{1,x},\Psi_{2,y},\Psi_{1,y}+\Psi_{2,x}\}^T$, where $\Psi_1(x,y)$ and $\Psi_2(x,y)$ are the moment potentials in plate problems.
4. 2-G vector. Examples are
   $\{\gamma_{xz},\gamma_{yz}\}^T=\{w_{,x},w_{,y}\}^T$, where $w(x,y)$ is the out-of-plane displacement in 2-D elasticity,
   $\{E_x,E_y\}^T=\{-\phi_{,x},-\phi_{,y}\}^T$, where $\phi(x,y)$ is the potential of the electric field,
   $\{H_x,H_y\}^T$, two-dimensional magnetic intensity vector in the absence of free current.

The vectors that belong to a particular category all satisfy governing differential equations with identical mathematical forms, though their physical meanings may be entirely different. Thus, the equations

$$\sigma_{x,x}+\tau_{xy,y}=0, \tag{2.1a}$$

$$\tau_{xy,x}+\sigma_{y,y}=0 \tag{2.1b}$$

are also satisfied when the two dimensional stress components σ_x, σ_y and τ_{xy} are replaced by the curvatures κ_y, κ_x and $-\kappa_{xy}$, respectively, of the plate bending problem. The equilibrium equation of the antiplane shearing stresses,

$$\tau_{xz,x}+\tau_{yz,y}=0, \tag{2.2}$$

is mathematically identical to $D_{x,x}+D_{y,y}=0$ for the electric displacement vector, and to a similar equation for the magnetic flux density vector $B_x\mathbf{i}+B_y\mathbf{j}$. The compatibility equation of the in-plane strains,

$$\varepsilon_{x,yy}+\varepsilon_{y,xx}-\gamma_{xy,xy}=0, \tag{2.3}$$

corresponds to the moment equilibrium equations of plates, $M_{y,yy}+M_{x,xx}-(-2M_{xy})_{,xy}=0$, whereas the compatibility of the out-of-plane shear strains

$$\gamma_{yz,x}-\gamma_{xz,y}=0 \tag{2.4}$$

is formally identical to the equation $E_{y,x}-E_{x,y}=0$ for the electric field. Such differential equations that define the mathematical character of a particular two-vector or three-vector of the S or G type will be called the ''inherent equations,'' in contrast to the material-dependent constitutive equations. In equilibrium problems, the inherent equations refer to the invariant spatial or balance relations that connect a dimensionally homogeneous set of physical variables independent of the specific material property. The constitutive equations are the relations among the different sets of field variables based on measurable material response. All transformation rules (invariance under rotation, etc.) that govern the constitutive equations belong to one of the two categories: (i) those that are derivable from the inherent transformation rules of the participating vectors or tensors, and therefore valid regardless of the material type, and (ii) symmetry relations that are valid only for special types of material defined by groups of symmetric transformation (orthotropy, isotropy, etc.).

In systems with an energy function, a 3-S vector always appears in the constitutive equation together with the conjugate 3-G vector. Similarly, 2-S and 2-G vectors also appear together in conjugate pairs. When a G-vector such as $\{\varepsilon_x,\varepsilon_y,\gamma_{xy}\}^T$ is kinematical, the conjugate S-vector $\{\sigma_x,\sigma_y,\tau_{xy}\}^T$ is kinetic. For the plate problem the roles are reversed: the 3-G vector $\{M_y,M_x,-2M_{xy}\}^T$ is kinetic, whereas the conjugate 3-S vector $\{w_{,yy},w_{,xx},-w_{,xy}\}^T$ is kinematical.

Constitutive Equations for Two-Dimensional Problems. In the literature of the various subjects, different constitutive equations involving the G-vectors and S-vectors have been widely used. The S-vectors may be expressed as linear functions of the G-vectors through a matrix of constitutive constants $\mathbf{C}$ (the stiffness matrix in elasticity). Conversely, the G-vectors are expressed in terms of the S-vectors using the inverse matrix $\mathbf{C}^{-1}$. Mixed formulations have also been developed where a combination of S-vectors and G-vectors is related to their conjugate G-vectors and S-vectors. Each formulation may have particular advantage compared to other formulations in certain special types of problems. But the adoption of a particular form of constitutive relation in the literature is often a matter of convention, and is not always based on compelling mathematical reason.

However, if one undertakes fundamental studies aiming at the general representation of the solutions of diverse classes of materials, Green's functions of various domains subjected to point loads and dislocations, as well as a thorough characterization of the mathematical structure of the solution space, then a suitable choice of the formalism dictated by the structure of the inherent equations may have decisive advantage over other choices, as pointed out in recent works on two-dimensional elasticity, unsymmetric laminated plates, and piezoelectricity by this author [4,5,7,8,14–17].

In these theories of two-dimensional continuum, all 3-G and 2-G vectors are expressed through a symmetric constitutive matrix $[\tilde{\omega}]$ in terms of the conjugate 3-S and 2-S vectors. The latter are called the primary variables, and their components are joined to form an S-vector of dimension $3m+2n$. The conjugate 3-G and 2-G vectors are called the secondary variables, and their components are joined in the corresponding manner to form a G-vector of dimension $3m+2n$.

The ith 3-S vector and the ith 3-G vector have the following expressions, respectively,

$$\mathbf{s}^{(i)}=\{F^{(i)}_{,yy},F^{(i)}_{,xx},-F^{(i)}_{,xy}\}^T, \tag{2.5a}$$

$$\mathbf{g}^{(i)}=\{U^{(i)}_{,x},V^{(i)}_{,y},U^{(i)}_{,y}+V^{(i)}_{,x}\}^T, \quad (i=1,\ldots,m) \tag{2.5b}$$

where $F^{(i)}$, $U^{(i)}$ and $V^{(i)}$ are scalar functions of x and y. The expressions of the jth 2-S vector and the jth 2-G vector are, respectively,

$$\mathbf{t}^{(j)}=\{\psi^{(j)}_{,y},-\psi^{(j)}_{,x}\}^T, \tag{2.6a}$$

$$\mathbf{h}^{(j)}=\{W^{(j)}_{,x},W^{(j)}_{,y}\}^T \quad (j=1,\ldots,n), \tag{2.6b}$$

where $\psi^{(j)}$ and $W^{(j)}$ are also scalar functions of x and y. Let

$$\mathbf{s}\equiv\{\mathbf{s}^{(1)T},\ldots,\mathbf{s}^{(m)T},\mathbf{t}^{(1)T},\ldots,\mathbf{t}^{(n)T}\}^T, \tag{2.7a}$$

$$\mathbf{g}\equiv\{\mathbf{g}^{(1)T},\ldots,\mathbf{g}^{(m)T},\mathbf{h}^{(1)T},\ldots,\mathbf{h}^{(n)T}\}^T. \tag{2.7b}$$

Then the following linear constitutive equation is postulated in terms of a symmetric, positive-definite matrix $[\tilde{\omega}]$ which may represent fully anisotropic, coupled response:

$$\mathbf{g}=[\tilde{\omega}]\mathbf{s}, \tag{2.8}$$

Consider the following two vectors composed of the anti-derivatives of the primary variables and the secondary variables, respectively:

$$\{\chi_1,\ldots,\chi_{2m+n}\}^T=\{F^{(1)}_{,y},-F^{(1)}_{,x},\ldots,F^{(m)}_{,y},-F^{(m)}_{,x},\psi^{(1)},\ldots,\psi^{(n)}\}^T, \tag{2.9a}$$

$$\{\chi_{2m+n+1},\ldots,\chi_{4m+2n}\}^T=\{U^{(1)},V^{(1)},\ldots,U^{(m)},V^{(m)},W^{(1)},\ldots,W^{(m)}\}^T. \tag{2.9b}$$

Notice that the first $2m$ elements of (2.9a) and (2.9b) occur in pairs, whose derivatives yield the components of (2.5a) and (2.5b), respectively. The derivatives of the last n elements of (2.9a) and (2.9b) yield the 2-S and 2-G vectors, respectively. Joining Eqs. (2.9a) and (2.9b), one obtains a vector function $\boldsymbol{\chi}$ of dimension $4m+2n$:

$$\boldsymbol{\chi}\equiv\{\chi_1,\ldots,\chi_{2m+n},\chi_{2m+n+1},\ldots,\chi_{4m+2n}\}^T. \tag{2.10}$$

Eigenrelation. We seek solutions $\boldsymbol{\chi}$ of the following form

$$\boldsymbol{\chi}=f(z,\mu)\boldsymbol{\xi}, \quad z\equiv x+\mu y, \tag{2.11}$$

where $\boldsymbol{\xi}$ is a complex constant vector, and f is a complex scalar function in which the complex parameter μ may occur both implicitly through $z=x+\mu y$ and explicitly as the second argument. Since $-F^{(i)}_{,yx}=-\xi^{(2i-1)}f_{,z}=-F^{(i)}_{,xy}=\xi^{(2i)}\mu f_{,z}$, the first m odd-numbered components of $\boldsymbol{\xi}$ may be expressed in terms of the first m even-numbered components, i.e.,

$$\xi^{(2i-1)}=-\mu\xi^{(2i)} \quad (i=1,\ldots,m). \tag{2.12}$$

Substitution of the spatial derivatives of (2.9a) and (2.9b) into Eqs. (2.5) and (2.6) gives

$$\mathbf{s}^{(i)}=\mathbf{s}_0\xi^{(2i)}f_{,z}, \tag{2.13a}$$

$$\mathbf{t}^{(j)}=\mathbf{t}_0\xi^{(2m+j)}f_{,z}, \tag{2.13b}$$

$$\mathbf{g}^{(i)}=\mathbf{g}_0\begin{Bmatrix}\xi^{(2m+n+2i-1)}\\ \xi^{(2m+n+2i)}\end{Bmatrix}f_{,z}, \tag{2.13c}$$

$$\mathbf{h}^{(j)}=\mathbf{h}_0\xi^{(4m+n+j)}f_{,z}, \tag{2.13d}$$

where

$$\mathbf{s}_0\equiv\{-\mu^2,-1,\mu\}^T, \tag{2.14a}$$

$$\mathbf{t}_0\equiv\{\mu,-1\}^T, \tag{2.14b}$$

$$\mathbf{h}_0\equiv\{1,\mu\}^T, \tag{2.14c}$$

$$\mathbf{g}_0\equiv\begin{bmatrix}1 & 0\\ 0 & \mu\\ \mu & 1\end{bmatrix}. \tag{2.14d}$$

Substituting Eqs. (2.13) and (2.14) into (2.7*a*) and (2.7*b*), one obtains

$$\mathbf{s}=f_{,z}\boldsymbol{\Psi}(\mu)\boldsymbol{\eta}, \tag{2.15a}$$

$$\mathbf{g}=f_{,z}\boldsymbol{\Theta}(\mu)\boldsymbol{\xi}_L \tag{2.15b}$$

where

$$\boldsymbol{\eta}\equiv\{\xi_2,\xi_4,\ldots,\xi_{2m},\xi_{2m+1},\ldots,\xi_{2m+n}\}^T, \tag{2.16a}$$

$$\boldsymbol{\xi}_L\equiv\{\xi_{2m+n+1},\ldots,\xi_{4m+2n}\}^T, \tag{2.16b}$$

i.e., $\boldsymbol{\xi}_L$ is the lower half of $\boldsymbol{\xi}$, whereas $\boldsymbol{\eta}$ is obtained by deleting the first m odd elements from the upper half of $\boldsymbol{\xi}$. Hence $\boldsymbol{\xi}_L$ and $\boldsymbol{\eta}$ have the dimensions $2m+n$ and $m+n$, respectively. $\boldsymbol{\Psi}(\mu)$ is a block-diagonal matrix formed by m identical blocks of 3×1 matrix $\mathbf{s}_0$ and n identical blocks of 2×1 matrix $\mathbf{t}_0$, and $\boldsymbol{\Theta}(\mu)$ is a block-diagonal matrix formed by m identical blocks of 3×2 matrix $\mathbf{g}_0$ and n identical blocks of 2×1 matrix $\mathbf{h}_0$. That is,

$$\boldsymbol{\Psi}_{(3m+2n)\times(m+n)}\equiv\langle\langle\mathbf{s}_0\rangle_m,\langle\mathbf{t}_0\rangle_n\rangle, \tag{2.17a}$$

$$\boldsymbol{\Theta}_{(3m+2n)\times(2m+n)}\equiv\langle\langle\mathbf{g}_0\rangle_m,\langle\mathbf{h}_0\rangle_n\rangle, \tag{2.17b}$$

where the notation $\langle\mathbf{s}_0\rangle_m$ stands for the block diagonal matrix containing m identical diagonal blocks of the submatrix $\mathbf{s}_0$, and $\langle\boldsymbol{\Phi}_1,\ldots,\boldsymbol{\Phi}_N\rangle$ indicates a block-diagonal matrix composed of N diagonal blocks $\boldsymbol{\Phi}_1,\ldots,\boldsymbol{\Phi}_N$. The subscripts of $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$ indicate the dimensions of the two matrices.

Since $\mathbf{t}_0$ is orthogonal to $\mathbf{h}_0$, and $\mathbf{s}_0$ is orthogonal to both columns of $\mathbf{g}_0$, it follows that each column of $\boldsymbol{\Psi}$ is orthogonal to all columns of $\boldsymbol{\Theta}$. Consequently,

$$\boldsymbol{\Psi}^T\boldsymbol{\Theta}=\mathbf{0},\quad \boldsymbol{\Theta}^T\boldsymbol{\Psi}=\mathbf{0}. \tag{2.18}$$

Substituting Eqs. (2.15*a*) and (2.15*b*) into (2.7), one obtains

$$\boldsymbol{\Theta}(\mu)\boldsymbol{\xi}_L=[\tilde{\omega}]\boldsymbol{\Psi}(\mu)\boldsymbol{\eta}. \tag{2.19}$$

Premultiplication of (2.19) by $\boldsymbol{\Theta}(\mu)^T[\tilde{\omega}]^{-1}$ and $\boldsymbol{\Psi}(\mu)^T$ yield, respectively [after using Eq. (2.18)],

$$\boldsymbol{\Theta}(\mu)^T[\tilde{\omega}]^{-1}\boldsymbol{\Theta}(\mu)\boldsymbol{\xi}_L=\mathbf{0}, \tag{2.20}$$

$$\mathbf{M}(\mu)\boldsymbol{\eta}=\mathbf{0}, \tag{2.21}$$

where

$$\mathbf{M}(\mu)\equiv\boldsymbol{\Psi}(\mu)^T[\tilde{\omega}]\boldsymbol{\Psi}(\mu) \tag{2.22}$$

is a symmetric matrix because $[\tilde{\omega}]$ is symmetric. $\mathbf{M}(\mu)$ is of central importance in the present theory, and it will be called the *eigenmatrix*. Since $\mathbf{M}(\mu)$ is a function, it is different from, but related to, the *constant* matrix of the same name usually found in the Stroh theory of plane elasticity. The relation between $\mathbf{M}(\mu)$ and the constitutive matrix $[\tilde{\omega}]$ may be clarified by writing

$$\mathbf{g}^{(i)}=\sum_{1\leqslant p\leqslant m}\mathbf{C}_{ip}\mathbf{s}^{(p)}+\sum_{1\leqslant q\leqslant n}\mathbf{E}_{iq}\mathbf{t}^{(q)}\quad(i=1,\ldots,m), \tag{2.23a}$$

$$\mathbf{h}^{(j)}=\sum_{1\leqslant p\leqslant m}\mathbf{E}_{jp}^T\mathbf{s}^{(p)}+\sum_{1\leqslant q\leqslant n}\mathbf{D}_{jq}\mathbf{t}^{(q)}\quad(j=1,\ldots,n). \tag{2.23b}$$

Then

$$[\tilde{\omega}]\equiv\begin{bmatrix}[\tilde{\omega}_{11}] & [\tilde{\omega}_{12}]\\ [\tilde{\omega}_{12}]^T & [\tilde{\omega}_{22}]\end{bmatrix}, \tag{2.24}$$

where $[\tilde{\omega}_{11}]$ is a $3m\times3m$ matrix with m^2 submatrices $\mathbf{C}_{ip}$ ($1\leqslant i,p\leqslant m$), $[\tilde{\omega}_{22}]$ is a $2n\times2n$ matrix with n^2 submatrices $\mathbf{D}_{jq}$ ($1\leqslant j,q\leqslant m$), and $[\tilde{\omega}_{12}]$ is a $3m\times2n$ matrix with mn submatrices $\mathbf{E}_{iq}$. The eigenmatrix $\mathbf{M}(\mu)$ is separated into four submatrices

$$\mathbf{M}(\mu)\equiv\begin{bmatrix}\mathbf{M}^{[4]}(\mu) & \mathbf{M}^{[3]}(\mu)\\ \mathbf{M}^{[3]}(\mu) & \mathbf{M}^{[2]}(\mu)\end{bmatrix}, \tag{2.25}$$

where $\mathbf{M}^{[4]}(\mu)$ is an $m\times m$ submatrix whose element in the ith row and pth column is

$$M_{ip}^{[4]}(\mu)=\{-\mu^2,-1,\mu\}\mathbf{C}_{ip}\{-\mu^2,-1,\mu\}^T\quad(1\leqslant i,p\leqslant m). \tag{2.26a}$$

$\mathbf{M}^{[2]}(\mu)$ is an $n\times n$ submatrix with the elements

$$M_{jq}^{[2]}(\mu)=\{\mu,-1\}\mathbf{D}_{jq}\{\mu,-1\}^T\quad(1\leqslant j,q\leqslant n), \tag{2.26b}$$

and $\mathbf{M}^{[3]}(\mu)$ has the dimension $m\times n$ and has elements of the form

$$M_{iq}^{[3]}(\mu)=\{-\mu^2,-1,\mu\}\mathbf{E}_{iq}\{\mu,-1\}^T\quad(1\leqslant i\leqslant m,1\leqslant q\leqslant n). \tag{2.26c}$$

Notice that the elements of $\mathbf{M}^{[4]}$, $\mathbf{M}^{[3]}$ and $\mathbf{M}^{[2]}$ are polynomials in μ of degrees 4, 3 and 2, respectively, and these elements have forms identical to the polynomials $l_4(\mu)$, $l_3(\mu)$ and $l_2(\mu)$, respectively, in Lekhnitskii's work in plane anisotropic elasticity.

3 Eigenvalues and Eigensolutions

Equation (2.19) has a nontrivial solution for $\boldsymbol{\eta}$ if and only if μ is a root of the characteristic equation

$$\delta(\mu)\equiv|\mathbf{M}(\mu)|=0. \tag{3.1}$$

It is clear from Eqs. (2.26*a*)–(2.26*c*) that $\delta(\mu)$ is a polynomial of degree $4m+2n$. Let $\mathbf{J}_1(\mu)$ be the block-diagonal matrix of dimension $(2m+n)\times(m+n)$, formed by m identical diagonal blocks of the matrix $-\mathbf{t}_0$ followed by the block $\mathbf{I}_n$:

$$\mathbf{J}_1(\mu)=\langle\langle-\mathbf{t}_0\rangle_m,\mathbf{I}_n\rangle. \tag{3.2}$$

Let $\mathbf{f}_0$ denote the 2×3 matrix having the first and second rows $\{1,0,0\}$ and $\{-\mu,0,1\}$, and let $\mathbf{J}_0(\mu)$ denote the following block diagonal matrix of dimension $(2m+n)\times(3m+2n)$:

$$\mathbf{J}_0\equiv\langle\langle\mathbf{f}_0\rangle_m,\langle[1,0]\rangle_n\rangle \tag{3.3}$$

Then it is easily verified that

$$\mathbf{J}_0(\mu)\boldsymbol{\Theta}(\mu)=\mathbf{I}_{2m+n}. \tag{3.4}$$

Define

$$\mathbf{J}_2(\mu)\equiv\mathbf{J}_0(\mu)[\tilde{\omega}]\boldsymbol{\Psi}(\mu), \tag{3.5a}$$

$$\mathbf{J}(\mu)\equiv\begin{Bmatrix}\mathbf{J}_1(\mu)\\ \mathbf{J}_2(\mu)\end{Bmatrix} \tag{3.5b}$$

Premultiplying Eq. (3.5*a*) by $\boldsymbol{\Theta}(\mu)$, one obtains, after some algebraic manipulation,

$$\boldsymbol{\Theta}(\mu)\mathbf{J}_2(\mu)=[\tilde{\omega}]\boldsymbol{\Psi}(\mu)+\mathbf{K}_0, \tag{3.6}$$

where $\mathbf{K}_0$ is a $(3m+2n)\times(m+n)$ matrix obtained from the eigenmatrix $\mathbf{M}(\mu)$ by inserting zero row vectors in the row positions $3j-2,3j(j=1,\ldots,m)$ and $3m+2k-1(k=1,\ldots,n)$. Hence, if $\boldsymbol{\eta}$ is a solution of Eq. (2.21), then $\mathbf{K}_0\boldsymbol{\eta}=\mathbf{0}$, so that Eq. (3.6) yields

$$\boldsymbol{\Theta}(\mu)\mathbf{J}_2(\mu)\boldsymbol{\eta}=[\tilde{\omega}]\boldsymbol{\Psi}(\mu)\boldsymbol{\eta}. \tag{3.7}$$

Premultiplying Eq. (2.19) by $\mathbf{J}_0(\mu)$, and using the last equation and Eqs. (3.4) and (3.5*a*), one has the following expression for the lower half of the vector $\boldsymbol{\xi}$:

$$\boldsymbol{\xi}_L = \mathbf{J}_3(\mu)[\tilde{\omega}]\boldsymbol{\Psi}(\mu)\boldsymbol{\eta} = \mathbf{J}_2(\mu)\boldsymbol{\eta}. \tag{3.8a}$$

On the other hand, Eqs. (2.12) and (3.2) give the upper half of $\boldsymbol{\xi}$:

$$\boldsymbol{\xi}_U \equiv \{\xi_1, \ldots, \xi_{2m+n}\}^T = \mathbf{J}_1(\xi)\boldsymbol{\eta}. \tag{3.8b}$$

The last two equations may be combined into one single expression, according to (3.5*b*):

$$\boldsymbol{\xi} = \mathbf{J}(\mu_k)\boldsymbol{\eta}. \tag{3.9}$$

If μ_k is a root of the characteristic equation (3.1), then Eq. (2.18) with $\mu = \mu_k$ has a nontrivial solution $\boldsymbol{\eta}$. The components of the vector $\boldsymbol{\xi} = \mathbf{J}(\mu_k)\boldsymbol{\eta}$ satisfy Eq. (2.15). Then, with the vector function $\boldsymbol{\chi}$ and its components defined by Eqs. (2.9)–(2.11), their spatial derivatives (2.5) and (2.6) satisfy the constitutive relation (2.8). The inherent equations governing the 3-S, 2-S, 3-G or 2-G vectors are automatically satisfied by virtue of the definitions of these vectors in terms of the potentials $F^{(i)}$, $\psi^{(j)}$, $U^{(i)}$, $V^{(i)}$ and $W^{(j)}$. Consequently, Eq. (2.11) yields the potentials of a two-dimensional field satisfying all field equations. We shall refer to the constant vector $\boldsymbol{\xi}$ and the vector-valued function $\boldsymbol{\chi}$, respectively, as a *zeroth-order eigenvector* and a *zeroth-order eigensolution* associated with the eigenvalue μ_k.

The eigenvalues occur in complex conjugate pairs because the characteristic equation has real coefficients. Hence there are $2m+n$ pairs of complex conjugate eigenvalues. Furthermore, $\overline{\boldsymbol{\xi}}$, the complex conjugate of $\boldsymbol{\xi}$, is an eigenvector associated with the conjugate eigenvalue $\overline{\mu}_0$. Then, according to Eq. (2.11), for any complex analytic function $f(x+\mu_0 y, \mu_0)$, the following expression

$$\boldsymbol{\chi} + \overline{\boldsymbol{\chi}} = 2\,\mathrm{Re}[f(x+\mu_0 y, \mu_0)\boldsymbol{\xi}] \tag{3.10}$$

yields real-valued potentials for Eqs. (3.8*a*) and (3.8*b*), whose spatial derivatives, as given by Eqs. (2.5) and (2.6), satisfy the constitutive relation and all inherent equations including equilibrium, compatibility, $\mathrm{div}(D_x\mathbf{i} + D_y\mathbf{j}) = 0$ and $\mathbf{curl}(E_x\mathbf{i} + E_y\mathbf{j}) = \mathbf{0}$.

Proof of the Complexity of Eigenvalues. We now show that, if the material has a positive-definite energy density, then the eigenvalues cannot be real. For suppose that Eq. (3.1) has a real root μ_0. Then $\mathbf{M}(\mu_0)$ is a real, singular matrix and Eq. (2.18), with $\mathbf{M}(\mu)$ replaced by $\mathbf{M}(\mu_0)$, must have a nontrivial real solution $\boldsymbol{\eta}$ (if $\boldsymbol{\eta}$ is a complex solution, then both the real and imaginary parts of $\boldsymbol{\eta}$ are also solutions) which yields a real eigenvector $\boldsymbol{\xi} = \mathbf{J}(\mu_0)\boldsymbol{\eta}$ because $\mathbf{J}(\mu_0)$ is also a real matrix. The choice $f \equiv x + \mu_0 y$ gives $f_{,z} \equiv 1$ and the real eigensolution $\boldsymbol{\chi} = (x+\mu_0 y)\boldsymbol{\xi}$. For this eigensolution, Eqs. (2.15) and (2.16) imply that the energy density vanishes:

$$(1/2)\mathbf{s}^T\mathbf{g} = (1/2)\boldsymbol{\eta}^T\boldsymbol{\Psi}^T\boldsymbol{\Theta}\{\xi_{2m+n+1}, \ldots, \xi_{4m+2n}\}^T = \mathbf{0}$$

for a nontrivial state $\mathbf{s}$ with

$$\mathbf{s}^T\mathbf{s} = \boldsymbol{\eta}^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}\boldsymbol{\eta} = (\mu_0^4 + \mu_0^2 + 1)(\xi_2^2 + \xi_4^2 + \ldots + \xi_{2m}^2) + (\mu_0^2+1)(\xi_{2m+1}^2 + \ldots + \xi_{2m+n}^2) > 0.$$

The contradiction proves the statement.

In the following, we assume that the energy density function is positive definite. Then there can be no real eigenvalues. We let $\{\mu\}_\perp$ denote the set of $2m+n$ eigenvalues with positive imaginary parts (not necessarily all distinct), and let $\{\overline{\mu}\}_\perp$ denote the complex conjugate set. The two sets are joined to form the complete set of $4m+2n$ eigenvalues $\{\mu\}$.

In this paper, we restrict the attention to the case where *all eigenvalues are distinct*. In this case, each eigenvalue has a (zeroth-order) eigenvector, and it will be shown that the $4m+2n$ eigenvectors form a linearly independent set to span the $(4m+2n)$-dimensional vector space. Then the corresponding eigensolutions give the two-dimensional general solution of the material. In the contrary case where the characteristic equation has repeated roots, the number of independent zeroth-order eigenvectors associated with a repeated root may or may not be equal to the multiplicity of the root, so that higher-order eigensolutions may be required to make up for the deficiency. Such cases are called degenerate of order d if the number of higher-order eigensolutions needed to supplement the zeroth-order eigensolutions is d. In a degenerate case, the solution space has a more general and complicated type of mathematical structure. The various degenerate cases, including the important case of isotropic materials, may be treated in a manner similar to [19,7,14] for the various subjects of plane anisotropic elasticity, piezoelectricity and laminated anisotropic plates.

4 General Solution of the Field Equations

Orthogonality of Eigenvectors. Equations (3.2), (3.3) and (3.5*a*) imply the relation

$$\mathbf{J}_1(\mu^*)^T\mathbf{J}_2(\mu) = (\mu^* - \mu)^{-1}\{\boldsymbol{\Psi}(\mu^*)^T - \boldsymbol{\Psi}(\mu)^T\}[\tilde{\omega}]\boldsymbol{\Psi}(\mu). \tag{4.1}$$

Due to the symmetry of $[\tilde{\omega}]$, one has

$$\begin{aligned}\mathbf{J}_1(\mu^*)^T\mathbf{J}_2(\mu) + \mathbf{J}_2(\mu^*)^T\mathbf{J}_1(\mu) &= (\mu^*-\mu)^{-1}\{\boldsymbol{\Psi}(\mu^*)^T[\tilde{\omega}]\boldsymbol{\Psi}(\mu^*) - \boldsymbol{\Psi}(\mu)^T[\tilde{\omega}]\boldsymbol{\Psi}(\mu^*) \\ &\quad + \boldsymbol{\Psi}(\mu^*)^T[\tilde{\omega}]\boldsymbol{\Psi}(\mu) - \boldsymbol{\Psi}(\mu)^T[\tilde{\omega}]\boldsymbol{\Psi}(\mu)\} \\ &= (\mu^*-\mu)^{-1}\{\mathbf{M}(\mu^*) - \mathbf{M}(\mu)\}.\end{aligned} \tag{4.2}$$

If μ and μ^* are distinct eigenvalues, with $\boldsymbol{\eta}$ and $\boldsymbol{\eta}'$ as the corresponding solutions of Eq. (2.21), then, by virtue of Eqs. (3.8*a*) and (3.8*b*), the eigenvectors $\boldsymbol{\xi} = \mathbf{J}(\mu)\boldsymbol{\eta}$ and $\boldsymbol{\xi}' = \mathbf{J}(\mu^*)\boldsymbol{\eta}'$ satisfy

$$[\![\boldsymbol{\xi}', \boldsymbol{\xi}]\!] \equiv \boldsymbol{\xi}'^T\mathbf{II}\boldsymbol{\xi} = \boldsymbol{\eta}'^T\{\mathbf{J}_1(\mu^*)^T\mathbf{J}_2(\mu) + \mathbf{J}_2(\mu^*)^T\mathbf{J}_1(\mu)\}\boldsymbol{\eta} = 0, \tag{4.3}$$

where

$$\mathbf{II} = \begin{bmatrix} \mathbf{0}_{(2m+n)\times(2m+n)} & \mathbf{I}_{2m+n} \\ \mathbf{I}_{2m+n} & \mathbf{0}_{(2m+n)\times(2m+n)} \end{bmatrix}, \tag{4.4}$$

and the subscripts of the zero matrices indicate their row and column dimensions. Clearly

$$\mathbf{II}\,\mathbf{II} = \mathbf{I}_{4m+2n}.$$

Hence any two zeroth-order eigenvectors $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ associated with distinct eigenvalues are orthogonal in the sense of Eq. (4.3). In particular, $\boldsymbol{\xi}$ is orthogonal to its complex conjugate vector:

$$[\![\boldsymbol{\xi}, \overline{\boldsymbol{\xi}}]\!] = \boldsymbol{\xi}^T\mathbf{II}\overline{\boldsymbol{\xi}} = 0. \tag{4.5}$$

In the following, $[\![\boldsymbol{\xi}', \boldsymbol{\xi}]\!] = [\![\boldsymbol{\xi}, \boldsymbol{\xi}']\!]$ will be called the binary product of the two vectors $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$. Orthogonality of eigenvectors is always defined in the sense of the binary product.

Complete Set of Eigenvectors and Eigensolutions. The binary product as defined for two vectors by the first equality of Eq. (4.3) may be extended to two matrices of the same row dimension $4m+2n$, regardless of their column dimensions. Equation (4.2) becomes

$$(\mu^* - \mu)[\![\mathbf{J}(\mu^*), \mathbf{J}(\mu)]\!] = \mathbf{M}(\mu^*) - \mathbf{M}(\mu), \tag{4.6}$$

Differentiation with respect to μ gives

$$(\mu^*-\mu)[\![\mathbf{J}(\mu^*), \mathbf{J}'(\mu)]\!] - [\![\mathbf{J}(\mu^*), \mathbf{J}(\mu)]\!] = -\mathbf{M}'(\mu). \tag{4.7}$$

Setting $\mu^* = \mu$, the last equation becomes

$$[\![\mathbf{J}(\mu), \mathbf{J}(\mu)]\!] = \mathbf{M}'(\mu). \tag{4.8}$$

The matrix $\mathbf{M}(\mu)$ and its adjoint matrix $\mathbf{W}(\mu)$ satisfy the polynomial identity

$$\mathbf{M}(\mu)\mathbf{W}(\mu)=\mathbf{W}(\mu)\mathbf{M}(\mu)=\delta(\mu)\mathbf{I}_{m+n}. \tag{4.9}$$

Differentiation yields

$$\mathbf{M}'(\mu)\mathbf{W}(\mu)+\mathbf{M}(\mu)\mathbf{W}'(\mu)=\delta'(\mu)\mathbf{I}_{m+n}. \tag{4.10}$$

If all eigenvalues μ_k are simple roots of the characteristic equation, then each matrix $\mathbf{M}(\mu_k)$ is of rank $m+n-1$. Equation (4.9) becomes

$$\mathbf{M}(\mu_k)\mathbf{W}(\mu_k)=\mathbf{0}. \tag{4.11}$$

Since the symmetric matrix $\mathbf{M}(\mu_k)$ has $m+n-1$ independent columns, and (4.11) implies that each column of $\mathbf{W}(\mu_k)$ has vanishing scalar products with all columns of $\mathbf{M}(\mu_k)$, it follows that $\mathbf{W}(\mu_k)$ has only one independent column, i.e., all columns of $\mathbf{W}(\mu_k)$ are proportional. Then it is clear that $\mathbf{W}(\mu_k)$ must have at least one nonvanishing diagonal element $W_k\equiv\boldsymbol{\rho}_j^T\mathbf{W}(\mu_k)\boldsymbol{\rho}_j$ at the jth diagonal position. Hence $\boldsymbol{\eta}_k\equiv\mathbf{W}(\mu_k)\boldsymbol{\rho}_j$ is a nontrivial vector, where $\boldsymbol{\rho}_j$ is the jth column of $\mathbf{I}_{m+n}$. Furthermore,

$$\mathbf{W}(\mu_k)=\mathbf{W}(\mu_k)\boldsymbol{\rho}_j\boldsymbol{\rho}_j^T\mathbf{W}(\mu_k)/W_k, \tag{4.12}$$

Equation (4.11) ensures that the nontrivial vector $\boldsymbol{\eta}_k\equiv\mathbf{W}(\mu_k)\boldsymbol{\rho}_j$ is a solution of Eq. (2.18) with $\mu=\mu_k$. This yields an eigenvector and an eigensolution for each eigenvalue μ_k in $\{\mu\}_\perp$:

$$\boldsymbol{\xi}_k=\mathbf{J}(\mu_k)\mathbf{W}(\mu_k)\boldsymbol{\rho}_j, \tag{4.13a}$$

$$\boldsymbol{\chi}_k=f_k(x+\mu_k y,\mu_k)\mathbf{J}(\mu_k)\mathbf{W}(\mu_k)\boldsymbol{\rho}_j. \tag{4.13b}$$

Since the eigenvalues are distinct, all eigenvectors may be obtained this way. Let

$$\mathbf{Z}_\perp=[\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{2m+n}], \tag{4.14a}$$

$$\mathbf{Z}=[\mathbf{Z}_\perp,\overline{\mathbf{Z}}_\perp]. \tag{4.14b}$$

The matrix $\mathbf{Z}$ contains the complete set of eigenvectors, chosen according to Eq. (4.13*a*) and arranged in the order of the corresponding μ_k in $\{\mu\}$. $\mathbf{Z}$ is called the *base matrix*. The eigenvalues and the base matrix together determine the two-dimensional general solution of the material:

$$\begin{aligned}\boldsymbol{\chi}&=\mathbf{Z}\langle\langle f_1(x+\mu_1 y,\mu_1),\ldots,f_{2m+n}(x+\mu_{2m+n}y,\mu_{2m+n})\rangle,\#\rangle\mathbf{c}\\&=\mathrm{Re}[2\mathbf{Z}_\perp\langle f_1(x+\mu_1 y,\mu_1),\ldots,\\&\quad f_{2m+n}(x+\mu_{2m+n}y,\mu_{2m+n})\rangle\mathbf{c}_\perp],\end{aligned} \tag{4.15a}$$

$$\mathbf{c}=\begin{Bmatrix}\mathbf{c}_\perp\\ \overline{\mathbf{c}}_\perp\end{Bmatrix}. \tag{4.15b}$$

Here $\mathbf{c}_\perp$ is an arbitrary complex constant vector of dimension $2m+n$. For any matrix $\boldsymbol{\sigma}$, $\langle\boldsymbol{\sigma},\#\rangle$ denotes the block diagonal matrix composed of two complex conjugate diagonal blocks $\boldsymbol{\sigma}$ and $\overline{\boldsymbol{\sigma}}$. The constant vector $\mathbf{c}$ in Eq. (4.15*a*) is redundant since it can be absorbed into the arbitrary analytic functions f_k. However, in many important problems, the functions f_k differ from one another only by constant multiplicative factors. Then it is convenient to use the same function form, and combine all multiplicative factors into an undeterminepd constant vector $\mathbf{c}$.

It will be shown in the next section that the base matrix $\mathbf{Z}$ is nonsingular, i.e., the eigenvectors in $\mathbf{Z}$ are linearly independent. Hence Eq. (4.15) may be rewritten as

$$\boldsymbol{\chi}=\mathbf{Z}\|f\|\mathbf{Z}^{-1}\mathbf{b}, \tag{4.16}$$

where

$$\|f\|\equiv\langle\|f\|_\perp,\#\rangle, \tag{4.17a}$$

$$\|f\|_\perp\equiv\langle f_1(x+\mu_1 y,\mu_1),\ldots,f_{2m+n}(x+\mu_{2m+n}y,\mu_{2m+n})\rangle, \tag{4.17b}$$

and $\mathbf{b}\equiv\mathbf{Z}\mathbf{c}$ is a real constant vector in view of Eqs. (4.14*b*) and (4.15*b*). Hence $\mathbf{Z}\|f\|\mathbf{Z}^{-1}$ is a real-valued function.

With the potentials of the general solution given by Eqs. (4.15*a*) and (4.15*b*), the physical variables of the general solution are obtained by taking spatial derivatives. Let $\boldsymbol{\chi}_U$ and $\boldsymbol{\chi}_L$ denote, respectively, the upper and lower halves of $\boldsymbol{\chi}$. Then

$$\begin{aligned}\partial_x\boldsymbol{\chi}_U&=\{F^{(1)}_{,xy},-F^{(1)}_{,xx},\ldots,F^{(m)}_{,xy},-F^{(m)}_{,xx},\psi^{(1)}_{,x},\ldots,\psi^{(n)}_{,x}\}^T\\&=[\mathbf{I}_{2m+n},\mathbf{0}_{(2m+n\}\times(2m+n\}}]\mathbf{Z}\|f_{,z}\|\mathbf{Z}^{-1}\mathbf{b},\end{aligned} \tag{4.18a}$$

$$\begin{aligned}\partial_y\boldsymbol{\chi}_U&=\{F^{(1)}_{,yy},-F^{(1)}_{,xy},\ldots,F^{(m)}_{,yy},-F^{(m)}_{,xy},\psi^{(1)}_{,y},\ldots,\psi^{(n)}_{,y}\}^T\\&=[\mathbf{I}_{2m+n},\mathbf{0}_{(2m+n\}\times(2m+n\}}]\mathbf{Z}\|\mu f_{,z}\|\mathbf{Z}^{-1}\mathbf{b},\end{aligned} \tag{4.18b}$$

$$\begin{aligned}\partial_x\boldsymbol{\chi}_L&=\{U^{(1)}_{,x},V^{(1)}_{,x},\ldots,U^{(m)}_{,x},V^{(m)}_{,x},W^{(1)}_{,x},\ldots,W^{(m)}_{,x}\}^T\\&=[\mathbf{0}_{(2m+n\}\times(2m+n\}},\mathbf{I}_{2m+n}]\mathbf{Z}\|f\|_{,z}\mathbf{Z}^{-1}\mathbf{b},\end{aligned} \tag{4.19a}$$

$$\begin{aligned}\partial_y\boldsymbol{\chi}_L&=\{U^{(1)}_{,y},V^{(1)}_{,y},\ldots,U^{(m)}_{,y},V^{(m)}_{,y},W^{(1)}_{,y},\ldots,W^{(m)}_{,y}\}^T\\&=[\mathbf{0}_{(2m+n\}\times(2m+n\}},\mathbf{I}_{2m+n}]\mathbf{Z}\|\mu f_{,z}\|\mathbf{Z}^{-1}\mathbf{b},\end{aligned} \tag{4.19b}$$

where $f_{,z}$ denotes the derivative of $f(x+\mu y,\mu)$ with respect to the first argument $z=x+\mu y$, evaluated at $\mu=\mu_k$. As the components of two-dimensional tensors or vectors, all variables given by Eqs. (4.18*a*) and (4.18*b*) have direct physical meanings. However, the variables $V^{(k)}_{,x}$ and $U^{(k)}_{,y}$ in (4.19*a*) and (4.19*b*) are not tensor components unless they are combined.

Taking the binary product of $\boldsymbol{\xi}_k$ with itself, and using Eqs. (4.8) and (4.10), one has

$$\begin{aligned}[\![\boldsymbol{\xi}_k,\boldsymbol{\xi}_k]\!]&=\boldsymbol{\rho}_j^T\mathbf{W}(\mu_k)[\![\mathbf{J}(\mu_k),\mathbf{J}(\mu_k)]\!]\mathbf{W}(\mu_k)\boldsymbol{\rho}_j=\boldsymbol{\rho}_j^T\mathbf{W}\mathrm{M}'\mathbf{W}\boldsymbol{\rho}_j\\&=\boldsymbol{\rho}_j^T\mathbf{W}\{\delta'\mathbf{I}_{m+n}-\mathbf{M}\mathbf{W}'\}\boldsymbol{\rho}_j=\delta'\boldsymbol{\rho}_j^T\mathbf{W}\boldsymbol{\rho}_j=\delta'(\mu_k)W_k\neq 0.\end{aligned} \tag{4.20}$$

Here $\delta'(\mu_k)$ does not vanish because μ_k is a simple root of $\delta(\mu)=0$, and W_k has been defined as a nonvanishing diagonal element of $\mathbf{W}(\mu_k)$.

Equations (4.3), (4.14) and (4.18) imply that the binary product of $\mathbf{Z}_\perp$ with itself is a diagonal matrix of the following form:

$$\boldsymbol{\Omega}_\perp\equiv[\![\mathbf{Z}_\perp,\mathbf{Z}_\perp]\!]=\langle\delta'(\mu_1)W_1,\ldots,\delta'(\mu_{2m+n})W_{2m+n}\rangle. \tag{4.21}$$

Furthermore, the binary product of $\mathbf{Z}_\perp$ with $\overline{\mathbf{Z}}_\perp$ yields the null matrix in view of Eq. (4.5). Hence

$$\begin{aligned}\boldsymbol{\Omega}\equiv[\![\mathbf{Z},\mathbf{Z}]\!]&=\mathbf{Z}^T\mathbf{I}\mathbf{I}\mathbf{Z}=\langle\langle\delta'(\mu_1)W_1,\ldots,\delta'(\mu_{2m+n})W_{2m+n}\rangle,\#\rangle\\&=\langle\boldsymbol{\Omega}_\perp,\#\rangle.\end{aligned} \tag{4.22}$$

Since $(Det[\mathbf{II}])^2=Det[\mathbf{I}_{4m+2n}]=1$, the last equation implies that

$$(Det[\mathbf{Z}])^2=\Pi_{1\leqslant k\leqslant 2m+n}|\delta'(\mu_k)W_k|^2\neq 0. \tag{4.23}$$

Hence the base matrix $\mathbf{Z}$ is nonsingular. Its $4m+2n$ columns form a complete set of *independent* vectors in the complex vector space of dimension $4m+2n$.

5 Intrinsic Tensors and Transformation Rules

Equations (4.12) and (4.13*a*) yield the following symmetric matrix with the real and imaginary parts $\mathbf{F}_k$ and $\mathbf{G}_k$:

$$\{\delta'(\mu_k)W_k\}^{-1}\boldsymbol{\xi}_k\boldsymbol{\xi}_k^T=\mathbf{J}(\mu_k)\mathbf{W}(\mu_k)\mathbf{J}(\mu_k)^T/\delta'(\mu_k)\equiv\mathbf{F}_k+i\mathbf{G}_k, \tag{5.1}$$

From Eqs. (4.3), (4.14) and (5.1) one obtains

$$(\mathbf{F}_k+i\mathbf{G}_k)\mathbf{I}\mathbf{I}\boldsymbol{\xi}_j=\mathbf{0}\quad\text{if } k\neq j;\qquad(\mathbf{F}_k+i\mathbf{G}_k)\mathbf{I}\mathbf{I}\boldsymbol{\xi}_k=\boldsymbol{\xi}_k, \tag{5.2}$$

so that $(\mathbf{F}_k+i\mathbf{G}_k)\mathbf{II}$ is the projection operator into the one-dimensional subspace of the eigenvector $\boldsymbol{\xi}_k$. The complex conjugate of Eq. (5.1) implies that $(\mathbf{F}_k-i\mathbf{G}_k)\mathbf{II}$ is the projection operator into the one-dimensional subspace of $\overline{\boldsymbol{\xi}}_k$. This yields the decomposition of the identity transformation into orthogonal projections:

$$\mathbf{I}_{4m+2n}=\sum_{1\leqslant k\leqslant 2m+n}(\mathbf{F}_k+i\mathbf{G}_k)\mathbf{\Pi}+\sum_{1\leqslant k\leqslant 2m+n}(\mathbf{F}_k-i\mathbf{G}_k)\mathbf{\Pi}. \tag{5.3}$$

Hence

$$2\sum_{1\leqslant k\leqslant 2m+n}\mathbf{F}_k=\mathbf{\Pi}, \tag{5.4}$$

and

$$\mathbf{\Gamma}\equiv 2\sum_{1\leqslant k\leqslant 2m+n}\mathbf{G}_k. \tag{5.5}$$

is a real symmetric matrix. Then

$$\mathbf{Z}\langle\mathbf{\Omega}_\perp^{-1},\overline{\mathbf{\Omega}}_\perp^{-1}\rangle\mathbf{Z}^T=\mathbf{\Pi}, \tag{5.6a}$$

$$\mathbf{Z}\langle-i\mathbf{\Omega}_\perp^{-1},i\overline{\mathbf{\Omega}}_\perp^{-1}\rangle\mathbf{Z}^T=\mathbf{\Gamma}. \tag{5.6b}$$

Furthermore, the following matrices are the projection operators into the multi-dimensional subspaces spanned, respectively, by the eigenvectors in $\mathbf{Z}_\perp$ and $\overline{\mathbf{Z}}_\perp$:

$$1/2(\mathbf{\Pi}+i\mathbf{\Gamma})\mathbf{\Pi}=\sum_{1\leqslant k\leqslant 2m+n}(\mathbf{F}_k+i\mathbf{G}_k)\mathbf{\Pi}, \tag{5.7a}$$

$$1/2(\mathbf{\Pi}-i\mathbf{\Gamma})\mathbf{\Pi}=\sum_{1\leqslant k\leqslant 2m+n}(\mathbf{F}_k-i\mathbf{G}_k)\mathbf{\Pi}. \tag{5.7b}$$

Hence one has the following equalities and their complex conjugates

$$1/2(\mathbf{\Pi}+i\mathbf{\Gamma})\mathbf{\Pi}\mathbf{Z}_\perp=\mathbf{Z}_\perp,\quad 1/2(\mathbf{\Pi}+i\mathbf{\Gamma})\mathbf{\Pi}\overline{\mathbf{Z}}_\perp=\mathbf{0}.$$

Clearly,

$$\mathbf{\Gamma}\mathbf{\Pi}\mathbf{Z}_\perp=-i\mathbf{Z}_\perp,\quad \mathbf{\Gamma}\mathbf{\Pi}\overline{\mathbf{Z}}_\perp=i\overline{\mathbf{Z}}_\perp. \tag{5.8}$$

Hence the matrix $\mathbf{\Gamma\Pi}$ has $-i$ and $+i$ as eigenvalues of multiplicity $2m+n$, and the corresponding eigenvectors are the column of $\mathbf{Z}_\perp$ and $\overline{\mathbf{Z}}_\perp$, respectively, i.e.,

$$\mathbf{\Gamma}\mathbf{\Pi}\mathbf{Z}=\mathbf{Z}\langle-i\mathbf{I}_{2m+n},i\mathbf{I}_{2m+n}\rangle, \tag{5.9}$$

which also follows directly from Eq. (5.6*b*). Postmultiplying Eq. (5.9) by $\langle\mathbf{\Omega}_\perp^{-1},-\overline{\mathbf{\Omega}}_\perp^{-1}\rangle\mathbf{Z}^T$, and using (5.6*a*) and (5.6*b*), one obtains $i\mathbf{\Gamma\Pi\Pi\Gamma}=\mathbf{Z}\langle-i\mathbf{\Omega}_\perp^{-1},-i\overline{\mathbf{\Omega}}_\perp^{-1}\rangle\mathbf{Z}^T=-i\mathbf{\Pi}$, i.e.,

$$\mathbf{\Gamma}\mathbf{\Pi}\mathbf{\Gamma}\mathbf{\Pi}=-\mathbf{I}_{4m+2n}. \tag{5.10}$$

The real symmetric matrix $\mathbf{\Gamma}$ has special importance in crack problems and in the fundamental solutions (Green's functions) of various regions. If $\mathbf{\Gamma}$ and the base matrix $\mathbf{Z}$ are each separated into four submatrices of dimension $(2m+n)\times(2m+n)$

$$\mathbf{\Gamma}\equiv\begin{bmatrix}-\mathbf{L} & \mathbf{S}^T\\ \mathbf{S} & \mathbf{H}\end{bmatrix}, \tag{5.11a}$$

$$\mathbf{Z}=\begin{bmatrix}\mathbf{B} & \overline{\mathbf{B}}\\ \mathbf{A} & \overline{\mathbf{A}}\end{bmatrix}, \tag{5.11b}$$

then $\mathbf{L}$ and $\mathbf{H}$ are symmetric but $\mathbf{S}$ is generally not. Equation (5.10) yields

$$\mathbf{HL}-\mathbf{SS}=\mathbf{LH}-\mathbf{S}^T\mathbf{S}^T=\mathbf{I}_{2m+n}, \tag{5.12}$$

$$\mathbf{LS}=-(\mathbf{LS})^T,\quad \mathbf{SH}=-(\mathbf{SH})^T. \tag{5.13}$$

This implies that $\mathbf{LS}$ and $\mathbf{SH}$ are skew-symmetric. Equations (4.20) and (5.11*b*) give

$$\mathbf{\Omega}_\perp=\mathbf{A}^T\mathbf{B}+\mathbf{B}^T\mathbf{A}, \tag{5.14a}$$

$$\mathbf{A}^T\overline{\mathbf{B}}+\mathbf{B}^T\overline{\mathbf{A}}=\mathbf{0}. \tag{5.14b}$$

From Eqs. (5.6*a*) and (5.6*b*) follow

$$\mathbf{L}=2i\mathbf{B}\mathbf{\Omega}_\perp^{-1}\mathbf{B}^T, \tag{5.15a}$$

$$\mathbf{H}=-2i\mathbf{A}\mathbf{\Omega}_\perp^{-1}\mathbf{A}^T, \tag{5.15b}$$

$$\mathbf{S}=-i(2\mathbf{A}\mathbf{\Omega}_\perp^{-1}\mathbf{B}^T-\mathbf{I}_{2m+n}), \tag{5.15c}$$

and Eq. (5.8) gives

$$\mathbf{LA}=(\mathbf{S}^T+i\mathbf{I}_{2m+n})\mathbf{B}, \tag{5.16a}$$

$$\mathbf{HB}=-(\mathbf{S}+i\mathbf{I}_{2m+n})\mathbf{A}. \tag{5.16b}$$

In 2-D elasticity, where $m=n=1$, $\mathbf{L}$, $\mathbf{H}$ and $\mathbf{S}$ are 3×3 matrices called Barnett-Lothe tensors, although one of the tensors appeared earlier in Stroh's work [2]. If $\mathbf{\Gamma}$ is found, then any one of the four submatrices of $\mathbf{Z}$ determines the other three, and hence also $\mathbf{Z}$.

There are two ways to obtain $\mathbf{\Gamma}$ explicitly. One is using Eqs. (5.1) and (5.5). This yields $\mathbf{\Pi}+i\mathbf{\Gamma}$ as the spectral sum of the contributions from all eigenmodes:

$$\mathbf{\Pi}+i\mathbf{\Gamma}=\sum_{1\leqslant k\leqslant 2m+n}2\mathbf{J}(\mu_k)\mathbf{W}(\mu_k)\mathbf{J}(\mu_k)^T/\delta'(\mu_k). \tag{5.17}$$

The other way is using Eq. (5.6*b*) in terms of the base matrix $\mathbf{Z}$ and

$$\mathbf{\Omega}_\perp^{-1}=\langle 1/\{\delta'(\mu_1)W_1\},\ldots,1/\{\delta'(\mu_{2m+n})W_{2m+n}\}\rangle. \tag{5.18}$$

Equation (5.17) is an intrinsic expression independent of the arbitrary scalar multiplicative factors contained in the eigenvectors. Equations (5.6) and (5.18) are easier to use, and, although $\mathbf{\Omega}_\perp^{-1}$ and $\mathbf{Z}_\perp$ both depend on the choice of eigenvectors, $\mathbf{\Gamma}$ does not.

An analytical expression of $\mathbf{Z}^{-1}$ is often needed in the exact expressions of Green's functions and in the solutions of multi-material wedge problems [20]. With $\mathbf{\Omega}_\perp^{-1}$ given by Eq. (5.18), the following expressions are obtained from Eqs. (5.6*a*), (5.6*b*) and (5.9):

$$\mathbf{Z}^{-1}=\langle\mathbf{\Omega}_\perp^{-1},\overline{\mathbf{\Omega}}_\perp^{-1}\rangle\mathbf{Z}^T\mathbf{\Pi}=\langle i\mathbf{\Omega}_\perp^{-1},-i\overline{\mathbf{\Omega}}_\perp^{-1}\rangle\mathbf{Z}^T\mathbf{\Pi}\mathbf{\Gamma}\mathbf{\Pi}. \tag{5.19}$$

Rotational Transformation. We define the matrices associated with two-dimensional rotation

$$\mathbf{Q}_2\equiv\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix},\quad \mathbf{Q}_{4m+2n}\equiv\langle\mathbf{Q}_2\rangle_{2m+n} \tag{5.20}$$

Under a coordinate transformation $\{x^*,y^*\}^T=\mathbf{Q}_2\{x,y\}^T$, the eigenvectors and the base matrix $\mathbf{Z}$ change in the following manner:

$$\boldsymbol{\xi}^*=\mathbf{Q}_{4m+2n}\boldsymbol{\xi}, \tag{5.21a}$$

$$\mathbf{Z}^*=\mathbf{Q}_{4m+2n}\mathbf{Z}. \tag{5.21b}$$

The binary product is invariant under the rotation transformation,

$$[\![\boldsymbol{\xi}_j^*,\boldsymbol{\xi}_k^*]\!]=[\![\boldsymbol{\xi}_j,\boldsymbol{\xi}_k]\!] \tag{5.22}$$

and it follows that $\mathbf{\Omega}_\perp=[\![\mathbf{Z}_\perp,\overline{\mathbf{Z}_\perp}]\!]$ is rotation invariant. Equations (4.12) implies that

$$\mathbf{F}_k^*=\mathbf{Q}_{4m+2n}\mathbf{F}_k\mathbf{Q}_{4m+2n}^T,\quad \mathbf{G}_k^*=\mathbf{Q}_{4m+2n}\mathbf{G}_k\mathbf{Q}_{4m+2n}^T. \tag{5.23}$$

This tensorial transformation rule must also be satisfied by $\mathbf{\Gamma}$.

Affine Transformation. Consider a nonsingular linear transformation, $\boldsymbol{\tau}$: $\mathbf{Z}_\perp\to(\mathbf{Z}_\perp)^*=\mathbf{Z}_\perp\boldsymbol{\tau}$, in the $(2m+n)$-dimensional space spanned by the column vectors of $\mathbf{Z}_\perp$, and the conjugate transformation $\overline{\boldsymbol{\tau}}$: $\overline{\mathbf{Z}}_\perp\to(\overline{\mathbf{Z}}_\perp)^*=\overline{\mathbf{Z}}_\perp\overline{\boldsymbol{\tau}}$. One has $\mathbf{Z}^*=\mathbf{Z}\langle\boldsymbol{\tau},\overline{\boldsymbol{\tau}}\rangle$ and $\mathbf{Z}^{*-1}=\langle\boldsymbol{\tau}^{-1},\overline{\boldsymbol{\tau}}^{-1}\rangle\mathbf{Z}^{-1}$. Equation (5.9) yields

$$\mathbf{\Gamma}=\mathbf{Z}\langle-i\mathbf{I},i\mathbf{I}\rangle\mathbf{Z}^{-1}\mathbf{\Pi}\to$$

$$\mathbf{\Gamma}^*=\mathbf{Z}^*\langle-i\mathbf{I},i\mathbf{I}\rangle\mathbf{Z}^{*-1}\mathbf{\Pi}=\mathbf{Z}\langle\boldsymbol{\tau},\overline{\boldsymbol{\tau}}\rangle\langle-i\mathbf{I},i\mathbf{I}\rangle\langle\boldsymbol{\tau}^{-1},\overline{\boldsymbol{\tau}}^{-1}\rangle\mathbf{Z}^{-1}\mathbf{\Pi}=\mathbf{\Gamma},$$

i.e., the real symmetric tensor $\boldsymbol{\Gamma}$ is invariant under an arbitrary affine transformation of the base vectors that preserves the complex conjugate relation. This implies, in particular, that $\boldsymbol{\Gamma}$ can be determined from any set of independent eigenvectors, and the practice of normalizing eigenvectors (usually performed in the Stroh formalism of anisotropic elasticity) is superfluous.

6 Green's Function of an Infinite Domain (Fundamental Singularity Solution)

Consider a solution $\boldsymbol{\chi}$ of the infinite two-dimensional domain that satisfies the requirements:

(i) $\mathbf{s}$ and $\mathbf{g}$ vanish at infinity;

(ii) $\boldsymbol{\chi}$ has a constant discontinuity $\boldsymbol{\chi}_0$ across the negative x-axis, i.e., in polar coordinates,

$$[\boldsymbol{\chi}] \equiv \oint d\boldsymbol{\chi} = \boldsymbol{\chi}(r,\pi) - \boldsymbol{\chi}(r,-\pi) = \boldsymbol{\chi}_0, \tag{6.1}$$

where the integral is along a closed path encircling the origin in the counterclockwise sense. We let

$$\boldsymbol{\chi} = (2\pi)^{-1}\mathbf{G}_\infty \boldsymbol{\chi}_0. \tag{6.2}$$

Then the matrix function $(2\pi)^{-1}\mathbf{G}_\infty$ of dimension $(4m+2n)\times(4m+2n)$ is called Green's function of the infinite domain with a singularity at the origin, and the constant vector $\boldsymbol{\chi}_0$ is called the strength of singularity. By setting

$$f(z,\mu) = (2\pi i)^{-1}\log[z], \tag{6.3}$$

Eq. (4.16) gives the function $\boldsymbol{\chi}$ of Eq. (6.2). Hence

$$\mathbf{G}_\infty = \mathbf{Z}\| -i\log[x+\mu y]\|\mathbf{Z}^{-1}, \tag{6.4}$$

On the upper and lower sides of the negative real axis, the function $-i\log[z]$ has the values $-i(\log[r]+i\pi)$ and $-i(\log[r]-i\pi)$, respectively, since $\mathrm{Im}[\mu]>0$. Consequently, both $-i\log[z]$ and its complex conjugate have the same constant jump 2π across the negative real axis. Hence Green's function $(2\pi)^{-1}\mathbf{G}_\infty$ has a constant discontinuity $\mathbf{I}_{4m+2n}$ across any branch cut emanating from the origin.

The 3-S vectors $\mathbf{s}^{(i)} = \{F^{(i)}_{,yy}, F^{(i)}_{,xx}, -F^{(i)}_{,xy}\}^T$ contained in the multi-dimensional vector $\mathbf{s}$ may be kinetic (stress field in plane elasticity) and may also be kinematical (e.g., curvatures in the plate bending problem). In the first case, discontinuities of $F^{(i)}_{,y}$ and of $-F^{(i)}_{,x}$ correspond, respectively, to concentrated forces in the x- and y-directions. In the second case, the discontinuities correspond to dislocations in the inclinations of the deformed plate surface. All such discontinuities, as well as the others associated with the 2-S vectors, appear as the first $2m+n$ elements of $\boldsymbol{\chi}_0 = \oint d\boldsymbol{\chi}$. For the 3-G vectors $\mathbf{g}^{(i)} = \{U^{(i)}_{,x}, V^{(i)}_{,y}, U^{(i)}_{,y} + V^{(i)}_{,x}\}^T$, discontinuities in $U^{(i)}$ and $V^{(i)}$ imply dislocations in the displacements of plane elasticity, or in the moment potentials of the plate bending problem, etc. For the 2-G vector $\mathbf{h}^{(j)} = \{W^{(j)}_{,x}, W^{(j)}_{,y}\}^T$, a discontinuity in $W^{(j)}$ corresponds to a dislocation in the antiplane displacement. All such discontinuities appear as the last $2m+n$ elements of $\boldsymbol{\chi}_0$. The first $2m+n$ columns of Green's function $(2\pi)^{-1}\mathbf{G}_\infty$ are the response functions of the infinite plane to the first group of elements of $\boldsymbol{\chi}_0$, i.e., discontinuities in $F^{(i)}_{,y}$, $-F^{(i)}_{,x}$ and $\psi^{(i)}$, whereas the last $2m+n$ columns of $(2\pi)^{-1}\mathbf{G}_\infty$ are the response functions to discontinuities in $U^{(i)}$, $V^{(i)}$ and $W^{(j)}$.

Since $\log[z] = \log[r] + \log[\cos\theta + \mu\sin\theta]$, Eq. (6.4) may be rewritten by using (5.9):

$$\mathbf{G}_\infty = \log[r]\boldsymbol{\Gamma}\mathbf{II} + \boldsymbol{\Gamma}\mathbf{II}\|\log[\cos\theta + \mu\sin\theta]\|\mathbf{Z}^{-1}. \tag{6.5}$$

Thus, in polar coordinates, $\mathbf{G}_\infty$ is separated into functions of r and θ.

Combining Eqs. (4.18) and (4.19), for the function $\boldsymbol{\chi}$ of Eqs. (6.2) and (6.4), one has

$$2\pi\boldsymbol{\chi}_{,x} = \mathbf{Z}\langle\langle -i/(x+\mu_1 y), \ldots, -i/(x+\mu_{2m+n}y)\rangle\rangle,\#\rangle\mathbf{Z}^{-1}\boldsymbol{\chi}_0, \tag{6.6a}$$

$$2\pi\boldsymbol{\chi}_{,y} = \mathbf{Z}\langle\langle -i\mu_1/(x+\mu_1 y), \ldots, -i\mu_{2m+n}/(x+\mu_{2m+n}y)\rangle\rangle,\#\rangle\mathbf{Z}^{-1}\boldsymbol{\chi}_0. \tag{6.6b}$$

On the positive x-axis, Eq. (6.6a) reduces to the following simple expression by virtue of Eq. (5.9):

$$\begin{aligned}\boldsymbol{\chi}_{,x} &= \{F^{(1)}_{,xy}, -F^{(1)}_{,xx}, \ldots, F^{(m)}_{,xy}, -F^{(m)}_{,xx}, \psi^{(1)}_{,x}, \ldots, \psi^{(n)}_{,x}, U^{(1)}_{,x}, V^{(1)}_{,x}, \ldots, U^{(m)}_{,x}, V^{(m)}_{,x}, W^{(1)}_{,x}, \ldots, W^{(m)}_{,x}\}^T \\ &= (2\pi r)^{-1}\boldsymbol{\Gamma}\mathbf{II}\boldsymbol{\chi}_0 = (2\pi r)^{-1}\boldsymbol{\Gamma}\mathbf{II}\oint d\boldsymbol{\chi}.\end{aligned} \tag{6.7}$$

Thus, the intrinsic matrix $\boldsymbol{\Gamma}$ has the physical significance as the influence matrix relating the various types of discontinuities at the singularity (concentrated forces, discontinuities of displacement and inclination, etc.) to the response along the positive x-axis of the physical variables $F^{(i)}_{,xy}$, $-F^{(i)}_{,xx}$, $\psi^{(j)}_{,x}$, $U^{(i)}_{,x}$, $V^{(i)}_{,x}$ and $W^{(j)}_{,x}$ (multiplied by $2\pi r$).

If Eq. (6.7) and the corresponding expression of $\boldsymbol{\chi}_{,y}$ are recast in polar coordinates, then the variables corresponding to Eq. (6.7) in polar coordinates are the components of the vector $(2\pi r)^{-1}\times\mathbf{Q}_{4m+2n}\boldsymbol{\Gamma}\mathbf{II}\boldsymbol{\chi}_0$, which is related to the right-hand side of Eq. (6.7) by a rotation transformation only (see Eq. (5.2) of [21] for plane anisotropic elasticity and Eq. (6.16) of [16] for anisotropic plate theory). This shows that, for Green's functions of the infinite domain in anisotropic elasticity and laminates, the membrane forces N_θ, $N_{r\theta}$, the circumferential bending moment M_θ, and the radial components of strain and curvature, ε_r and $w_{,rr}$ all have the simple $\cos\theta$ and $\sin\theta$ dependence, and their values on any radial line are related to the corresponding values on the positive x-axis by the rotation transformation. However, other variables including N_r, M_r, $M_{r\theta}$, ε_θ, $2\varepsilon_{r\theta}$ and the curvature components with the θ and r-θ subscripts have much more complicated angular dependence involving trigonometric functions with complex arguments.

7 Green's Function of the Semi-Infinite Region With Interior or Edge Singularities

Consider the problem of the half plane $y \geqslant 0$, with a singularity of strength $\boldsymbol{\chi}_0$ at an interior point $(x,y) = (0,h)$, and with $2m+n$ homogeneous boundary conditions on $y = 0$:

$$\mathbf{K}\boldsymbol{\chi} = \mathbf{0}, \tag{7.1a}$$

$$\mathbf{K} \equiv [\mathbf{K}_1, \mathbf{I}_{2m+n} - \mathbf{K}_1], \tag{7.1b}$$

where $\mathbf{K}_1$ is a *diagonal* matrix of dimension $2m+n$ whose diagonal elements are either 0 or 1. Examples of $\mathbf{K}_1$ for various fixed, free and mixed boundary conditions of anisotropic plates with bending-stretching coupling are shown in [17].

By shifting the singularity from the origin to the point $(0,h)$, Green's function of the infinite domain becomes $(2\pi)^{-1}\mathbf{G}_{(0,h)}$, where

$$\mathbf{G}_{(0,h)}=\mathbf{Z}\langle\langle -\log[x+\mu_1 y-\mu_1 h],\ldots,-i\log[x+\mu_{2m+n}(y-h)\rangle\rangle,\#\rangle\mathbf{Z}^{-1}. \tag{7.2}$$

Although Eq. (7.2) has the required singularity at $(0,h)$, $\mathbf{KG}_{(0,h)}$ does not vanish on $y=0$. One must combine (7.2) with an appropriate nonsingular solution so that the combination $\mathbf{G}$ satisfies $\mathbf{KG}|_{y=0}=0$. Clearly,

$$\begin{aligned}\mathbf{K}(\mathbf{G}|_{y=0})\mathbf{Z}=&-\sum\ i\log[x-\mu_j h]\mathbf{KZ}\langle\boldsymbol{\Delta}_j,\mathbf{0}_{(2m+n)\times(2m+n)}\rangle\\&+\sum\ i\log[x-\overline{\mu}_j h]\mathbf{KZ}\langle\mathbf{0}_{(2m+n)\times(2m+n)},\boldsymbol{\Delta}_j\rangle=\\&-\sum\ i\log[x-\mu_j h][\mathbf{KZ}_\perp\boldsymbol{\Delta}_j,\mathbf{0}_{(2m+n)\times(2m+n)}]\\&+\sum\ i\log[x-\overline{\mu}_j h][\mathbf{0}_{(2m+n)\times(2m+n)},\mathbf{K}\overline{\mathbf{Z}}_\perp\boldsymbol{\Delta}_j],\end{aligned} \tag{7.3}$$

where the summations extend over $1\leqslant j\leqslant 2m+n$, and $\boldsymbol{\Delta}_j$ denotes the diagonal matrix of dimension $2m+n$ with all elements vanish except the jth diagonal element 1.

Consider the following nonsingular matrix functions $\mathbf{G}_j(j=1,\ldots,2m+n)$ and their boundary values:

$$\mathbf{G}_j\equiv\mathbf{Z}\langle\langle -i\log[x+\mu_1 y-\overline{\mu}_j h],\ldots,-i\log[x+\mu_{2m+n}y-\overline{\mu}_j h]\rangle\rangle,\#\rangle\langle\mathbf{C}_j,\overline{\mathbf{C}}_j\rangle\mathbf{II}\mathbf{Z}^{-1}, \tag{7.4}$$

$$\begin{aligned}\mathbf{K}(\mathbf{G}_j|_{y=0})\mathbf{Z}=\mathbf{KZ}\langle -i\log[x-\overline{\mu}_j h]\mathbf{I}_{2m+n},\#\rangle\langle\mathbf{C}_j,\overline{\mathbf{C}}_j\rangle\mathbf{II}=i\log[x-\mu_j h][\mathbf{K}\overline{\mathbf{Z}}_\perp\overline{\mathbf{C}}_j,\mathbf{0}_{(2m+n)\times(2m+n)}]-i\log[x-\overline{\mu}_j h]\\\times[\mathbf{0}_{(2m+n)\times(2m+n)},\mathbf{KZ}_\perp\mathbf{C}_j],\end{aligned} \tag{7.5}$$

where $\mathbf{C}_j\,(j=1,\ldots,2m+n)$ are constant matrices of dimension $(2m+n)\times(2m+n)$ to be determined from the boundary condition

$$\mathbf{K}\left(\mathbf{G}_{(0,h)}+\sum\ \mathbf{G}_j\right)|_{y=0}=\mathbf{0}. \tag{7.6}$$

Substituting Eqs. (7.2) and (7.4) into the last equation, one obtains

$$\mathbf{C}_j=\mathbf{T}\boldsymbol{\Delta}_j\ \ (j=1,\ldots,2m+n), \tag{7.7a}$$

$$\mathbf{T}\equiv(\mathbf{KZ}_\perp)^{-1}\mathbf{K}\overline{\mathbf{Z}}_\perp, \tag{7.7b}$$

and *Green's function for the half plane* $y\geqslant 0$ *with a singularity at* $(0,h)$ *is given by* $(2\pi)^{-1}\mathbf{G}$, *where*

$$\begin{aligned}\mathbf{G}=&\mathbf{Z}\langle\langle -i\log[x+\mu_1 y-\mu_1 h],\ldots,-i\log[x+\mu_{2m+n}(y-h)]\rangle\rangle,\\&\#\rangle\mathbf{Z}^{-1}+\sum\nolimits_{1\leqslant j\leqslant 2m+n}\mathbf{Z}\langle\langle -i\log[x+\mu_1 y-\overline{\mu}_j h],\ldots,\\&-i\log[x+\mu_{2m+n}y-\overline{\mu}_j h]\rangle\rangle,\#\rangle\langle\mathbf{T}\boldsymbol{\Delta}_j,\overline{\mathbf{T}}\boldsymbol{\Delta}_j\rangle\mathbf{II}\mathbf{Z}^{-1}.\end{aligned} \tag{7.8}$$

This solution is valid only for the nondegenerate case, including the case when all eigenvalues are distinct. Setting $h=0$ in Eq. (7.8), one obtains Green's function for the half plane subjected to an edge singularity at the origin.

We recall that the analytic functions $f_k(x+\mu_k y,\mu_k)$ in the general solution of Eq. (4.15) may depend on μ_k implicitly through the complex variable $x+\mu_k y$ and explicitly through the second argument. An example of the general dependence is given by the first line of Eq. (7.8).

8 Particular Cases and Concluding Remarks

A unified formalism is developed for general types of coupled two-dimensional problems in linear theories of continua with a positive-definite energy density function. The formalism encompasses anisotropic thin plate theory with arbitrary bending-stretching coupling, two-dimensional anisotropic elasticity and piezoelectricity, anisotropic piezoelectric plates, and other theories. In these two-dimensional theories of continua, the physical variables appear in groups of 3-S vectors and 2-S vectors, and their conjugate 3-G and 2-G vectors. By expressing the 3-G and 2-G vectors as functions of the conjugate 3-S and 2-S vectors (arranged in the same order) through the constitutive matrix $[\tilde{\omega}]$, one obtains an eigenrelation of Eq. (2.19). This relation leads immediately to Eq. (2.21), $\mathbf{M}(\mu)\boldsymbol{\eta}=\mathbf{0}$, and Eqs. (3.2), (3.5) and (3.9) for the eigenvectors of the materials. If all material eigenvalues are distinct, as is assumed in the present paper, then a complete set of eigenvectors may be obtained in this simple manner, and assembled as the column vectors of the base matrix $\mathbf{Z}$. The general solution of the theory is then given by Eq. (4.15) in terms of a set of arbitrary analytic functions, each containing a distinct eigenvalue μ_k. The physical variables of stress, strain, plate curvatures and bending and twisting moments, electric field, electric displacement, etc., are obtained from the x- and y-derivatives of the general solution $\boldsymbol{\chi}$ for the potentials. In principle, all solutions of 2-D boundary-value problems for any domain may be obtained by determining the appropriate set of analytic functions $f_k(x+\mu_k y,\mu_k)$ using various mathematical methods including power and Laurent series, analytic continuation, the mapping method in complex planes, boundary integral equations and boundary element methods, etc. A number of these powerful analytical tools have been meticulously developed in isotropic elasticity and systematically applied to a wide range of problems by Muskhelishivili and others [22]. They have been extended to anisotropic media with the use of multiple complex variables. The adoption of such methods to the multi-variable coupled continua in the present formalism is straightforward, though the relatively large size of the problem would certainly require all analytical derivations be performed by symbolic algebra.

When applied to the simple case of two-dimensional anisotropic elasticity, the unified formalism reduces to Lekhnitskii's formalism, *not* to the Eshelby-Stroh formalism in terms of the stiffness constants. The constitutive matrix $[\tilde{\omega}]$ reduces exactly to the 5×5 anisotropic compliance matrix $[\beta_{ij}]$ in the Lekhnitskii theory. The in-plane stresses and the anti-plane shearing stresses form 3-S and 2-S vectors, respectively, and their conjugate 3-G and 2-G vectors are $\{u_{,x},v_{,y},u_{,y}+v_{,x}\}^T$ and $\{w_{,x},w_{,y}\}^T$. The elements of the 2×2 eigenmatrix $\mathbf{M}(\mu)$ are the three well-known polynomials $l_4(\mu)$, $-l_3(\mu)$ and $l_2(\mu)$. The complete derivation of the Eshelby-Stroh formalism may also be given based on the eigenrelation (2.19), as shown by Yin [5]. However, instead of using Eq. (2.21) with a 2×2 matrix $\mathbf{M}(\mu)$, the Eshelby-Stroh formalism uses Eq. (2.20) involving the stiffness matrix. The dimension of the eigenmatrix $\boldsymbol{\Theta}(\mu)^T[\tilde{\omega}]^{-1}\boldsymbol{\Theta}(\mu)$ is 3×3 rather than 2×2. As a result, analytical expression of the eigenvectors become too complicated to be given explicitly. The complexity of the stiffness-based formulation relative to the compliance-based formulation was pointed out by Stroh [2]. Moreover, the analytical difficulty of the Stroh formalism is drastically aggravated in the degenerate cases requiring the determination of higher-order eigenvectors [19].

For the anisotropic plate theory with bending-extension coupling, one has two 3-S vectors, $\{F_{,yy},F_{,xx},-F_{,xy}\}^T$ and $\{w_{,yy},w_{,xx},-w_{,xy}\}^T$, and the conjugate 3-G vectors, $\{u_{,x},v_{,y},u_{,y}+v_{,x}\}^T$ and $\{M_y,M_x,-2M_{xy}\}^T=\{\Psi_{1,x},\Psi_{2,y},\Psi_{1,y}+\Psi_{2,x}\}^T$, where $\Psi_1(x,y)$ and $\Psi_2(x,y)$ are the moment potentials. It is a mixed formulation, in contrast to the purely displacement formulation of the conventional laminated plate theory in terms of the three stiffness matrices **A**, **B** and **D**. Notice that, in

the present formulation, the orders of the moment variables and of the curvature variables are different from those in the conventional laminated pate theory, and M_{xy} and $w_{,xy}$ have been replaced by $-2M_{xy}$ and $-w_{,xy}$, respectively. The resulting eigenrelation also involves a 2×2 eigenmatrix $\mathbf{M}(\mu)$ but all its elements are quartic functions. The general solutions of *all types of nondegenerate and degenerate* anisotropic plates have been determined explicitly [14,15], and so are Green's functions for the infinite plate, a semi-infinite plate with various types of boundary conditions including but not limited to fixed, free and simple supports, and an infinite plate with an elliptical hole or inclusion [14,16,17]. These explicit results for the degenerate cases would be very difficult to obtain, and could be obtained only in unduly cumbersome expressions, if the conventional displacement formulation is used. The displacement formulation for unsymmetric laminates [9–11] yields an eigenmatrix of dimension 3×3.

In two-dimensional piezoelectricity, the electric displacement and the electric field are 2-S and 2-G vectors that participate in the constitutive relation in addition to the stresses and strains. This results in an eigenmatrix $\mathbf{M}(\mu)$ with the dimension 3×3, containing elements that are polynomial functions in μ of degrees varying from 2 to 4. There are 14 distinct types of nondegenerate and degenerate piezoelectric materials, each with a different representation of the general solution. The results are given explicitly by Yin [7,8].

The unified formalism of the present theory determines the eigenvectors and the general solution of two-dimensional problems by eliminating the first *m* odd-numbered components of the eigenvector $\boldsymbol{\xi}$ on the basis of the relation $F^{(i)}_{,xy}=F^{(i)}_{,yx}$. This elimination cannot be implemented in the Stroh formalism of plane elasticity, nor can it be done in the conventional displacement formulation of the laminated plate theory based on the stiffness matrices **A**, **B**, and **D**. Thus, the present theory yields a systematic procedure for choosing the primary and secondary variables and the constitutive matrix that lead to the eigenvectors and the general solution in the simplest manner. This choice of variables yields results and expressions that have the same standard form for the different theories of two-dimensional continua, so that any solution of one set of field equations (and the accompanying boundary conditions) generates a corresponding solution of another set of field equations with a similar mathematical structure. Without the unified formalism, the exact formal analogy among the various theories remains obscure, due to the improper choice of the variables and the constitutive matrices (e.g., stiffness-based Stroh formalism in plane elasticity and **A**, **B** and **D** matrices in the conventional laminated plate theory).

Notice that in Green's functions of the infinite plane and the half plane, the analytic functions $f(x+\mu_k y,\mu_k)$ occur as the elements of a diagonal matrix premultiplied by $\mathbf{Z}$ and postmultiplied by either $\mathbf{Z}^{-1}$ or $\langle \mathbf{T}\boldsymbol{\Delta}_j, \overline{\mathbf{T}\boldsymbol{\Delta}_j}\rangle \mathbf{I}\mathbf{I}\mathbf{Z}^{-1}$. The functional form of f_k is the same for all *k*, and it is dependent on the shape of the region and the boundary conditions, but not on the constitutive matrix and the number of three-vectors and two-vectors in the theory. Thus, when the function form of f_k has been determined for Green's function of plane anisotropic elasticity, one readily obtains corresponding Green's function for the anisotropic plate theory, for two-dimensional problems of piezoelectricity, and even for piezoelectric laminated plates without the need to repeat the analysis for the different theories. One simply uses the same function form f_k in Eq. (6.4) or (7.8), depending on the domain, while using the base matrix $\mathbf{Z}$ specific to the material, and a matrix $\mathbf{T}$ of Eq. (7.7*b*) which characterizes the material and the boundary conditions. In this way, the unified formalism allows the fundamental singularity solutions and the analytical solutions of boundary value problems to be achieved in a single analysis regardless of the dimension of the problem, the number and physical roles of the variables, as well as the constitutive matrix.

In contrast to the relatively simple results of the present case with all distinct eigenvalues, the various degenerate cases are analytically more complicated. The main complication arises from intrinsic coupling of any high-order eigenmode with all lower-order eigenmodes that share a common eigenvalue. The situation may be compared to the intrinsic coupling of the three fracture modes in interface fracture problems, which affects the solutions and yields qualitatively different results from the uncoupled case. Isotropic elastic materials gives one important example of the degenerate case. Their general solution and Green's functions for the various domains cannot be obtained from the analytical solution of the nondegenerate case by mere substitution of the isotropic elastic constants for the anisotropic ones. Complete analyses of the various degenerate cases and the resulting general solution and Green's function have been developed in recent papers for the separate subjects of plane elasticity, piezoelectricity and anisotropic plates [4,5,7,8,16,17,19]. The unified formalism of the present work may be extended to various degenerate cases according to the derivative rule, which has been established for the various special theories in [7,16,19].

References

[1] Lekhnitskii, S. G., 1963, *Theory of Elasticity of an Anisotropic Body*, Holden-Day, San Francisco.

[2] Stroh, A. N., 1958, "Dislocations and Cracks in Anisotropic Elasticity," Philos. Mag., **3**, pp. 625–646.

[3] Ting, T. C. T., 1996, *Anisotropic Elasticity: Theory and Application*, Oxford University Press, New York.

[4] Yin, W.-L., 2000, "Deconstructing Plane Anisotropic Elasticity, Part I: The Latent Structure of Lekhnitskii's Formalism," Int. J. Solids Struct., **37**, pp. 5257–5276.

[5] Yin, W.-L., 2000, "Deconstructing Plane Anisotropic Elasticity, Part II: Stroh's Formalism Sans Frills," Int. J. Solids Struct., **37**, pp. 5277–5296.

[6] Sosa, H., 1991, "Plane Problems in Piezoelectric Media With Defects," Int. J. Solids Struct., **28**, pp. 491–505.

[7] Yin, W.-L., 2005, "Two-Dimensional Piezoelectricity, Part I: Eigensolutions of Nondegenerate and Degenerate Materials," Int. J. Solids Struct., **42**, pp. 2645–2668.

[8] Yin, W.-L., 2005, "Two-Dimensional Piezoelectricity, Part II: General Solution, Green's Function and Interface Cracks," Int. J. Solids Struct., **42**, pp. 2669–2687.

[9] Becker, W., 1991, "A Complex Potential Method for Plate Problems With Bending Extension Coupling," Arch. Appl. Mech., **61**, pp. 318–326.

[10] Lu, P., and Mahrenholtz, O., 1994, "Extension of the Stroh Formalism to the Analysis of Bending of Anisotropic Elastic Plates," J. Mech. Phys. Solids, **42**, pp. 1725–1741.

[11] Cheng, Z.-Q., and Reddy, J. N., 2002, "Octet Formalism for Kirchhoff Anisotropic Plates," Proc. R. Soc. London, Ser. A, **458**, pp. 1499–1517.

[12] Chen, P., and Shen, Z., 2001, "Extension of Lekhnitskii's Complex Potential Approach to Unsymmetric Composite Laminates," Mech. Res. Commun., **28**, pp. 423–428.

[13] Hwu, C., 2003, "Stroh-Like Formalism for the Coupled Stretching-Bending Analysis of Composite Laminates," Int. J. Solids Struct., **40**, pp. 3681–3705.

[14] Yin, W.-L., 2003, "General Solutions of Laminated Anisotropic Plates," ASME J. Appl. Mech., **70**, pp. 496–504.

[15] Yin, W.-L., 2003, "Structure and Properties of the Solution Space of General Anisotropic Laminates," Int. J. Solids Struct., **40**, pp. 1825–1852.

[16] Yin, W.-L., 2005, "Green's Function of Anisotropic Plates With Unrestricted Coupling and Degeneracy, Part 1: The Infinite Plate," Composite Struct., in press.

[17] Yin, W.-L., 2005, "Green's Function of Anisotropic Plates With Unrestricted Coupling and Degeneracy, Part 2: Other Domains and Special Laminates," Composite Struct., in press.

[18] Ting, T. C. T., 1992, "Anatomy of Green's Functions for Line Forces and Dislocations in Anisotropic Media and Degenerate Materials," *The Jens Lothe Symposium Volume*, Phys. Scr., T, **T44**, pp. 137–144.

[19] Yin, W.-L., 2004, "Degeneracy, Derivative Rule, and Green's Function of Anisotropic Elasticity," ASME J. Appl. Mech., **71**, pp. 273–282.

[20] Yin, W.-L., 2003, "Anisotropic Elasticity and Multi-Material Singularities," J. Elast., **71**, pp. 263–292.

[21] Yin, W.-L., 2005, "Green's Function of Bimaterials Comprising all Cases of Material Degeneracy," Int. J. Solids Struct., **42**, pp. 1–19.

[22] Muskhelishvili, N. I., 1953, *Some Basic Problems of the Mathematical Theory of Elasticity*, Noordhoof, Leyden.

Control of Bending Vibrations Within Subdomains of Thin Plates—Part I: Theory and Exact Solution

Michael Krommer
Institute for Technical Mechanics,
Johannes Kepler University Linz,
Altenbergerstr. 69,
A-4040 Linz, Austria
e-mail: krommer@mechatronik.uni-linz.ac.at

Vasundara V. Varadan
George & Boyce Billingsley Endowed Chair
and Distinguished Professor of Electrical
Engineering,
University of Arkansas,
3217 Bell Engineering Center,
Fayetteville, AR 72701
e-mail: vvvesm@engr.uark.edu

Control of the surface profile or shape of structures that deform under externally applied dynamical loads is important in many applications where no control can be exercised on the applied loads. The only recourse is to make the structure adaptive by the action of smart actuators that can null or nearly null the resulting deformation. The class of problems, to which shape control may be applied, is huge and in this paper a theoretical approach is presented for a special subset of such problems, wherein, suitable actuation can be applied in order to keep a subdomain of the structure in its nondeformed state under the action of external dynamical loads. A suitable actuation to achieve this goal is the complement of the self-stress. An appropriate distribution of the self-stress should result in an elimination of the motion of the subdomain of the structure. Moreover, we seek a solution of the problem, which only requires the application of the self-stress in the subdomains or in a slightly larger domain. This is also a practical approach to such problems where it would be prohibitively expensive to design and power actuators to control the entire domain. We choose a linear, thin elastic plate to present the basics of our methodology. The main part of the paper is devoted to the theoretical foundation of the method; however, to show its validity, we also present exact results for the simple case of a circular plate in axisymmetric bending. [DOI: 10.1115/1.1839185]

1 Introduction

Shape control of structures is concerned with methods that will result in a desired shape by applying a suitable actuation. A desired shape may be a prescribed new shape or may be the nondeformed shape for a structure that is under the influence of external disturbances. No matter what the actual desired shape is, the problem formulation will always result in the question: Given an external disturbance, how does a suitable distribution of the actuation result in the structure assuming the desired shape.

In a fundamental contribution Haftka and Adelman [1] were the first to introduce the notion of shape control into the journal literature. They noted that disturbances that affect the shape of the structure could be subdivided into two types; one type is transient, whereas the other type is associated with fixed deformations or those that vary slowly in time. The first refers to dynamic shape control while the latter refers to static shape control. Haftka and Adelman addressed the problem of static shape control of a large spacecraft structure by applying temperature as an actuating mechanism. Irschik and Pichler [2] reported on results for dynamic shape control of solids and structures by applied thermal expansion strains. Irschik [3] gave a detailed discussion and review of both static and dynamic shape control using piezoelectric eigenstrains as the actuation mechanism. In addition to thermal expansion strains and piezoelectric strains, other actuation methods may be used. However, all of these actuation methods belong to the general class of what is usually referred to as eigenstrain actuation or self-stress actuation methods. Early reports on self-stresses were given by Reissner [4] and Nemenyi [5]. Presently, self-stresses are discussed in connection with micromechanics of solids (Mura [6]) and, more recently, in connection with shape control and active (noise) control of structures (Rao and Sunar [7], Gopinathan et al. [8], Irschik [3]). Hence, a unified approach can be used for all members of the general class of self-stress actuation methods.

In the present Part (Part I) of this paper we present the theoretical basics of a new method to control the bending vibrations of a subdomain of a thin plate; suitably distributed sources of self-stress are only applied in the subdomain itself. We calculate exact solutions in the framework of thin circular plates, which bend into an axisymmetric surface, and we consider control of arbitrary subdomains; however, we do not pay any attention to the problem of how to practically achieve the required distributed control. We only assume the control agency to be any possible type of sources of self-stress. In the forthcoming part (Part II, Krommer and Varadan [9]) of this paper we will especially focus on this latter practical aspect of the problem.

We begin the first part of this paper by discussing a special problem of dynamic shape control. Our goal is to eliminate the motion of a subdomain of a linear elastic plate by applying control by means of self-stress actuation. We apply the actuation only in the subdomain we want to control, or in a slightly larger domain. Dynamic shape control of plates has been studied intensively in the literature; however, elimination of the total motion of the plate was the desired goal. Nader et al. [10] and Irschik et al. [11] published results for circular plates and rectangular plates with homogenous kinematic boundary conditions. In order to be applicable to dynamic shape control of subdomains, we have to extend the method developed by Nader et al. [10] and Irschik et al. [11] to the case of nonhomogenous kinematic boundary conditions in a first step. Our method for the dynamic shape control of the subdomain is based on the free-body diagram of the subdomain; releasing the subdomain from the rest of the plate requires that we account for the influence of the rest of the plate by nonhomog-

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, June 11, 2004; final revision, July 16, 2004. Editor: R. M. McMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California–Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

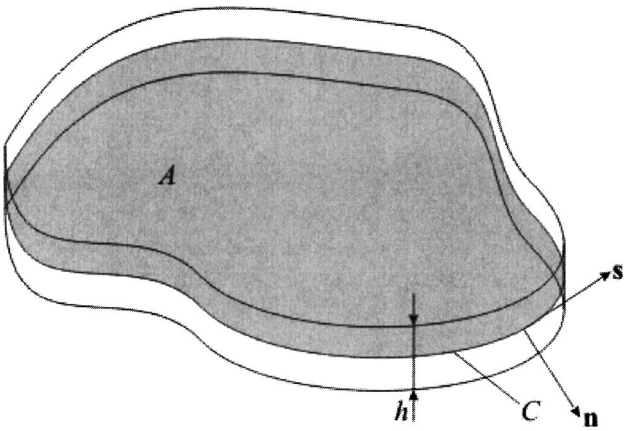


Fig. 1 Geometry of the plate

enous boundary conditions. These have to be chosen such that a solution for the statically admissible stress in the released subdomain exists. Therefore, nonhomogenous kinematic boundary conditions enter the formulation in this step. To eliminate the nonhomogenous kinematic boundary conditions, we introduce an additive decomposition of the motion into two parts; one accounts for the nonhomogenous kinematic boundary conditions, the other for the external force loading. In order to eliminate both parts of the bending motion in the subdomain, we have to release a slightly larger domain of the plate; then we are able to eliminate the influence of the nonhomogenous kinematic boundary conditions within the additional, so-called transition domain, such that the motion within the subdomain is exactly eliminated or controlled to result in the desired shape of the subdomain. Therefore, our solution is optimal. Finally, we demonstrate the validity of the method for the axisymmetric bending motion of a thin, circular plate, for which we are able to calculate exact analytic solutions.

The motivation for starting to work on this problem was the problem of controlling the shape of an antenna that is conformal to a deformable structure. For the performance of such antennas it is important to keep the domain of a structure, where the antenna is situated, in its nondeformed state. Preliminary results for the antenna problem can be found in Krommer and Varadan [12] in which we have not provided the theoretical foundation. Extended results will be presented in the second part of this paper (Krommer and Varadan [9]). Other applications belong to the field of noise reduction in structures; see, for instance, Gopinathan et al. [8]. In these applications it is often not necessary to control the whole structure, but a significant noise reduction can be achieved by focusing on critical parts of the structure. An example would be the funnel of a magnetic resonance imaging (MRI) unit, which is responsible for the highly annoying noise experienced by patients (Nader et al. [13]).

2 Mathematical Formulation

The equations that govern the small bending motion of a thin plate (or possibly only a subdomain of a thin plate) are

$$A:\ \operatorname{div}(\operatorname{div}\mathbf{M})+p_z=\rho^{(0)}\ddot{w}_0,$$

$$C:\ [\operatorname{div}\mathbf{M}\cdot\mathbf{n}+\nabla(\mathbf{M}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s}-\overline{q}]\delta w_0=0,\ (\mathbf{M}\mathbf{n}\cdot\mathbf{n}-\overline{m})(\nabla\delta w_0\cdot\mathbf{n})=0, \tag{1}$$

$$P:\ [(\mathbf{M}\mathbf{n}\cdot\mathbf{s})\delta w_0]_{P^+}^{P^-}=0,$$

together with initial conditions for w_0 and $\dot{w}_0$. A is the area of the plate, which is bounded by the curve C; $\mathbf{n}$ and $\mathbf{s}$ are the unit normal vector and the unit tangential vector of C, and P is any point of C. Figure 1 shows the geometry of the plate. Further, $\mathbf{M}$ is the second-order moment tensor, p_z is the transverse force loading, $\rho^{(0)}$ denotes linear inertia w_0 denotes the deflection of the plate, and $\overline{m}$ and $\overline{q}$ stand for the prescribed moment and the prescribed transverse force at the boundary. Details of the notation used in Eq. (1) can be found in Krommer [14]. The constitutive relation for the moment tensor $\mathbf{M}$ is

$$\mathbf{M}=\mathbf{D}:\boldsymbol{\kappa}-\mathbf{M}^*, \tag{2}$$

where $\boldsymbol{\kappa}=-[\nabla\nabla w_0+(\nabla\nabla w_0)^T]/2$ is the curvature tensor. The second order self-stress tensor $\mathbf{M}^*$ denotes a source of self-stress acting in the elastic background plate. Sources of self-stresses may be of different nature, like thermal stresses or piezoelectric stresses; however, self-stresses are not restricted to these latter two sources; see Mura [6]. $\mathbf{D}$ is the fourth order bending stiffness tensor. Equation (1) can be cast into a convolution integral

$$\begin{aligned}\int_A \hat{p}_z^d\hat{w}_0 dA-\int_{C_\psi}(\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{n})\hat{\bar{\psi}}dC+\int_{C_w}[\operatorname{div}\hat{\mathbf{M}}^d\cdot\mathbf{n}\\ +\nabla(\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s}]\hat{\bar{w}}_0 dC\\ =\int_A\hat{\bar{p}}_z\hat{w}_0^d dA+\int_A\hat{\mathbf{M}}^*:\hat{\boldsymbol{\kappa}}^d dA-\int_{C_m}\hat{\bar{m}}(\nabla\hat{w}_0^d\cdot\mathbf{n})dC\\ +\int_{C_q}\hat{\bar{q}}\hat{w}_0^d dC+\sum_i[\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s}\hat{w}_0]_{P_{wi}^+}^{P_{wi}^-},\end{aligned} \tag{3}$$

see the Appendix for the derivation. Equation (3) is formulated in the Laplace domain, characterized by a hat. The superscript d denotes a state of the plate due to an arbitrary loading p_z^d and with homogenous initial and boundary conditions as well as with $\mathbf{M}^{*d}=\mathbf{0}$. We refer to this problem as the ''dummy problem.'' $\hat{\bar{p}}_z=\hat{p}_z+\rho^{(0)}(sw_0(t=0)+\dot{w}_0(t=0))$ accounts for nonhomogenous initial conditions of the original problem; s is the Laplace variable. The last term represents the work done by the corner forces of the dummy problem with respect to the deflection of the original problem in those corners, where the deflection is prescribed. The boundary has been split into four parts corresponding to different types of boundary conditions. $\overline{w}_0$ and $\overline{\psi}$ are prescribed functions for displacement and normal slope at the boundary in the original problem. Equation (3) represents an extension of the principle of virtual forces and Maysel's formula to dynamic problems involving nonhomogenous initial conditions and nonhomogenous boundary conditions. For reference see Ziegler [15] and Ziegler and Irschik [16]. The problem with nonhomogenous kinematic boundary conditions can be transformed into a problem with homogenous kinematic boundary conditions by splitting the total deflection into two parts $w_0=\tilde{w}_0+\breve{w}_0$, where $\breve{w}_0$ accounts for the nonhomogenous kinematic boundary conditions by means of

$$C_w:\ \breve{w}_0=\overline{w}_0\quad\text{and}\quad C_\psi:\ \nabla\breve{w}_0\cdot\mathbf{n}=\overline{\psi}. \tag{4}$$

In addition $\breve{w}_0$ has to be an admissible plate deflection. An integral representation for the bending motion $\tilde{w}_0$ is

$$\begin{aligned}\int_A\hat{p}_z^d\hat{\tilde{w}}_0 dA=\int_A\hat{\tilde{p}}_z\hat{w}_0^d dA+\int_A\hat{\mathbf{M}}^*:\hat{\boldsymbol{\kappa}}^d dA-\int_{C_m}\hat{\tilde{m}}(\nabla\hat{w}_0^d\cdot\mathbf{n})dC\\ +\int_{C_q}\hat{\tilde{q}}\hat{w}_0^d dC,\end{aligned} \tag{5}$$

see the Appendix for the derivation. In Eq. (5) a transformed force $\tilde{p}_z$, a transformed moment $\tilde{m}$, and a transformed transverse force $\tilde{q}$ have been introduced. Definitions are

$$\begin{aligned}A:\ \hat{\tilde{p}}_z=\hat{p}_z+\rho^{(0)}(sw_0(t=0)+\dot{w}_0(t=0))-\rho^{(0)}s^2\hat{\tilde{w}}_0\\ +\operatorname{div}(\operatorname{div}\hat{\tilde{\mathbf{M}}}),\end{aligned} \tag{6}$$

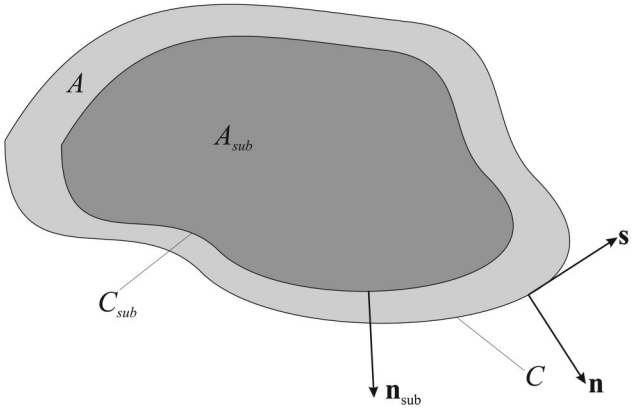


Fig. 2 Plate with subdomain

$$C_m:\ \hat{\tilde{m}}=\hat{\tilde{m}}-\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{n},\quad C_q:\ \hat{\tilde{q}}=\hat{\tilde{q}}-\operatorname{div}\hat{\breve{\mathbf{M}}}\cdot\mathbf{n}-\nabla(\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s}.$$

$\breve{\mathbf{M}}$ is a second-order moment tensor, which is calculated from the deflection $\breve{w}_0$ as

$$\breve{\mathbf{M}}=\mathbf{D}:\breve{\boldsymbol{\kappa}},\quad \breve{\boldsymbol{\kappa}}=-\frac{1}{2}[\nabla\nabla\breve{w}_0+(\nabla\nabla\breve{w}_0)^T]. \tag{7}$$

Finally, we define the problem of a statically admissible moment tensor $\mathbf{M}^p$ as

$$\begin{aligned} &A:\ \operatorname{div}(\operatorname{div}\mathbf{M}^p)+\tilde{p}_z=0,\\ &C_q:\ \operatorname{div}\mathbf{M}^p\cdot\mathbf{n}+\nabla(\mathbf{M}^p\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s}=\tilde{q},\quad C_m:\ \mathbf{M}^p\mathbf{n}\cdot\mathbf{n}=\tilde{m},\\ &\overline{P}_i:\ [\mathbf{M}^p\mathbf{n}\cdot\mathbf{s}]_{P_i^+}^{\overline{P_i^-}}=0 \end{aligned} \tag{8}$$

The last term has to be satisfied for all corner points $\overline{P}_i$, where the deflection is not prescribed. The integral representation of Eq. (5) can be reformulated by taking care of Eq. (8). The relations for $\tilde{p}_z$, $\tilde{q}$, and $\tilde{m}$, as stated in Eq. (8), are inserted into Eq. (5) and the Gauss theorem is applied twice. The result is

$$\int_A \hat{p}_z^d\hat{\tilde{w}}dA=\int_A[\hat{\mathbf{M}}^p+\hat{\mathbf{M}}^*]:\hat{\boldsymbol{\kappa}}^d dA. \tag{9}$$

Due to the arbitrariness of p_z^d the deflection $\tilde{w}_0=w_0-\breve{w}_0$ vanishes identically, if $\hat{\mathbf{M}}^p+\hat{\mathbf{M}}^*=\mathbf{0}$ is satisfied. Therefore, applying control by means of self-stress actuation can be used to eliminate the bending motion $\tilde{w}_0$ of the plate that is induced by transverse forces. The motion $\breve{w}_0$, which we like to account to the nonhomogenous kinematic boundary conditions, is not eliminated. The required distribution of the self-stress actuation is calculated from Eq. (8). This is a simple task, because without any loss of generality the tensor $\mathbf{M}^p$ can be taken as spherical, $\mathbf{M}^p=m^p\mathbf{I}$. Hence, $\mathbf{M}^p\mathbf{n}\cdot\mathbf{s}=m^p\mathbf{I}\mathbf{n}\cdot\mathbf{s}=0$ and Eq. (8) changes to

$$\begin{aligned} &A:\ \Delta m^p+\tilde{p}_z=0,\\ &C_q:\ \nabla m^p\cdot\mathbf{n}=\tilde{q},\quad C_m:\ m^p=\tilde{m}. \end{aligned} \tag{10}$$

Depending on the boundary conditions, the solution to Eq. (10) is either unique, not unique, or it does not exist. The nonuniqueness is not a problem, but the nonexistence is.

A crucial point for realizing control of a subdomain of the plate is the choice of $\breve{w}_0$, which, besides being an admissible plate deflection, only has to satisfy the nonhomogenous kinematic boundary conditions. The deflection of the controlled plate will exactly coincide with this latter deflection $\breve{w}_0$. We therefore consider the following procedure. We chose a subdomain A_{sub} of the plate, bounded by C_{sub}; see Fig. 2. Within A_{sub} we take the deflection $\breve{w}_0$ to be zero. $\breve{w}_0$ at least has to satisfy

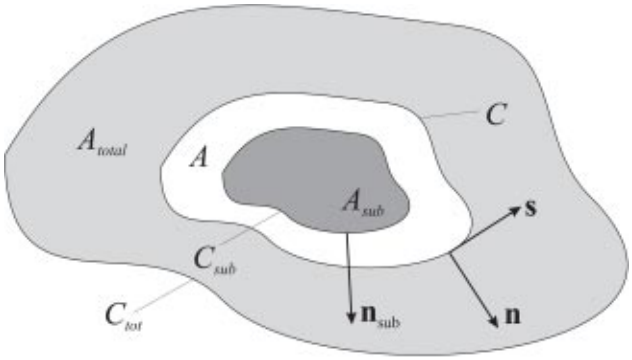


Fig. 3 Plate with subdomain to be controlled

$$\begin{aligned} &A_{\text{sub}}:\ \breve{w}_0=0,\\ &C_w:\ \breve{w}_0=\overline{w}_0,\quad C_\psi:\ \nabla\breve{w}_0\cdot\mathbf{n}=\overline{\psi},\\ &C_{\text{sub}}:\ \breve{w}_0=0\quad\text{and}\quad \nabla\breve{w}_0\cdot\mathbf{n}_{\text{sub}}=0. \end{aligned} \tag{11}$$

Any deflection satisfying the relations of Eq. (11) is a proper choice for $\breve{w}_0$, resulting in $\breve{w}_0=0$ inside A_{sub}; however, it has to be an admissible plate deflection. The required distribution of the control by means of self-stress actuation then is obtained from Eq. (10), where the transformed force $\tilde{p}_z$, the transformed moment $\tilde{m}$, and the transformed transverse force $\tilde{q}$ are

$$\begin{aligned} &A_{\text{sub}}:\ \hat{\tilde{p}}_z=\hat{p}_z+\rho^{(0)}(sw_0(t=0)+\dot{w}_0(t=0)),\\ &A-A_{\text{sub}}:\ \hat{\tilde{p}}_z=\hat{p}_z+\rho^{(0)}(sw_0(t=0)+\dot{w}_0(t=0))-\rho^{(0)}s^2\hat{\tilde{w}}_0\\ &\qquad+\operatorname{div}(\operatorname{div}\hat{\breve{\mathbf{M}}}),\\ &C_m:\ \hat{\tilde{m}}=\hat{\tilde{m}}-\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{n},\quad C_q:\ \hat{\tilde{q}}=\hat{\tilde{q}}-\operatorname{div}\hat{\breve{\mathbf{M}}}\cdot\mathbf{n}-\nabla(\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s}. \end{aligned} \tag{12}$$

Then, the bending motion of the controlled plate is $\breve{w}_0$. If the kinematic boundary conditions are homogenous, the problem simplifies dramatically; Nader et al. [10] and Irschik et al. [11] published results for circular plates and rectangular plates with homogenous kinematic boundary conditions.

3 Subdomain Control

We study a plate within the domain A_{total} and with a boundary defined by C_{total}; the kinematic boundary conditions are homogenous. Our scope is to control the bending motion within a subdomain A of A_{total}; moreover, we apply control only in the subdomain A. We release the subdomain from the rest of the plate and we account for the effect of the rest of the plate, that is $A_{\text{total}}-A$, by applying, yet unknown, boundary conditions at the interfacing curve C. Then we apply the method we have just developed in the preceding section. But, the boundary conditions have to ensure a solution of Eq. (10). Three possible types of boundary conditions that are proper are

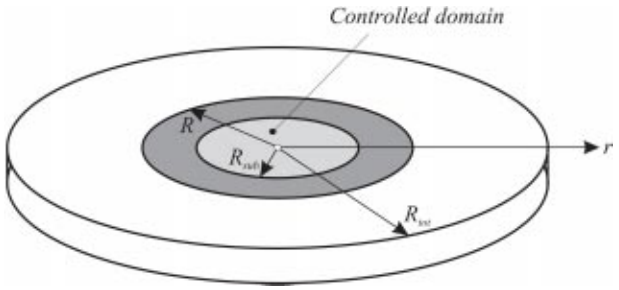


Fig. 4 Clamped circular plate

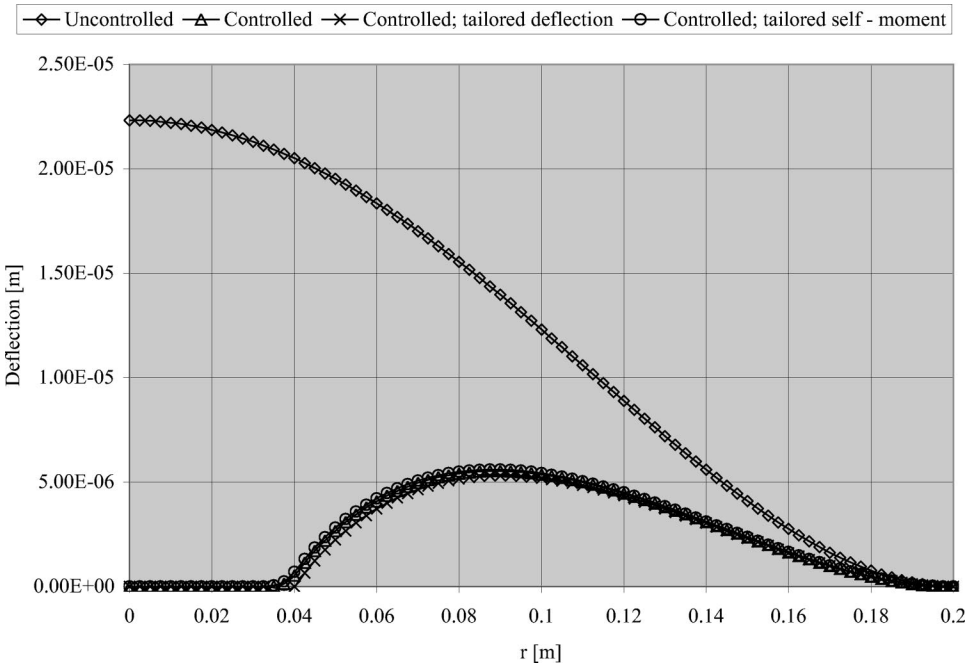


Fig. 5 Deflection of the clamped plate; $\omega=2\pi 100\ \mathrm{s}^{-1}$

$$C:\ \begin{cases} w_0=\overline{w}_0, & \mathbf{Mn}\cdot\mathbf{n}=\overline{m} \\ \operatorname{div}\mathbf{M}\cdot\mathbf{n}+\nabla(\mathbf{Mn}\cdot\mathbf{s})\cdot\mathbf{s}=\overline{q}, & \nabla w_0\cdot\mathbf{n}=\overline{\psi} \\ w_0=\overline{w}_0, & \nabla w_0\cdot\mathbf{n}=\overline{\psi} \end{cases} \tag{13}$$

As one can see, each of the three types of boundary conditions has at least one nonhomogenous kinematic boundary condition that has to be satisfied. Therefore, we have to apply the methodology that we have developed in the preceding section. We introduce the additive separation of the deflection $w_0=\tilde{w}_0+\breve{w}_0$ within A. Then the statically admissible moment tensor $\mathbf{M}^p=m^p\mathbf{I}$ has to satisfy

$$\begin{aligned} &A:\ \Delta m^p+\tilde{p}_z=0, \\ &C:\ \begin{cases} m^p=\tilde{m} \\ \nabla m^p\cdot\mathbf{n}=\tilde{q} \\ \text{no boundary conditions to be satisfied,} \end{cases} \end{aligned} \tag{14}$$

where $\tilde{p}_z$ and $\tilde{m}$, $\tilde{q}$ have already been defined in Eq. (6). If $\mathbf{M}^p+\mathbf{M}^*=\mathbf{0}$ is satisfied within A, the deflection $\tilde{w}_0=w_0-\breve{w}_0$ vanishes. Again the crucial point is the choice of $\breve{w}_0$. For that sake we

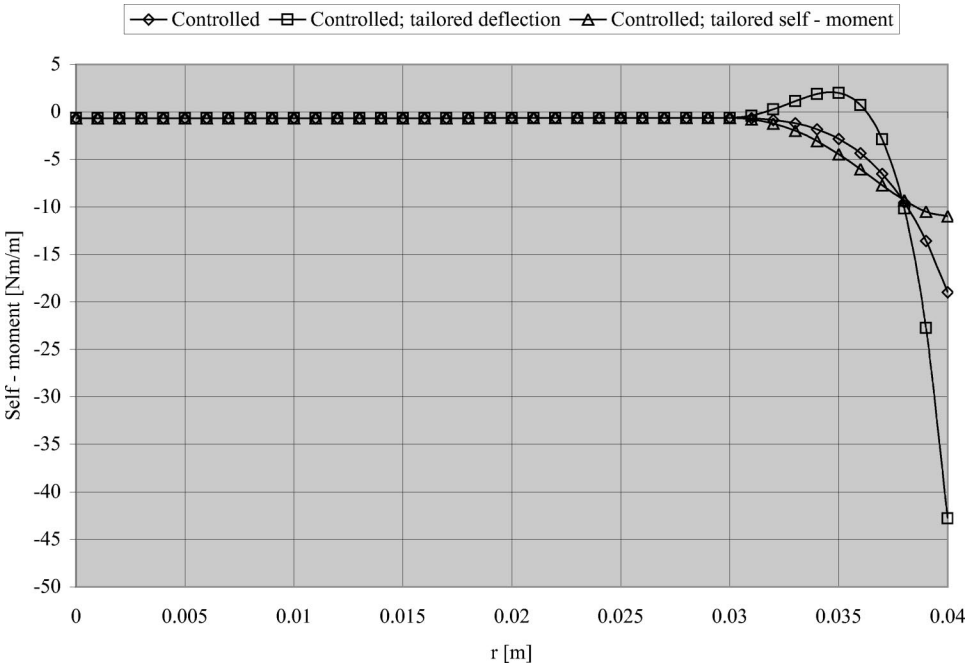
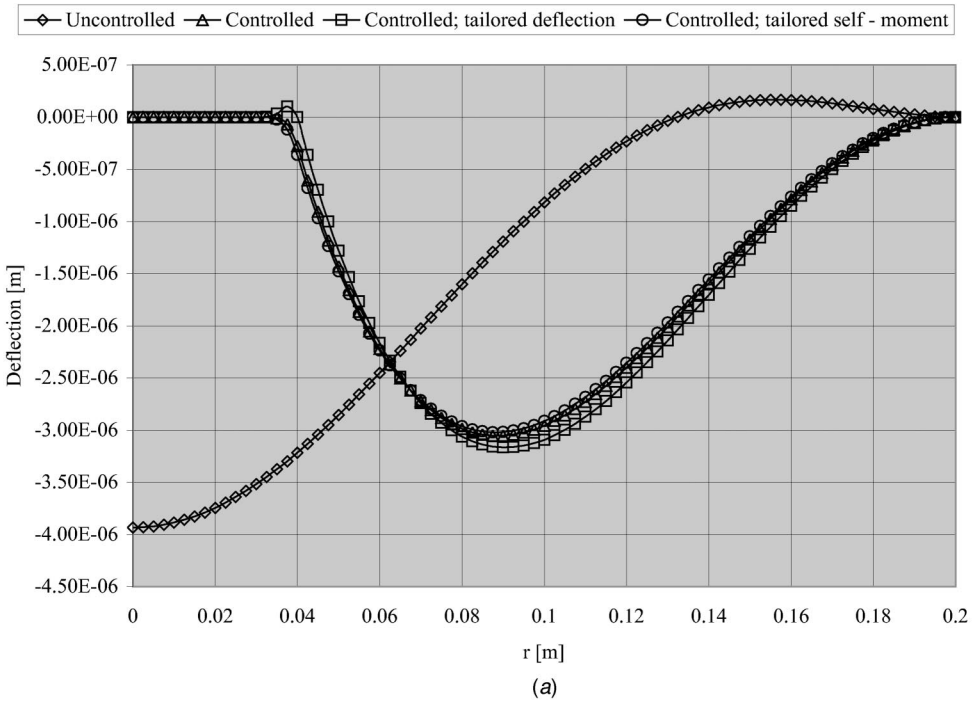


Fig. 6 Self-moment of the clamped plate; $\omega=2\pi 100\ \mathrm{s}^{-1}$



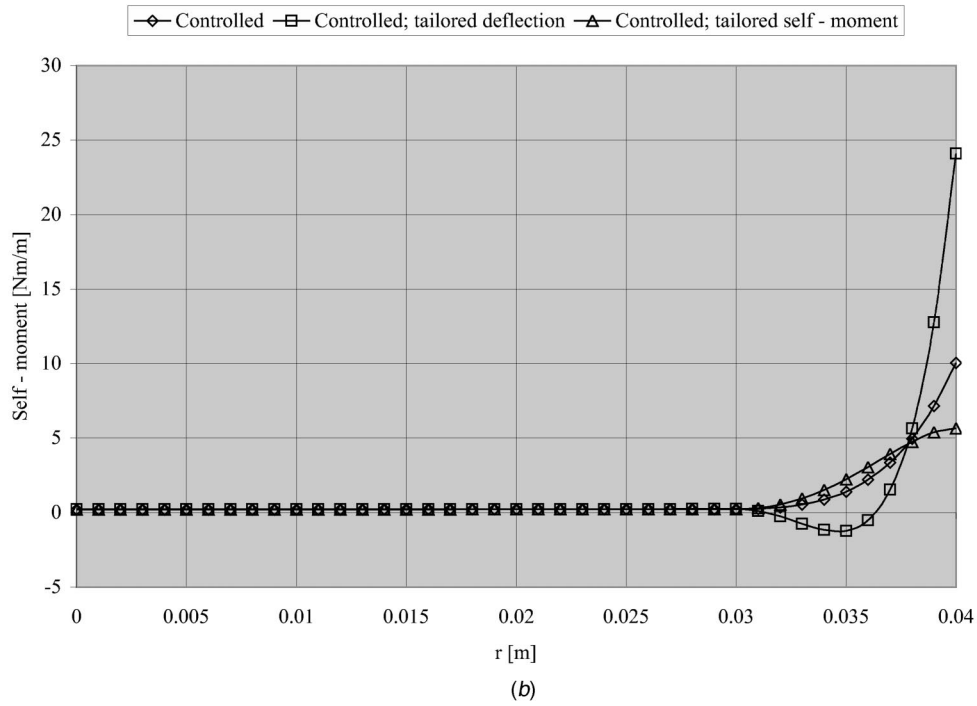


Fig. 7 (*a*) **Deflection and** (*b*) **self-moment of the clamped plate;** $\omega = 2\pi 500\ \mathrm{s}^{-1}$

use a subdomain A_{sub} of the subdomain A; within A_{sub} we consider $\breve{w}_0 = 0$. Figure 3 shows a sketch of the plate. Therefore, $\breve{w}_0$ has to satisfy

$$A_{\mathrm{sub}}: \ \breve{w}_0 = 0,$$

$$C: \begin{cases} \breve{w}_0 = \overline{w}_0 \\ \nabla \breve{w}_0 \cdot \mathbf{n} = \overline{\psi} \\ \breve{w}_0 = \overline{w}_0, \ \nabla \breve{w}_0 \cdot \mathbf{n} = \overline{\psi} \end{cases} \tag{15}$$

$$C_{\mathrm{sub}}: \ \breve{w}_0 = 0, \ \nabla \breve{w}_0 \cdot \mathbf{n}_{\mathrm{sub}} = 0.$$

However, note that $\overline{w}_0$, $\overline{\psi}$, $\overline{m}$, and $\overline{q}$ are not known, but account for the influence of the remaining (uncontrolled) portion of the plate. To calculate these latter unknowns, the whole dynamic problem has to be solved. In the next section we present solutions for the case of axisymmetric bending of clamped circular plates; a case for which we are able to calculate exact analytical solutions.

4 Illustrative Example

4.1 Clamped Circular Plate in Axisymmetric Bending. We study a clamped circular plate with radius R_{tot}; the initial conditions are homogenous and the material parameters are iso-

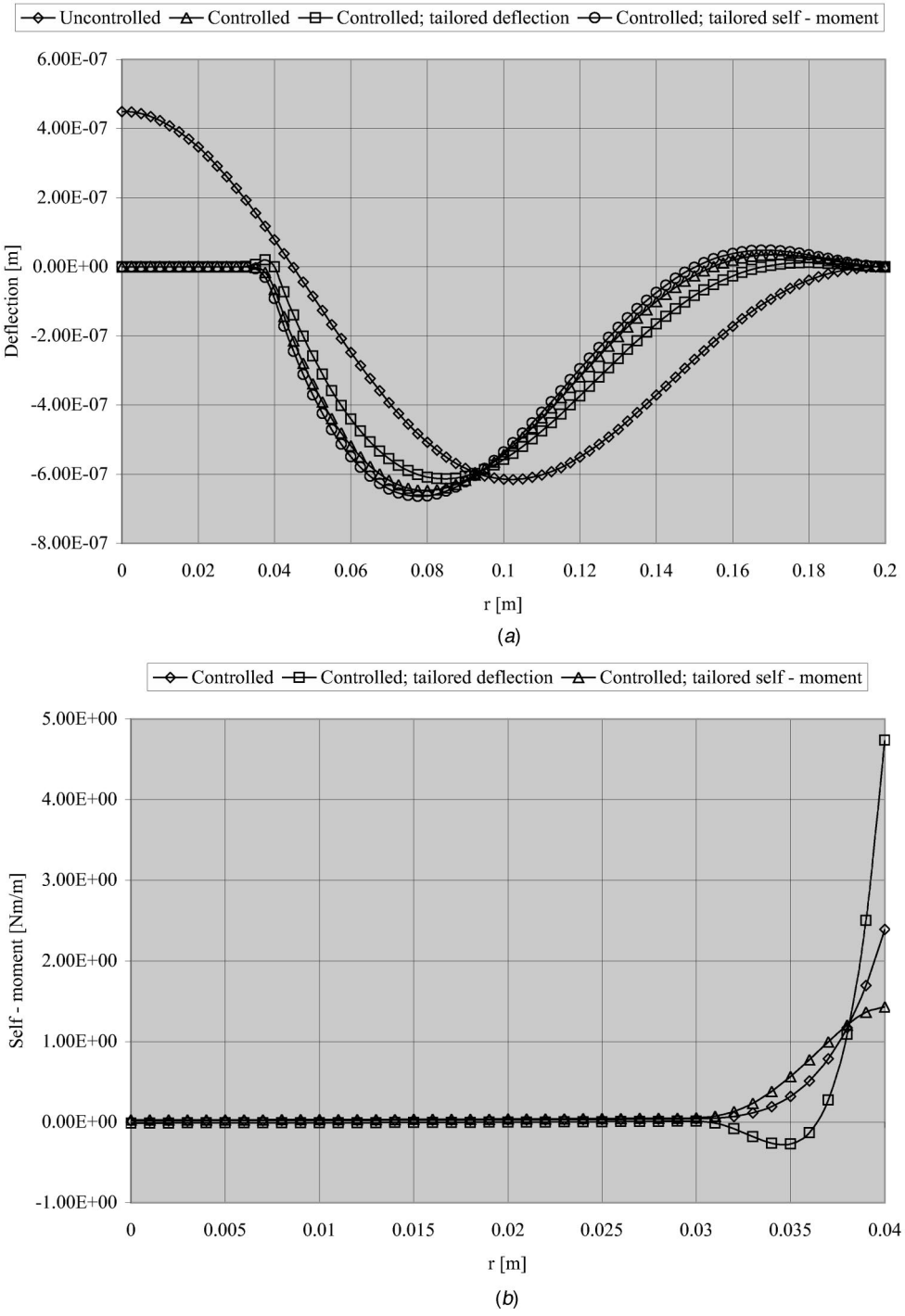


Fig. 8 (*a*) Deflection and (*b*) self-moment of the clamped plate; $\omega = 2\pi 1000\ \mathrm{s}^{-1}$

tropic. A space-wise constant force $p_z(r,\varphi,t) = p_0(t)$ is applied. Indeed, this is not the general case; we should apply a single ring load at an arbitrary location, for instance $p_z(r,\varphi,t) = \delta(r - \xi)p_0(t)$, and calculate the Green's function. However, in this paper we are interested in presenting a theoretical foundation of the method, and to show the validity in a simple example problem. The domain to be controlled has radius R_{sub}. We apply control only within $r \leqslant R$, with $R_{\mathrm{sub}} \leqslant R \leqslant R_{\mathrm{tot}}$. Figure 4 shows the circular plate. At $r = R$ we consider $w_0 = \overline{w}_0$ and $\partial w_0 / \partial r = \overline{\psi}$. $\breve{w}_0$ is calculated from

$$
\begin{gathered}
r \leqslant R_{\mathrm{sub}}: \quad \breve{w}_0 = 0, \\
R_{\mathrm{sub}} \leqslant r \leqslant R: \quad \breve{w}_0 = C_1 + C_2 r^2 + C_3 r^4 + C_4 r^6 + C_5 r^8 + C_6 r^{10}, \quad (16) \\
r = R_{\mathrm{sub}}: \quad \breve{w}_0 = 0, \quad \frac{\partial \breve{w}_0}{\partial r} = 0, \quad r = R: \quad \breve{w}_0 = \overline{w}_0, \quad \frac{\partial \breve{w}_0}{\partial r} = \overline{\psi}.
\end{gathered}
$$

After adjusting the series of Eq. (16) to the boundary conditions, two unknowns remain to be calculated. We shall use these to ensure the deflection be admissible. The next step is to calculate the statically admissible moment tensor for $r \leqslant R$ from Eq. (14).

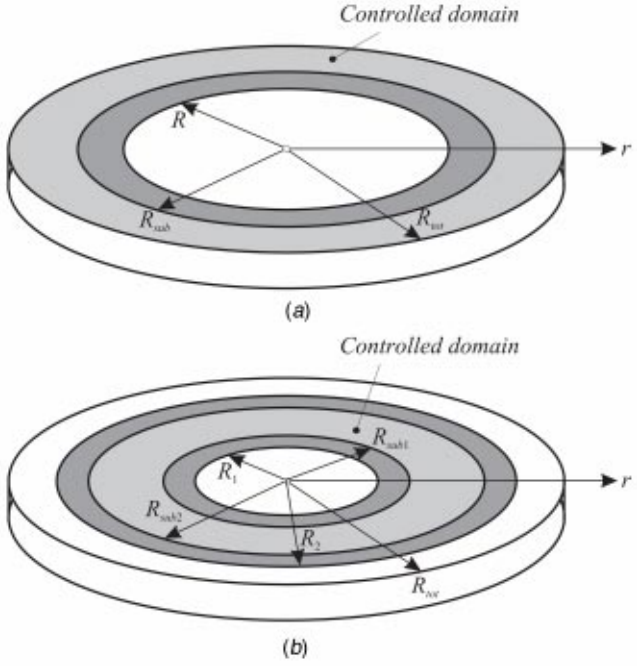


Fig. 9 Alternative domains to be controlled; (*a*) $R \leqslant R_{sub} \leqslant R_{tot}$, (*b*) $R_1 \leqslant R_{sub1} \leqslant R_{sub2} \leqslant R_2$

$$r \leqslant R: \ \frac{1}{r}\frac{\partial m^p}{\partial r} + \frac{\partial^2 m^p}{\partial r^2} + \tilde{p}_0 = 0 \tag{17}$$

No boundary condition has to be satisfied by the solution of Eq. (17); therefore, the solution is not unique. One constant shall remain in the solution for m^p. The transformed force loading is obtained from the definitions of Eq. (12).

$$\begin{aligned} r \leqslant R_{\text{sub}}: \ & \tilde{p}_0 = p_0, \\ R_{\text{sub}} \leqslant r \leqslant R: \ & \tilde{p}_0 = p_0 - \rho^{(0)}\ddot{w}_0 - D\Delta\Delta\breve{w}_0. \end{aligned} \tag{18}$$

The solution to Eq. (17) can be calculated in the two domains separately; the continuity of the solution and its directional derivative are required to be satisfied. If the self-stress moment tensor $\mathbf{M}^*$ is taken as

$$\mathbf{M}^* = -\mathbf{M}^p = -m^p\mathbf{I}, \tag{19}$$

then the motion for $r \leqslant R_{\text{sub}}$ vanishes, $\breve{w}_0 = 0$; for $R_{\text{sub}} \leqslant r \leqslant R$ the motion is $\breve{w}_0$, as it has been specified in the series of Eq. (16). However, the solution for $m^* = -m^p$ contains five unknowns; $\overline{w}_0$ and $\overline{\psi}$, two constants from the power series of Eq. (16), and one constant, because the solution of Eq. (17) is not unique. We eliminate the two constants in the power series by ensuring that the deflection we want the controlled plate to perform is an admissible plate deflection. To calculate $\overline{w}_0$ and $\overline{\psi}$ we solve the dynamic problem for the whole plate. The remaining constant we keep as a parameter to tailor the deflection of the controlled plate. The dynamic problem to be solved is

$$\begin{aligned} & 0 \leqslant r \leqslant R_{\text{tot}}: \\ & D\left[\frac{\partial^4 w_0(r,t)}{\partial r^4} + \frac{2}{r}\frac{\partial^3 w_0(r,t)}{\partial r^3} + \frac{1}{r^2}\frac{\partial^2 w_0(r,t)}{\partial r^2} + \frac{1}{r^3}\frac{\partial w_0(r,t)}{\partial r}\right] \\ & \quad + \rho^{(0)}\ddot{w}_0(r,t) \\ & \quad = p_0(t) - \frac{1}{r}\frac{\partial m^*(r,t)}{\partial r} - \frac{\partial^2 m^*(r,t)}{\partial r^2}, \\ & \quad r = R_{\text{tot}}: \ w_0 = 0, \ \frac{\partial w_0}{\partial t} = 0. \end{aligned} \tag{20}$$

The time variation of the force loading $p_0(t)$ may be considered general. However, by considering the force loading is harmonic, we may find the solution for an arbitrary time variation by means of the Fourier integral; this argumentation is taken from Graff [17]. Therefore, we proceed with studying harmonic loading $p_0(t) = P_0 e^{i\omega t}$; ω is the driving frequency. The response may be written as $w_0(r,t) = W(r)e^{i\omega t}$ and $\breve{w}_0(r,t) = \breve{W}_0(r)e^{i\omega t}$. Inserting Eq. (17) into Eq. (20) results in the following boundary value problem in the frequency range.

$$\begin{aligned} & D\left[\frac{\partial^4 W(r)}{\partial r^4} + \frac{2}{r}\frac{\partial^3 W(r)}{\partial r^3} + \frac{1}{r^2}\frac{\partial^2 W(r)}{\partial r^2} + \frac{1}{r^3}\frac{\partial W(r)}{\partial r}\right] \\ & \quad - \rho^{(0)} P\omega^2 W(r) 0 \leqslant r \leqslant R_{\text{sub}}: \ = 0, \\ & R_{\text{sub}} \leqslant r \leqslant R: \ = -\rho^{(0)} P\omega^2 \breve{W}_0 + D\left[\frac{\partial^4 \breve{W}_0(r)}{\partial r^4} + \frac{2}{r}\frac{\partial^3 \breve{W}_0(r)}{\partial r^3}\right. \\ & \quad \left. + \frac{1}{r^2}\frac{\partial^2 \breve{W}_0(r)}{\partial r^2} + \frac{1}{r^3}\frac{\partial \breve{W}_0(r)}{\partial r}\right], \\ & \quad R \leqslant r \leqslant R_{\text{tot}}: \ = P_0, \\ & \quad r = R_{\text{tot}}: \ W = 0, \ \frac{\partial W}{\partial r} = 0 \end{aligned} \tag{21}$$

Thereby, we have taken care of Eqs. (18) and (19). Solving Eq. (21) separately in the three domains adds ten additional unknowns. These can be eliminated by adjusting the solution to the eight continuity conditions and to the two boundary conditions. Three unknowns still have to be calculated. The solution of Eq. (21), when evaluated at $r = R$, is exactly the unknown deflection at this location. Also, the directional derivative of the solution at $r = R$ is the unknown slope. Hence,

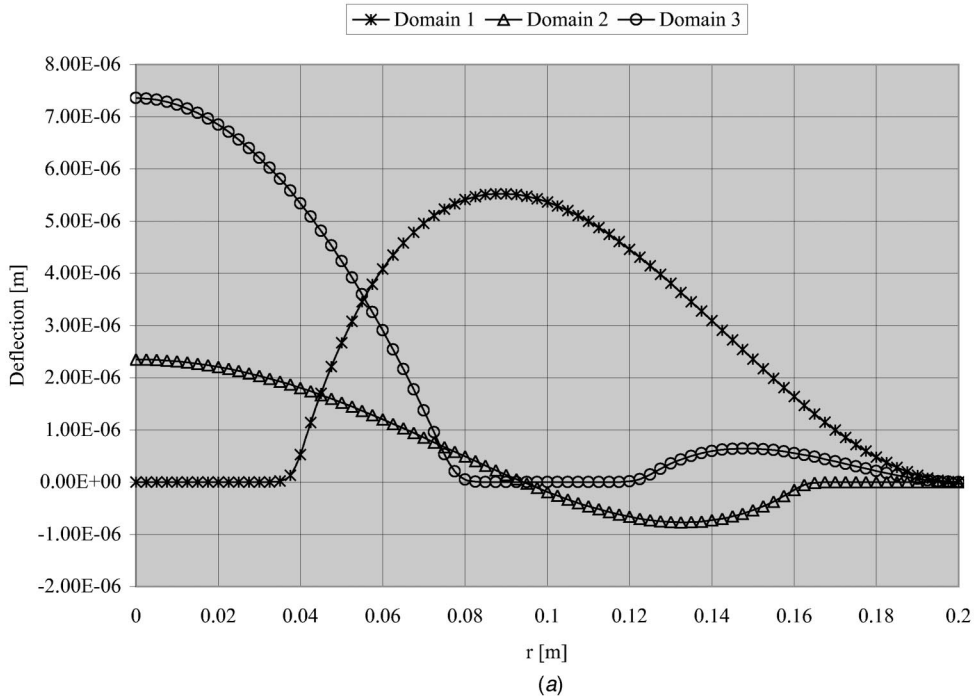
$$W(r=R)e^{i\omega t} = \overline{w}_0 = \overline{W}_0 e^{i\omega t}, \ \frac{\partial W}{\partial r}(r=R)e^{i\omega t} = \overline{\psi} = \overline{\Psi}e^{i\omega t}. \tag{22}$$

The remaining unknown reflects the non-uniqueness of the statically admissible moment. We use this latter unknown to tailor the deflection of the controlled plate in a desired manner.

4.2 Numerical Results. For a numerical simulation we consider a typical plate, with the following stiffness parameters, inertia parameter, geometry, and loading:

$$\begin{aligned} & D = 162.45 \ \text{N m}, \ \rho^{(0)} = 8.1 \ \text{kg m}^{-2}, \ P_0 = 100 \ \text{N m}^{-2}, \\ & R_{\text{tot}} = 0.2 \ \text{m}, \ R = 0.04 \ \text{m}, \ R_{\text{sub}} = 0.03 \ \text{m}. \end{aligned} \tag{23}$$

The first three natural frequencies of this plate are $f_1 = 182.0$ Hz, $f_2 = 708.7$ Hz, and $f_3 = 1588$ Hz. We consider different driving frequencies $\omega = 2\pi f = 2\pi(100,500,1000)\ \text{s}^{-1}$. The deflection of the plate $W(r)$ for $2\pi 100\ \text{s}^{-1}$ is presented in Fig. 5. The uncontrolled deflection is presented together with three types of controlled deflection. In the first case (''Controlled'') the remaining constant was set to zero. For ''Controlled; tailored deflection'' we



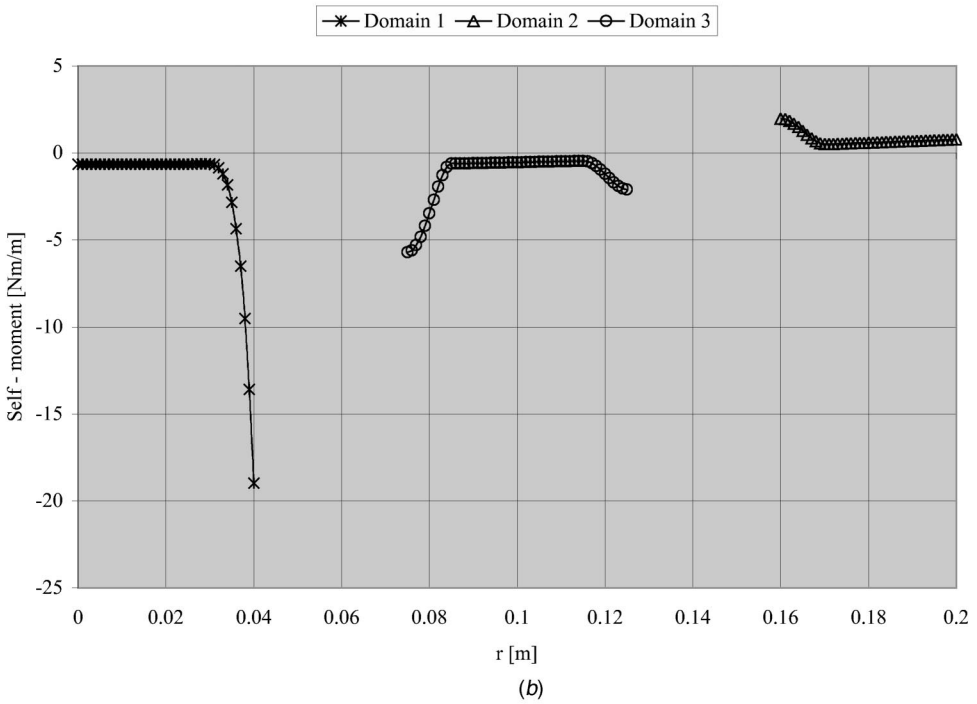
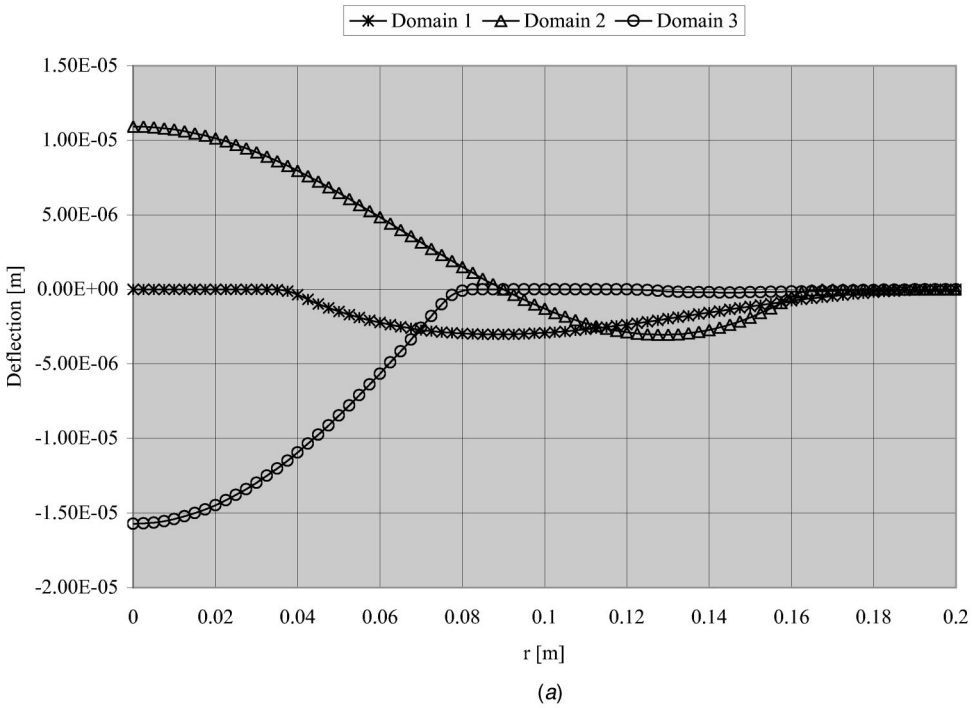
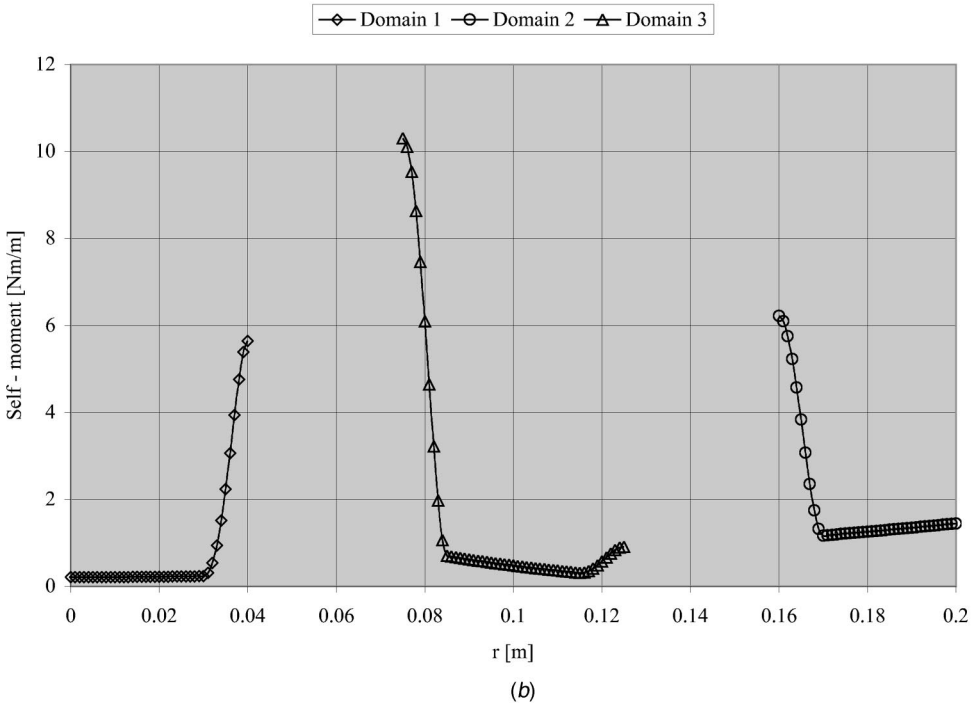


Fig. 10 (*a*) **Deflection and** (*b*) **self-moment of the clamped plate;** $\omega=2\pi 100\ \mathrm{s}^{-1}$

used the constant to make the deflection at $r=R$ zero and for ''Controlled; tailored self-moment'' we made the self-moment to have a horizontal tangent at $r=R$. Figure 6 shows the distribution of the self-moment that has to be applied to obtain the deflection of Fig. 5. It is obvious that the self-moment for ''Controlled; tailored self-moment'' requires the least effort to achieve the desired goal of eliminating the deflection in the domain $r \leqslant R_{\mathrm{sub}}$. The idea of enforcing the deflection at $r=R$ to be zero does not seem to be such a good idea and taking the remaining constant to be zero is not preferable to the case of the tailored self-moment. Whether this is also true for different driving frequencies remains to be studied. For that reason we consider driving frequencies $\omega=2\pi 500\ \mathrm{s}^{-1}$ and $\omega=2\pi 1000\ \mathrm{s}^{-1}$ next. Figures 7 and 8 show the results. With respect to the deflection the three methods are more or less identical. However, if we take a look at both the deflection and the self-moment distribution, we may conclude that the method of the tailored self-moment is the most suitable one; the deflection is nearly identical for the three methods and the required control effort is least for the method of the tailored self-moment. Our solution is exact; therefore, our solution is optimal



(*a*)



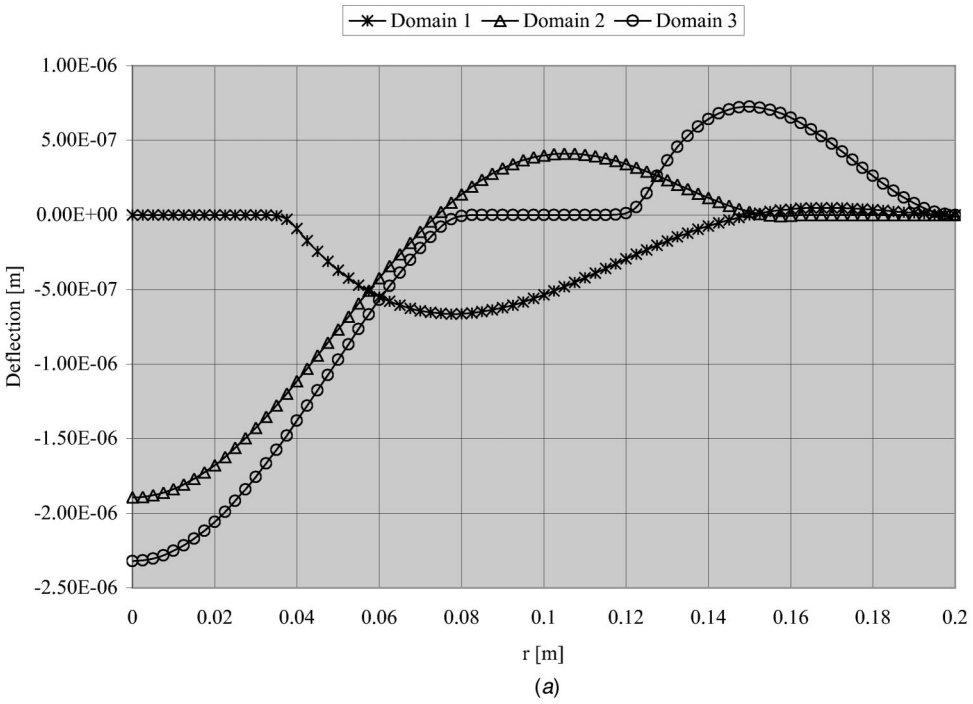
(*b*)

Fig. 11 (*a*) Deflection and (*b*) self-moment of the clamped plate; $\omega = 2\pi 500\ \mathrm{s}^{-1}$

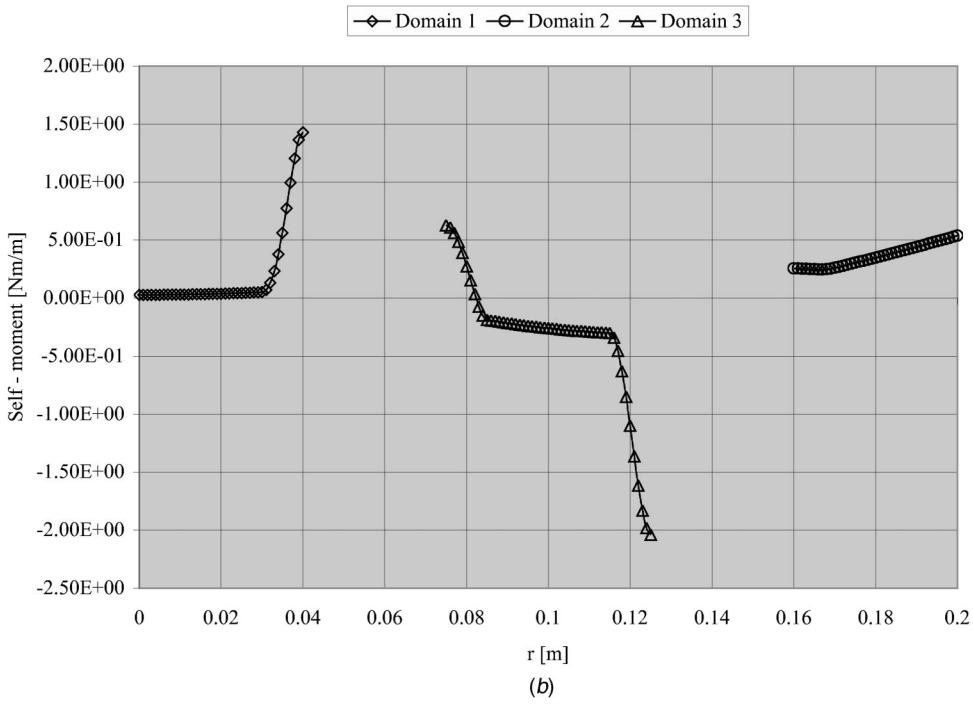
with respect to the proposed goal of eliminating the deflection in the subdomain. However, the deflection of the uncontrolled domain is significantly different to the deflection in the case of no control applied. Ideas to overcome this possibly undesirable behavior have been developed in Krommer and Varadan [12] and will also be discussed in Part II of this paper (Krommer and Varadan [9]).

4.2.1 Control of Alternative Subdomains. Finally, our interest is to control different domains of the circular plate. Those are either within $R_{\mathrm{sub}} \leqslant r \leqslant R_{\mathrm{tot}}$ or $R_{\mathrm{sub1}} \leqslant r \leqslant R_{\mathrm{sub2}}$; see Fig. 9. The domains where we apply control are $R \leqslant R_{\mathrm{sub}} \leqslant R_{\mathrm{tot}}$ or $R_1 \leqslant R_{\mathrm{sub1}} \leqslant R_{\mathrm{sub2}} \leqslant R_2$. The dimensions of the plate are

$$R_{\mathrm{tot}} = 0.2\ \mathrm{m},\ R = 0.16\ \mathrm{m},\ R_{\mathrm{sub}} = 0.17\ \mathrm{m},$$
$$R_1 = 0.075\ \mathrm{m},\ R_{\mathrm{sub1}} = 0.085\ \mathrm{m},\ R_{\mathrm{sub2}} = 0.115\ \mathrm{m},\tag{24}$$
$$R_2 = 0.125\ \mathrm{m}.$$



(*a*)



(*b*)

Fig. 12 (*a*) Deflection and (*b*) self-moment of the clamped plate; $\omega=2\pi 1000\ s^{-1}$

Without going into detail, we would like to mention that the method we used in the first example is applied in a straightforward manner. For the domain we want to control, we set $\breve{w}_0=0$, and for the transition domain, where we apply control, but which we do not control, we assume $\breve{w}_0$ in the form of a series similar to the one of Eq. (16). Furthermore, we shall only use that method, for which the self-moment has a horizontal tangent at those locations, where we release the part we apply control to from the rest of the plate. In Fig. 10 the results for $\omega=2\pi 100\ s^{-1}$ are presented. Results for the three different domains we are controlling are shown. ''Domain 1'' refers to the previous example, ''Domain 2'' is the domain next to the clamped boundary of the plate, and ''Domain 3'' is the interior domain. Figure 10(*a*) shows the deflection and the Fig. 10(*b*) shows the self-moment that has to be applied to achieve the corresponding deflection. Again, our method results in an exact elimination of the deflection; the solution is optimal. Finally, Figs. 11 and 12 show the results for $\omega=2\pi 500\ s^{-1}$ and $\omega=2\pi 1000\ s^{-1}$.

5 Conclusions

In the present paper we have presented a solution for the dynamic shape control of subdomains of linear elastic plates by applying self-stresses, which act in the linear elastic background plate. The solution is exact; hence, it represents an optimal solution to the following problem: How do we distribute a source of self-stress in the subdomain in order to eliminate the motion of the subdomain? It was our intention to present the theoretical foundation of the method in the present paper. For that reason we avoided a longer discussion of its possible practical applications; however, there is a number of applications we have in mind. We have mentioned a few in the introduction section and we will discuss them in more detail in the second part of the paper.

Acknowledgment

Support for M. Krommer from the MAX-KADE Foundation and the Austrian Academy of Sciences is gratefully acknowledged. Support for V. V. Varadan from the National Science Foundation is gratefully acknowledged.

Appendix

The constitutive relations (in the Laplace domain) for the two loading cases (original and dummy) are

$$\hat{\mathbf{M}}=\mathbf{D}:\hat{\boldsymbol{\kappa}}-\hat{\mathbf{M}}^{*},\quad \hat{\mathbf{M}}^{d}=\mathbf{D}:\hat{\boldsymbol{\kappa}}^{d}. \tag{A1}$$

We contract the first relation with the curvature tensor of the dummy loading case and we contract the second relation with the curvature tensor of the original loading case. The results are subtracted and integrated with respect to the plate area.

$$\int_A[\hat{\mathbf{M}}:\hat{\boldsymbol{\kappa}}^{d}-\hat{\mathbf{M}}^{d}:\hat{\boldsymbol{\kappa}}]dA=-\int_A\hat{\mathbf{M}}^{*}:\hat{\boldsymbol{\kappa}}^{d}dA \tag{A2}$$

The moment tensor is symmetric and the curvature tensor represents the symmetric part of the non-symmetric tensor $-\nabla\nabla w_0$. Hence, we have

$$\int_A[\hat{\mathbf{M}}:\hat{\boldsymbol{\kappa}}^{d}-\hat{\mathbf{M}}^{d}:\hat{\boldsymbol{\kappa}}]dA=-\int_A[\hat{\mathbf{M}}:\nabla\nabla\hat{w}_0^{d}-\hat{\mathbf{M}}^{d}:\nabla\nabla\hat{w}_0]dA. \tag{A3}$$

We reformulate Eq. (*A*3) by using the Gauss theorem twice.

$$\begin{aligned}
&-\int_A[\hat{\mathbf{M}}:\nabla\nabla\hat{w}_0^{d}-\hat{\mathbf{M}}^{d}:\nabla\nabla\hat{w}_0]dA\\
&=\int_A[\operatorname{div}\hat{\mathbf{M}}\cdot\nabla\hat{w}_0^{d}-\operatorname{div}\hat{\mathbf{M}}^{d}\cdot\nabla\hat{w}_0]dA\\
&\quad-\int_C[\hat{\mathbf{M}}\mathbf{n}\cdot\nabla\hat{w}_0^{d}-\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\nabla\hat{w}_0]dC\\
&=-\int_A[\operatorname{div}(\operatorname{div}\hat{\mathbf{M}})\hat{w}_0^{d}-\operatorname{div}(\operatorname{div}\hat{\mathbf{M}}^{d})\hat{w}_0]dA\\
&\quad+\int_C[\operatorname{div}\hat{\mathbf{M}}\cdot\mathbf{n}\hat{w}_0^{d}-\operatorname{div}\hat{\mathbf{M}}^{d}\cdot\mathbf{n}\hat{w}_0]dC\\
&\quad-\int_C[\hat{\mathbf{M}}\mathbf{n}\cdot(\nabla\hat{w}_0^{d}\cdot\mathbf{n})\mathbf{n}-\hat{\mathbf{M}}^{d}\mathbf{n}\cdot(\nabla\hat{w}_0\cdot\mathbf{n})\mathbf{n}]dC\\
&\quad-\int_C[\hat{\mathbf{M}}\mathbf{n}\cdot(\nabla\hat{w}_0^{d}\cdot\mathbf{s})\mathbf{s}-\hat{\mathbf{M}}^{d}\mathbf{n}\cdot(\nabla\hat{w}_0\cdot\mathbf{s})\mathbf{s}]dC
\end{aligned} \tag{A4}$$

Note the decomposition of the displacement gradient into the normal and tangential direction. This step has been performed in order to proceed with

$$\begin{aligned}
&-\int_C[\hat{\mathbf{M}}\mathbf{n}\cdot(\nabla\hat{w}_0^{d}\cdot\mathbf{s})\mathbf{s}-\hat{\mathbf{M}}^{d}\mathbf{n}\cdot(\nabla\hat{w}_0\cdot\mathbf{s})\mathbf{s}]dC\\
&=\int_C[\nabla(\hat{\mathbf{M}}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s}\hat{w}_0^{d}-\nabla(\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s}\hat{w}_0]dC\\
&\quad-\sum_i[\hat{\mathbf{M}}\mathbf{n}\cdot\mathbf{s}\hat{w}_0^{d}-\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\mathbf{s}\hat{w}_0]_{P_i^+}^{P_i^-}.
\end{aligned} \tag{A5}$$

The last term accounts for all points P_i of C that are located at a corner. We use the partial differential equation, which governs the bending motion, to reformulate the term $-\int_A[\operatorname{div}(\operatorname{div}\hat{\mathbf{M}})\hat{w}_0^{d}-\operatorname{div}(\operatorname{div}\hat{\mathbf{M}}^{d})\hat{w}_0]dA$. For the two loading cases we have $\operatorname{div}(\operatorname{div}\hat{\mathbf{M}})+\hat{p}_z=\rho^{(0)}(s^2\hat{w}_0-sw_0(0)-\dot{w}_0(0))$ and $\operatorname{div}(\operatorname{div}\hat{\mathbf{M}}^{d})+\hat{p}_z^{d}=\rho^{(0)}s^2\hat{w}_0^{d}$, because the initial conditions in the dummy case are homogenous. The result is

$$\begin{aligned}
&-\int_A[\operatorname{div}(\operatorname{div}\hat{\mathbf{M}})\hat{w}_0^{d}-\operatorname{div}(\operatorname{div}\hat{\mathbf{M}}^{d})\hat{w}_0]dA\\
&=\int_A[(\hat{p}_z+\rho^{(0)}(sw_0(0)+\dot{w}_0(0)))\hat{w}_0^{d}-\hat{p}_z^{d}\hat{w}_0]dA\\
&\quad-\int_A\Big[\underbrace{\rho^{(0)}s^2\hat{w}_0\hat{w}_0^{d}-\rho^{(0)}s^2\hat{w}_0^{d}\hat{w}_0}_{=\mathbf{0}}\Big]dA.
\end{aligned} \tag{A6}$$

We insert Eqs. (*A*3), (*A*4), (*A*5), and (*A*6) into Eq. (*A*2).

$$\begin{aligned}
&\int_A[(\hat{p}_z+\rho^{(0)}(sw_0(0)+\dot{w}_0(0)))\hat{w}_0^{d}-\hat{p}_z^{d}\hat{w}_0]dA+\int_C[(\operatorname{div}\hat{\mathbf{M}}\cdot\mathbf{n}\\
&\quad+\nabla(\hat{\mathbf{M}}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s})\hat{w}_0^{d}-(\operatorname{div}\hat{\mathbf{M}}^{d}\cdot\mathbf{n}+\nabla(\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s})\hat{w}_0]dC\\
&=\int_C[\hat{\mathbf{M}}\mathbf{n}\cdot(\nabla\hat{w}_0^{d}\cdot\mathbf{n})\mathbf{n}-\hat{\mathbf{M}}^{d}\mathbf{n}\cdot(\nabla\hat{w}_0\cdot\mathbf{n})\mathbf{n}]dC+\sum_i[\hat{\mathbf{M}}\mathbf{n}\cdot\mathbf{s}\hat{w}_0^{d}\\
&\quad-\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\mathbf{s}\hat{w}_0]_{P_i^+}^{P_i^-}-\int_A\hat{\mathbf{M}}^{*}:\hat{\boldsymbol{\kappa}}^{d}dA
\end{aligned} \tag{A7}$$

Next, we introduce the transformed transverse force $\hat{\bar{p}}_z=\hat{p}_z+\rho^{(0)}(sw_0(t=0)+\dot{w}_0(t=0))$, which has already been introduced in the main part of the paper. Furthermore, by taking into account the boundary conditions in the original problem and that the boundary conditions in the dummy loading case are homogenous, we can reformulate the boundary integrals in Eq. (*A*7) as

$$\begin{aligned}
&\int_C[(\operatorname{div}\hat{\mathbf{M}}\cdot\mathbf{n}+\nabla(\hat{\mathbf{M}}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s})\hat{w}_0^{d}-(\operatorname{div}\hat{\mathbf{M}}^{d}\cdot\mathbf{n}\\
&\quad+\nabla(\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s})\hat{w}_0]dC\\
&=\int_{C_q}\hat{\bar{q}}\hat{w}_0^{d}dC-\int_{C_w}(\operatorname{div}\hat{\mathbf{M}}^{d}\cdot\mathbf{n}+\nabla(\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s})\hat{\bar{w}}_0dC,\\
&\quad\int_C[\hat{\mathbf{M}}\mathbf{n}\cdot(\nabla\hat{w}_0^{d}\cdot\mathbf{n})\mathbf{n}-\hat{\mathbf{M}}^{d}\mathbf{n}\cdot(\nabla\hat{w}_0\cdot\mathbf{n})\mathbf{n}]dC\\
&=\int_{C_m}\hat{\bar{m}}(\nabla\hat{w}_0^{d}\cdot\mathbf{n})dC-\int_{C_\psi}(\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\mathbf{n}\hat{\bar{\psi}}dC,\\
&\quad\times\sum_i[\hat{\mathbf{M}}\mathbf{n}\cdot\mathbf{s}\hat{w}_0^{d}-\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\mathbf{s}\hat{w}_0]_{P_i^+}^{P_i^-}\\
&=-\sum_i[\hat{\mathbf{M}}^{d}\mathbf{n}\cdot\mathbf{s}\hat{w}_0]_{P_{wi}^+}^{P_{wi}^-}.
\end{aligned} \tag{A8}$$

The last term takes into account that, if the deflection at a corner is not prescribed, then the corner forces $[\hat{\mathbf{M}}\mathbf{n}\cdot\mathbf{s}]_{P_i^+}^{P_i^-}$ and $[\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s}]_{P_i^+}^{P_i^-}$ have to vanish. If the deflection is prescribed at the corner, then $\hat{w}_0^d=0$, but $\hat{w}_0$ may be nontrivial; such points are denoted as P_{wi}. We insert the relations of Eq. (*A*8) into Eq. (*A*7), such that we finally end up with the integral representation that has been stated in Eq. (3).

$$\int_A [\hat{\tilde{p}}_z\hat{w}_0^d-\hat{p}_z^d\hat{w}_0]dA+\int_{C_q}\hat{\tilde{q}}\hat{w}_0^d dC - \int_{C_w}(\operatorname{div}\hat{\mathbf{M}}^d\cdot\mathbf{n}+\nabla(\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s})\hat{\tilde{w}}_0 dC = \int_{C_m}\hat{\tilde{m}}(\nabla\hat{w}_0^d\cdot\mathbf{n})dC-\int_{C_\psi}\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{n}\hat{\tilde{\psi}}dC - \int_A\hat{\mathbf{M}}^*:\hat{\boldsymbol{\kappa}}^d dA\hat{\tilde{\psi}}\hat{\tilde{\psi}}dC-\int_A\hat{\mathbf{M}}^*:\hat{\boldsymbol{\kappa}}^d dA - \sum_i[\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s}\hat{w}_0]_{P_{wi}^+}^{P_{wi}^-} \tag{A9}$$

In order the derive the integral representation of Eq. (5), we utilize the decomposition of the transverse displacement, $w_0=\tilde{w}_0+\breve{w}_0$, as well as the governing partial differential equation for the motion in the dummy loading case, $\operatorname{div}(\operatorname{div}\hat{\mathbf{M}}^d)+\hat{p}_z^d=\rho^{(0)}s^2\hat{w}_0^d$. We find

$$\int_A\hat{p}_z^d\hat{w}_0dA=\int_A\hat{p}_z^d\hat{\tilde{w}}_0dA+\int_A\hat{p}_z^d\hat{\breve{w}}_0dA = \int_A\hat{p}_z^d\hat{\tilde{w}}_0dA+\int_A\rho^{(0)}s^2\hat{w}_0^d\hat{\tilde{w}}_0dA - \int_A\operatorname{div}(\operatorname{div}\hat{\mathbf{M}}^d)\hat{\tilde{w}}_0dA. \tag{A10}$$

If we apply the Gauss theorem twice to the last term on the right hand side of Eq. (*A*10), then we end up with

$$-\int_A\operatorname{div}(\operatorname{div}\hat{\mathbf{M}}^d)\hat{\tilde{w}}_0dA=\int_A\hat{\mathbf{M}}^d:\hat{\tilde{\boldsymbol{\kappa}}}dA-\int_C(\operatorname{div}\hat{\mathbf{M}}^d\cdot\mathbf{n} +\nabla(\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s})\hat{\tilde{w}}_0dC+\int_C(\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{n}) \times(\nabla\hat{\tilde{w}}_0\cdot\mathbf{n})dC+\sum_i[\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s}\hat{\tilde{w}}_0]_{P_i^+}^{P_i^-}. \tag{A11}$$

Note that we considered the definition of the curvature tensor and that we decomposed the gradient of the transverse displacement into its normal and tangential direction. The last term in Eq. (*A*11) becomes $\Sigma_i[\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s}\hat{\tilde{w}}_0]_{P_{wi}^+}^{P_{wi}^-}$ (see the argumentation above) and the boundary integrals simplify if we account for the boundary conditions. In the dummy loading case the dynamic ones are homogenous and in the original loading case we have $C_w:\hat{\tilde{w}}_0=\hat{\breve{w}}_0$ and $C_\psi:\nabla\hat{\tilde{w}}_0\cdot\mathbf{n}=\hat{\tilde{\psi}}$. Hence,

$$\int_A\hat{p}_z^d\hat{w}_0dA=\int_A\hat{p}_z^d\hat{\tilde{w}}_0dA+\int_A\rho^{(0)}s^2\hat{w}_0^d\hat{\tilde{w}}_0dA+\int_A\hat{\mathbf{M}}^d:\hat{\tilde{\boldsymbol{\kappa}}}dAz +\sum_i[\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s}\hat{\tilde{w}}_0]_{P_{wi}^+}^{P_{wi}^-} -\int_{C_w}(\operatorname{div}\hat{\mathbf{M}}^d\cdot\mathbf{n}+\nabla(\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s})\hat{\breve{w}}_0dC +\int_{C_\psi}(\hat{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{n})\hat{\breve{\psi}}dC. \tag{A12}$$

Due to symmetry, the relation $\int_A\hat{\mathbf{M}}^d:\hat{\tilde{\boldsymbol{\kappa}}}dA=\int_A\hat{\breve{\mathbf{M}}}:\hat{\boldsymbol{\kappa}}^d dA$, in which $\breve{\mathbf{M}}=\mathbf{D}:\breve{\boldsymbol{\kappa}}$ has been introduced, holds. It remains to reformulate the term $\int_A\hat{\breve{\mathbf{M}}}:\hat{\boldsymbol{\kappa}}^d dA$. For that sake we insert the definition of the dummy curvature tensor, apply the Gauss theorem twice, and keep in mind that kinematic boundary conditions in the dummy loading case are homogenous

$$\int_A\hat{\breve{\mathbf{M}}}:\hat{\boldsymbol{\kappa}}^d dA=-\int_A\hat{\breve{\mathbf{M}}}:\left[\frac{1}{2}\nabla\nabla\hat{w}_0^d+(\nabla\nabla\hat{w}_0^d)^T\right]dA =-\int_A\hat{\breve{\mathbf{M}}}:\nabla\nabla\hat{w}_0^d dA =-\int_A\operatorname{div}(\operatorname{div}\hat{\breve{\mathbf{M}}})\hat{w}_0^d dA-\int_{C_m}\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{n}(\nabla\hat{w}_0^d\cdot\mathbf{n})dC +\int_{C_q}(\operatorname{div}\hat{\breve{\mathbf{M}}}\cdot\mathbf{n}+\nabla(\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s})\hat{w}_0^d dC -\sum_i[(\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{s})\hat{w}_0^d]_{P_i^+}^{P_i^-}. \tag{A13}$$

To obtain the integral representation of Eq. (5) and the definitions of Eq. (6) we insert Eq. (*A*13) into Eq. (*A*12) and we insert the result into Eq. (*A*9).

$$\int_A\underbrace{\hat{\tilde{p}}_z-\rho^{(0)}s^2\hat{\tilde{w}}_0+\operatorname{div}(\operatorname{div}\hat{\breve{\mathbf{M}}})}_{\hat{\tilde{p}}_z}\hat{w}_0^d dA -\int_{C_n}\left[\underbrace{\hat{\tilde{m}}-\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{n}}_{\hat{\tilde{m}}}\right](\nabla\hat{w}_0^d\cdot\mathbf{n})dC +\int_{C_q}\underbrace{\hat{\tilde{q}}-\operatorname{div}\hat{\breve{\mathbf{M}}}\cdot\mathbf{n}+\nabla(\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{s})\cdot\mathbf{s}}_{\hat{\tilde{q}}}\hat{w}_0^d dC +\int_A\hat{\mathbf{M}}^*:\hat{\boldsymbol{\kappa}}^d dA=\int_A\hat{p}_z^d\hat{\tilde{w}}_0dA+\sum_i\left[(\breve{\mathbf{M}}^d\mathbf{n}\cdot\mathbf{s})\underbrace{(\hat{\tilde{w}}_0-\hat{w}_0)}_{0}\right]_{P_{wi}^+}^{P_{wi}^-} -\sum_i[(\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{s})\hat{w}_0^d]_{P_i^+}^{P_i^-}. \tag{A14}$$

The last but one term vanishes, because in every corner where the deflection is prescribed the relation $\hat{\tilde{w}}_0=\hat{w}_0=\hat{\breve{w}}_0$ must be satisfied. The last term vanishes, if $\hat{\tilde{w}}_0$ is an admissible plate deflection. By admissible we mean that we are able to find an associated bending problem for the plate, which satisfies the non-homogenous kinematical boundary conditions of the original problem and for which $\hat{\tilde{w}}_0$ is a solution. If so, $(\hat{\breve{\mathbf{M}}}\mathbf{n}\cdot\mathbf{s})\delta\hat{\tilde{w}}_0]_{P_i^+}^{P_i^-}$

$=0$. In every corner point P_i, in which the deflection is not prescribed $(\hat{\tilde{\mathbf{M}}}\mathbf{n}\cdot\mathbf{s})]_{P_i^+}^{P_i^-}=0$ is satisfied; for corner points with the deflection prescribed $\hat{w}_0^d=0$ is satisfied. Hence, the last term in Eq. (*A*14) vanishes. Then Eq. (*A*14) is identical to Eq. (5).

References

[1] Haftka, R. T., and Adelman, H. M., 1985, ''An Analytical Investigation of Static Shape Control of Large Space Structures by Applied Temperature,'' AIAA J., **23**, pp. 450–457.

[2] Irschik, H., and Pichler, U., 2001, ''Dynamic Shape Control of Solids and Structures by Thermal Expansion Strains,'' J. Therm. Stresses, **24**, pp. 565–578.

[3] Irschik, H., 2002, ''A Review on Static and Dynamic Shape Control of Structures by Piezoelectric Actuation,'' Eng. Struct., **24**, pp. 5–11.

[4] Reissner, H., 1931, ''Selbstspannungen Elastischer Gebilde,'' Z. Angew. Math. Mech., **11**, pp. 59–70.

[5] Nemenyi, P., 1931, ''Eigenspannungen und Eigenspannungsquellen,'' Z. Angew. Math. Mech., **11**, pp. 1–8.

[6] Mura, T., 1991, *Micromechanics of Defects in Solids*, 2nd ed., Kluwer, Dordrecht.

[7] Rao, S. S., and Sunar, M., 1994, ''Piezoelectricity and Its use in Disturbance Sensing and Control of Flexible Structures: a Survey,'' Appl. Mech. Rev., **47**, pp. 113–123.

[8] Gopinathan, S. V., Varadan, V. V., and Varadan, V. K., 2001, ''Active Noise Control Studies Using the Rayleigh-Ritz Method,'' *Proceedings of IUTAM-Symposium on Smart Structures and Structronic Systems*, U. Gabbert and H. S. Tzou, eds., Kluwer, Dordrecht, pp. 169–177.

[9] Krommer, M., and Varadan, V. V., 2004, ''Control of Bending Vibrations Within Sub-Domains of Thin Plates. Part II: Piezoelectric Actuation and Optimization,'' ASME J. Appl. Mech, to be submitted.

[10] Nader, M., Gattringer, H., Krommer, M., and Irschik, H., 2003, ''Shape Control of Flexural Vibrations of Circular Plates by Piezoelectric Actuation,'' ASME J. Vibr. Acoust., **125**, pp. 88–94.

[11] Irschik, H., Schlacher, K., and Haas, W., 1997, ''Output Annihilation and Optimal H2 Control of Plate Vibrations by Piezoelectric Actuation,'' *Proceedings of the IUTAM-Symposium on Interactions Between Dynamics and Control in Advanced Mechanical Systems*, D. H. van Campen, ed., Kluwer, Dordrecht, pp. 159–166.

[12] Krommer, M., and Varadan, V. V., 2003, ''Dynamic Shape Control of Conformal Antennas,'' *Proc. of SPIE's 10th Annual International Symposium on Smart Materials and Structures*, R. C. Smith, ed., Proc. SPIE Vol. 5049 (Smart Structures and Materials 2003: Modeling, Signal Processing and Control), pp. 622–630.

[13] Nader, M., von Gassen, H. G., Krommer, M., and Irschik, H., 2003, ''Piezoelectric Actuation of Thin Shells With Support Actuation,'' *Proc. of SPIE's 10th Annual International Symposium on Smart Materials and Structures*, R. C. Smith, ed., Proc. SPIE Vol. 5049 (Smart Structures and Materials 2003: Modeling, Signal Processing and Control), pp. 180–189.

[14] Krommer, M., 2003, ''On the Significance of Non-Local Constitutive Relations for Composite Thin Plates Including Piezoelastic Layers With Prescribed Electric Charge,'' Smart Mater. Struct., **12**, pp. 318–330.

[15] Ziegler, F., 1991, *Mechanics of Solids and Fluids*, Springer, New York.

[16] Ziegler, F., and Irschik, H., 1987, ''Thermal Stress Analysis Based on Maysel's Formula,'' *Thermal Stresses 2*, R. B. Hetnarski, ed., North-Holland, Amsterdam, pp. 120–188.

[17] Graff, K. F., 1975, *Wave Motion in Elastic Solids*, Clarendon Press, Oxford.

Analysis of Doubly Clamped Nanotube Devices in the Finite Deformation Regime

N. Pugno[1]

C. H. Ke

H. D. Espinosa[2]
e-mail: espinosa@northwestern.edu

Department of Mechanical Engineering,
Northwestern University,
Evanston, IL 60208-3111

In this paper, a nonlinear theory applicable to the design of nanotube based devices is presented. The role of finite kinematics for a doubly clamped nanotube device is investigated. In particular, we analyze the continuous deformation and instability (pull in) of a clamped-clamped nanotube suspended over an electrode from which a potential differential is imposed. The transformation of an applied voltage into a nanomechanical deformation indeed represents a key step toward the design of innovative nanodevices. Likewise, accurate prediction of pull-in/pull-out voltages is highly needed. We show that an energy-based method can be conveniently used to predict the structural behavior and instability corresponding to the ON/OFF states of the device at the so-called pull-in voltage. The analysis reveals that finite kinematics effects can result in a significant increase of the pull-in voltage. This increase results from a ropelike behavior of the nanotube as a consequence of the stretching imposed by the actuation.
[DOI: 10.1115/1.1875452]

1 Introduction

Nanoelectromechanical systems (NEMS) are attracting much interest in the scientific community, since the discovery of nanotubes [1]. The first really true carbon-nanotube-based NEMS device, fully integrating electronic control and mechanical response, was developed only some months ago [2]. The authors reported the construction and successful operation of a fully synthetic nanoscale electromechanical actuator incorporating a rotable metal plate, with a multi-walled carbon nanotube serving as the key motion-enabling element. Rueckes et al. [3] investigated a carbon nanotube-based nonvolatile random access memory by developing an innovative bistable nanoswitch based on electrostatic and van der Waals forces. The authors emphasized the extreme high integration level of the nanoswitches, approaching 10^{12} elements per square centimeter, and an element operation frequency in excess of 100 GHz. The viability of the concept was demonstrated by the experimental realization of a reversible bistable nanotube-based bit. In [4] the development of nanotweezers was reported. The mechanical capabilities of the nanotweezers were demonstrated by grapping and manipulating submicron clusters and nanowires.

In this context, the characterization of mechanical and electronic properties of nanotubes has been the subject of intense research. Their small size, low density, high stiffness, flexibility, and strength, as well as excellent electronic properties, suggest that nanotubes and nanowires are the most promising nanoscopic elements in the implementation of NEMS. For a recent review on the mechanics of carbon nanotubes the reader should refer to the paper by Qian et al. [5], and references therein. The strength of carbon nanotubes [6], was found to be of the order of 10–100 GPa. Furthermore, nanotubes (as well as nanoropes—composed of several nanotubes—and nanowires—having differenltly shaped cross sections) possess an extremely high stiffness (Young's modulus of the order of 1 TPa [7,8]) and flexibility (strain at tensile failure of the order of 30% [9]). As a consequence of this large flexibility, the effect of the large displacements, usually neglected in analytical calculations, has to be considered in the analysis of NEMS.

In spite of the described fast acceleration in developing NEMS structures, key formulas needed in their design are still absent in the literature. The first extensive investigation of the behavior of nanotube-based devices has been recently reported [10]. In that paper, the differential equation of the elastic line of a nanotube suspended over an electrode and from which a differential in potential is imposed, was numerically solved according to continuum mechanics, assuming small displacements. The corresponding pull-in voltages, at the structural instability, were evaluated for different case studies. In addition, the first attempt to obtain an analytical formula for the pull-in voltage of the nanotube was also proposed, assuming for the nanotube a *platelike* undeformed shape, connected via a lamped stiffness to the ground electrode. As emphasized by the same authors, the proposed formula was not able to reproduce accuratetly all their numerical results.

In this paper we present a nonlinear energy-based theory for the prediction of the pull-in voltage of doubly clamped nanotubes under stretching. The equilibrium condition as well as the instability of the nanotube is obtained, respectively, by setting to zero the first and the second derivatives of the free energy of the system. A comparison between analytically predicted pull-in voltages and those obtained by numerically solving the corresponding governing equations is also provided.

2 Elastic Line Equation of the Nanotube Under Finite Kinematics

In this section we derive, in the finite deformation regime, the elastic line equation for a nanotube. We focus the attention on a doubly clamped nanotube suspended over an electrode at a distance $r=H$ from which a difference V in the electrostatic potential is imposed (nanoswitch) which is schematically shown in Fig. 1. Note that this is equivalent to the problem of two identical doubly clamped nanotubes placed at distance $2r=2H$ under a difference in voltage $2V$, as imposed by the symmetry.

[1]On leave from the Department of Structural Engineering, Politecnico di Torino, Torino, Italy.

[2]To whom correspondence should be addressed.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, August 23, 2004; final revision, September 23, 2004. Editor: K. R. M. MeMeeking. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California - Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication in the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

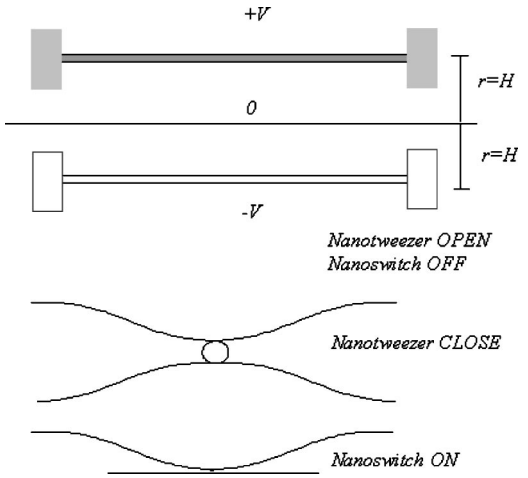


Fig. 1 Schematics of doubly clamped nanotube based nanoswitches and nanotweezers

The electrostatic and van der Waals energies per unit length can be evaluated by the following relationships [10]:

$$\frac{dE_{elec}}{dz} = \frac{\pi\varepsilon_0 V^2}{\cosh^{-1}\left(1+\frac{r}{R}\right)},$$

$$\frac{dE_{vdW}}{dz} = \sum_{R=R_{int}}^{R_{ext}} \sum_{r=r_{int}}^{r_{int}+(N_G-1)d} \frac{\pi^2 C_6 n^2 d^2 R(r+R)[3R^2+2(r+R)^2]}{2[(r+R)^2-R^2]^{7/2}}, \tag{1}$$

where z is the axial coordinate of the cantilever nanotube, R_{int} and $R \equiv R_{ext}$ are the inner and outer radius of a multiwalled nanotube, N_G is the number of layers in the substrate (graphene), d is the interlayer distance (for graphite $d=0.335$ nm). In addition, $r \equiv r_{int}$ is the gap between the nanotube (external wall) and the surface layer of the substrate, where n is the atomic density, that for graphite is equal to $n=1.14\times10^{29}\ m^{-3}$, and $\varepsilon_0=8.85\times10^{-12}\ C^2\ N^{-1}\ m^{-2}$ is the vacuum permittivity.

The corresponding forces per unit length q_{elec} and q_{vdW} can be evaluated, according to Eq. (1) as

$$q_{elec} = -\frac{d(dE_{elec}/dz)}{dr}, \quad q_{vdW} = -\frac{d(dE_{vdW}/dz)}{dr}. \tag{2}$$

Based on continuum mechanics, the quasistatic structural behavior of the nanotube, can be obtained solving the elastic line equation, namely,

$$EI\frac{d^4w}{dz^4} = q_{vdW} + q_{elec}, \quad I = \frac{\pi(R_{ext}^4 - R_{int}^4)}{4}, \tag{3}$$

where $w(z)=H-r(z)$ is the nanotube deflection, H is the nominal gap between nanotube and electrode, and E is the Young's modulus of the nanotube, with moment of inertia I.

It is important to underline that Eq. (3) assumes small displacements. On the other hand, due to the large flexibility of the nanotube, the role of the finite kinematics (large displacements) could become relevant. According to these considerations, we have to consider the complete expression for the elastic curvature. In addition, it is important to note that large deformations could imply, for doubly clamped nanotubes, also the stretching of the element. Finally, under large deformations, the electrostatic forces, orthogonal to the surface of the nanotube, have to be considered with respect to the deformed configuration (we neglect the effect of the finite kinematics on the Lennard-Jones forces per unit length, i.e., van der Waals and Pauli forces, which become significant only for very small gaps, i.e., few nanometers). In the dynamic regime, the damping and inertia forces must be also added (e.g., to consider variable applied voltages, thermal vibrations, free vibrations, and so on). According to these considerations, the complete expression of the elastic line equation for a nanotube device is

$$EI\frac{\partial^2}{\partial z^2}\left(\frac{\frac{\partial^2 w}{\partial z^2}}{\left[1+\left(\frac{\partial w}{\partial z}\right)^2\right]^{3/2}}\right) - \frac{EA}{2L}\int_0^L \left(\frac{\partial w}{\partial z}\right)^2 dz \frac{\frac{\partial^2 w}{\partial z^2}}{\left[1+\left(\frac{\partial w}{\partial z}\right)^2\right]^{3/2}}$$
$$= q_{vdW} + q_P + \left(q_{elec} - \gamma\frac{\partial w}{\partial t} - \mu\frac{\partial^2 w}{\partial t^2}\right)\left[1+\left(\frac{\partial w}{\partial z}\right)^2\right]^{1/2}, \tag{4}$$

where μ and γ represent the mass and the damping per unit length of the nanotube and t is the time. The Pauli force per unit length q_P is obtained from the repulsive part of the Lennard-Jones potential [10]. From the Q-factor of the nanotube (between 170–500 [11]), $\gamma=\mu\omega_0/Q$, where ω_0 is its fundamental rotating frequency. The term $[1+(\partial w/\partial z)^2]^{3/2}$ represents the correction for the curvature, that must be considered under large displacements. The term $\cos\vartheta=(1+(\partial w/\partial z)^2)^{-1/2}$ has to be introduced to consider the changing in the positions of the loads that remain pependicular to the nanotube axis, as a consequence of the large displacements, involving not necessarily small rotations of the cross-section by an angle ϑ. For a clamped-clamped nanotube the axial force is equal to $N(w)=EA/2L\int_0^L(\partial w/\partial z)^2dz$.

Some interesting results were obtained [10] by solving numerically Eq. (3). The more general Eq. (4) was also solved numerically [12,13]. On the other hand, here we prefer obtaining an analytical solution under simplified hypotheses for the pull-in voltage, corresponding to the quasistatic collapse of the nanotube, i.e., assuming $\gamma=\mu=0$. Instead of solving Eq. (4) in an approximate way, we will obtain the equilibrium and the instability of the nanosystem by minimizing the free energy and its first derivative.

3 Small Deformation

We consider a clamped-clamped nanotube of length L. For the small deflection case of a clamped-clamped nanotube loaded by a constant force per unit length, we assume a function $w(z)$ satisfying the boundary conditions $w(z=0,L)=w'(z=0,L)=0$, namely,

$$w(z) \approx 16\left[\left(\frac{z}{L}\right)^2 - 2\left(\frac{z}{L}\right)^3 + \left(\frac{z}{L}\right)^4\right]c, \tag{5}$$

where $w(z=L/2)=c$ is here an unknown constant that represents the displacement of the central point.

As a consequence, the elastic energy, assuming small displacements, as well as the electrostatic and van der Waals energies stored in the nanotube can be obtained by integration as

$$E_{elast}(c) = \frac{EI}{2}\int_0^L \left(\frac{d^2w}{dz^2}\right)^2 dz, \tag{6a}$$

$$E_{elec}(c) \approx \int_0^L \frac{dE_{elec,vdW,P}\{r[w(z)]\}}{dz}dz. \tag{6b}$$

We investigate the validity of the form of Eq. (5) by evaluating the associated fundamental frequency and by comparing it with the well-known value for a clamped-clamped nanotube. Equating the maximum values of the elastic strain energy of Eq. (6*a*) and of the kinetic energy $K(t)=1/2\int_0^L(dw/dt)^2\mu dz$, with μ mass per unit length of the nanotube, during its free-vibration with $w(z,t) \approx w(z)\sin\omega_0 t$, one finds the estimation of the fundamental frequency ω_0 of the nanotube. The ratio between the estimated fundamental frequency and the real one is found to be close to 1. We conclude that the form of Eq. (5) is good for our scope.

The free-energy (or total potential energy) of the system can be written as

$$W(c) = E_{\text{elast}}(c) - E_{\text{elec}}(c) - E_{\text{vdW}}(c) - E_{\text{P}}(c). \tag{7a}$$

Equilibrium and stability are obtained from

$$\frac{\mathrm{d}W(c)}{\mathrm{d}c} = 0, \tag{7b}$$

$$\frac{\mathrm{d}^2W(c)}{\mathrm{d}c^2} = 0. \tag{7c}$$

The equilibrium condition is reached when the free-energy reaches a minimum value (Eq. (7*b*)). On the other hand, the structural instability occurs at the so-called pull-in voltage, when the second order of the free-energy becomes zero (Eq. (7*c*)). According to [10], the effects of the van der Walls and Pauli forces for these boundary conditions is negligible, even for small gaps; hence, we take $E_{\text{vdW,P}} \approx 0$.

The electrostatic energy per unit length can be approximated as

$$\frac{\mathrm{d}E_{\text{elec}}}{\mathrm{d}z} \approx \frac{\pi\varepsilon_0 V^2}{\ln\left(\frac{2(R+H-w)}{R}\right)} = \frac{\pi\varepsilon_0 V^2}{\ln\left(\frac{2(H+R)}{R}\right)} \times \left[1 + \sum_{i=1}^{\infty}\left(\frac{1}{\ln\left(\frac{2(H+R)}{R}\right)}\sum_{j=1}^{\infty}\frac{1}{i}\left(\frac{w}{(H+R)}\right)^j\right)^i\right]. \tag{8}$$

Employing Eq. (6*b*), the total electrostatic energy can be expressed as

$$E_{\text{elec}}(c) = \frac{\pi\varepsilon_0 V^2 L}{\ln\left(\frac{2(H+R)}{R}\right)} \times \left[1 + \sum_{i=1}^{\infty}\left(\frac{1}{\left[\ln\left(\frac{2(H+R)}{R}\right)\right]^i}\sum_{j=i}^{\infty}a_{ij}\left(\frac{c}{(H+R)}\right)^j\right)\right], \tag{9}$$

where $\{a_{ij}\}$ are constants. Let $S(c) = \sum_{i=1}^{\infty}\left(\frac{1}{\left[\ln\left(\frac{2(H+R)}{R}\right)\right]^i}\sum_{j=i}^{\infty}a_{ij}\left(\frac{c}{(H+R)}\right)^j\right)$. From Eqs. (5) and (6*a*), the total elastic energy of the nanotube can be obtained as

$$E_{\text{elast}} = \frac{512}{5}\frac{EI}{L^3}c^2. \tag{10}$$

From Eqs. (7*a*) and (7*b*), the equilibrium condition provides

$$V(c) = \frac{H+R}{L^2}\ln\left(\frac{2(H+R)}{R}\right)\sqrt{\frac{1024EI}{5\pi\varepsilon_0 S'(c)}\left(\frac{c}{H+R}\right)}. \tag{11}$$

The central displacement of the nanotube at pull-in c_{PI} can be obtained from

$$\frac{\mathrm{d}V(c)}{\mathrm{d}c} = 0, \tag{12}$$

which means the pull-in corresponds to a maximum in V. Hence, the pull-in voltage can be written as

$$V_{\text{PI}} = k\frac{H+R}{L^2}\ln\left(\frac{2(H+R)}{R}\right)\sqrt{\frac{EI}{\varepsilon_0}}, \tag{13}$$

where $k = \sqrt{\frac{1024}{5\pi S'(c_{\text{PI}})}\left(\frac{c_{\text{PI}}}{H+R}\right)}$.

4 Finite Kinematics

To take into account the nonlinear effect arising from finite kinematics, we have to evaluate the energy stored in the beam not only due to bending but also due to stretching. This represents the predominant effect of the finite kinematics for the doubly clamped nanotube.

The strain due to bending is

$$\varepsilon_b = -y\frac{\mathrm{d}^2w}{\mathrm{d}z^2}, \tag{14}$$

where y has the origin in the centroid of the cross section, and is parallel to the direction of the loads. In addition, the mean value of stretching due to the displacement w, noting that $\mathrm{d}s^2 = \mathrm{d}z^2 + \mathrm{d}w^2$, is [14]

$$\varepsilon_s = \frac{\mathrm{d}s - \mathrm{d}z}{\mathrm{d}s} \approx \frac{1}{2L}\int_0^L\left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)^2\mathrm{d}z. \tag{15}$$

As a consequence, the elastic energy stored in the nanotube, is

$$E_{\text{elast}} = \frac{E}{2}\int_A\int_0^L(\varepsilon_s + \varepsilon_b)^2\mathrm{d}A\mathrm{d}z, \tag{16}$$

where $A = \pi(R_{\text{ext}}^2 - R_{\text{int}}^2)$ is the cross-section area of the nanotube. Considering Eq. (5), the result is

$$E_{\text{elast}} = \frac{512}{5}\frac{EI}{L^3}c^2\left(1 + \frac{128}{3003}\frac{c^2}{\rho^2}\right), \tag{17}$$

where the radius of inertia ρ is defined as $I = A\rho^2$. The first term corresponds to the bending, whereas the second nonlinear term represents the elastic strain energy stored in the beam due to the stretching of the nanotube. Note that the force (derivative of the energy) due to bending is linear, while the one due to stretching is cubic.

Considering the energy as the fundamental quantity to derive a nonlinear correction for the stretching, we have to consider the increase in beam stiffness as

$$EI \rightarrow \left(1 + \frac{128}{3003}\frac{c^2}{\rho^2}\right)EI. \tag{18}$$

Therefore, the equilibrium condition gives

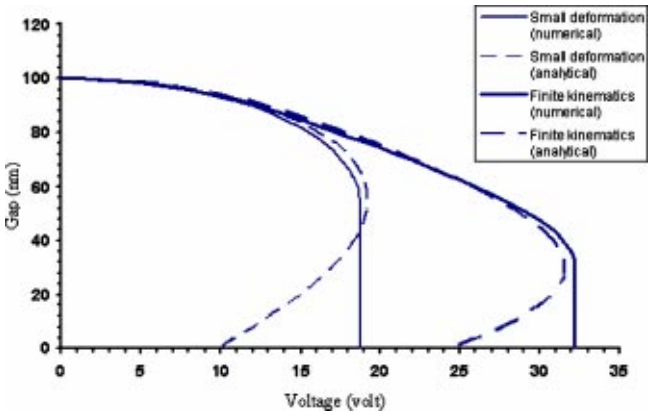


Fig. 2 Comparison between analytical predictions and numerical results. Plot of applied voltage versus gap for both small deformation and finite kinematics. The gap is measured between the axis of the nanotube and the electrode in the middle of the span.

Table 1 Comparison between pull-in voltages evaluated numerically and analytically by Eqs. (13) and (20) for doubly clamped nanotube devices, respectively. $E=1.0$ TPa, $R_{int}=0$. SD refers to small deformation, FK refers to finite kinematics.

Case	H[nm]	L[nm]	$R=R_{ext}$[nm]	V_{PI}[V] (theo-SD)	V_{PI}[V] (num-SD)	V_{PI}[V] (theo-FK)	V_{PI}[V] (num-FK)
1	100	4000	10	3.20	3.18	9.06	9.54
2	100	3000	10	5.69	5.66	16.14	16.95
3	100	2000	10	12.81	12.73	36.31	38.14
4	150	3000	10	9.45	9.43	38.93	40.92
5	200	3000	10	13.53	13.52	73.50	77.09
6	100	3000	20	19.21	18.74	31.57	32.16
7	100	3000	30	38.57	37.72	51.96	50.63

$$V^{FK}(c)=V(c)\sqrt{\left(1+\frac{128}{3003}\left(\frac{c}{\rho}\right)^2\right)}=\frac{H+R}{L^2}\ln\left(\frac{2(H+R)}{R}\right)\times\sqrt{\frac{1024EI\left[1+\frac{128}{3003}\left(\frac{c}{\rho}\right)^2\right]\left(\frac{c}{H+R}\right)}{5\pi\varepsilon_0 S'(c)}}. \quad (19)$$

Stationary condition, Eq. (12), applied to V^{FK} (FK refers to finite kinematics), provides the value of c at pull in. The pull-in voltage can then be expressed as

$$V_{PI}^{FK}=k^{FK}\frac{H+R}{L^2}\ln\left(\frac{2(H+R)}{R}\right)\sqrt{\frac{EI}{\varepsilon_0}}, \quad (20)$$

where $k^{FK}=\sqrt{\frac{1024}{5\pi S'(c_{PI})}\left(\frac{c_{PI}}{H+R}\right)\left[1+\frac{128}{3003}\left(\frac{c_{PI}}{\rho}\right)^2\right]}$.

5 Comparison Between Analytical Prediction and Numerical Simulations

An assessment of the derived analytical formulas is performed by comparing the results obtained solving numerically the corresponding elastic line equations, for both small deformation (only bending) and finite kinematics (bending+stretching). The nanotube properties and dimensions used here are Young's modulus $E=1.0$ TPa, $R_{ext}=20$ nm, $R_{int}=0$ nm, and $L=3000$ nm. The initial gap $H=100$ nm is also employed. Note that the theory does not involve a best fit parameter. The detailed comparison is reported in Fig. 2. In this figure, the applied voltage versus the gap, in the middle of the span between the nanotube and the electrode, is theoretically evaluated and plotted for both the small deformation model Eq. (11) and the finite kinematics model Eq. (19). These plots are compared in the same figure with the small deformation and finite kinematics numerical results obtained by solving Eq. (3) using a finite difference scheme. When evaluating the analytical solution, $i=1$ to 4 and $j=1$ to 10 are employed in the series of Eq. (8). The corresponding constants $\{a_{ij}\}$ in Eq. (9) are obtained using Mathematica®. From Fig. 2, it is clear that the finite kinematics effect is indistinguishable when the deformation is small and it gradually becomes significant with the increase in deformation. It is noted that the theoretical prediction curve can be divided into two parts with the separation point at $V=V_{PI}$ and $c=c_{PI}$. The part that corresponds to $c\leq c_{PI}$ follows the numerical results and can be experimentally implemented. The part that corresponds to $c>c_{PI}$ could be experimentally captured only by a displacement-control device. On the other hand, if the NEMS is voltage controlled, it will follow the unstable path (at $V=V_{PI}$) until reaching the contact. The difference between the two paths is related to the kinetic energy released by the structure after pull in when the device is actuated under voltage control. From Fig. 2, it can be concluded that the analytical results are in excellent agreement with the numerical results.

The effects of the geometry of the nanotube L and R and the step height H on the pull-in voltage of the NEMS device have also been examined both analyticall and numerically. The results are reported in Table 1. Columns five and six in Table 1 compare analytical and numerical pull-in voltage predictions under the assumption of small deformations. Columns seven and eight in Table 1 compare analytical and numerical pull-in voltage predictions under the assumption of finite kinematics. The agreement between the analytical predictions and numerical results is satisfactory (with a maximum discrepancy of 5%).

Note that an oversimplified model, e.g., assuming a capacitance of two parallel plates and a concentrated stiffness [10], can result in significant errors in the evaluation of the pull-in voltage. The importance of a more accurate model, that is the aim of this paper, has been recently emphasized in [4] where, by assuming a parallel plate capacitance, a pull-in voltage of 9.4 V was predicted in contrast to the experimental measurement of 8.5 V.

6 Closure

We have presented a theory to analyze nanotube structures, which is particularly suited to the design of NEMS and nanosensors. Comparison with numerical results shows good agreement. The formulas here reported could represent a considerable step forward in the understanding and development of nanosensors and NEMS. Note that the analysis is also applicable to microelectromechanical systems (MEMS).

With improvements in nanomanipulation and manufacturing of nanodevices we hope experimental measurements will become available, which will confirm or identify limitations of the theoretical predictions here reported.

Acknowledgments

The authors acknowledge the support from the FAA through Award No. DTFA03-01-C-00031 and the NSF through Award No. CMS-0120866. The authors would like to express our appreciation to Dr. J. Newcomb and Dr. J. Larsen-Base for supporting this work. Work was also supported in part by the Nanoscale Science and Engineering Initiative of the National Science Foundation under NSF Award No. EEC-0118025. The authors also express special thanks to Dr. N. Moldovan for helpful discussions.

References

[1] Iijima, S., 1991, "Helical Microtubules of Graphitic Carbon," Nature (London) **354**, pp. 56-58.
[2] Fennimore, A. M., Yuzvinsky, T. D., Han, W.-Q., Fuhrer, M. S., Cumings, J., and Zettl., A., 2003, "Rotational Actuators Based on Carbon Nanotubes," Nature (London), **424**, pp. 408-410.
[3] Rueckes, T., Kim, K., Joselevich, E., Tseng, G. Y., Cheung C.-L., and Lieber, C. M., 2000, "Carbon Nanotube-Based Nonvolatile Random Access Memory for Molecular Computing," Science, **289**, pp. 94-97.
[4] Kim, P., and Lieber, C. M., 1999, "Nanotube Nanotweezers," Science, **286**, pp. 2148-2150.
[5] Qian, D., Wagner, G. J., Liu, W. K., Yu, M.-F., and Ruoff R. S., 2002, "Mechanics of Carbon Nanotubes," Appl. Mech. Rev., **55**, pp. 495-532.
[6] Yu, M.-F., Lourie, O., Dyer, M. J., Moloni, K., Kelly, T. F., and Ruoff R. S., 2000, "Strength and Breaking Mechanics of Multiwalled Carbon Nanotubes Under Tensile Load," Science, **287**, pp. 637-640.
[7] Treacy, M. M., Ebbesen, T. W., and Gibson, J. M., 1996, "Exceptionally High Young's Modulus Observed for Individual Carbon Nanotubes," Nature (London), **381**, pp. 678-680.
[8] Chopra, N. G., and Zettl, A., 1998, "Measurement of the Elasticity of a Multiwall Boron Nitride Nanotube," Solid State Commun. **105**, pp. 297-300.
[9] Yakobson, B. I., Campbell, M. P., Brabec, C. J., and Bernholc, J., 1997, "High Strain Rate Fracture and C-chain Unraveling in Carbon Nanotubes," Comput. Mater. Sci., **8**, 1997, pp. 341-348.

[10] Desquenes, M., Rotkin S. V., and Alaru, N. R., 2002, "Calculation of Pull-in Voltages for Carbon-Nanotube-Based Nanoelectromechanical Switches," Nanotechnology **13**, pp. 120-131.
[11] Kinaret, J., Nord, T., and Viefers, S., 2003, "A Carbon-Nanotube-Based Nanorelay," Appl. Phys. Lett., **82**, pp. 1287-1289.
[12] Desquenes, M., Tang, Z., and Aluru, N. R., 2000, "Static and Dynamic Analysis of Carbon Nanotube-Based Switches," ASME J. Eng. Mater. Technol., **126**, pp.230-237.
[13] Ke, C.–H., Espinosa, H. D., and Pugno, N., 2004, "Numerical Analysis of Nanotube Based NEMS devices—Part II: Role of Finite Kinematics, Stretching, and Charge Concentrations," ASME J. Appl. Mech. (accepted for publication).
[14] Sathyamoorthy, M., 1998, *Nonlinear Analysis of Structures*, CRC Press, Boca Raton, FL.

Journal of Applied Mechanics

Brief Note

A Paradox in Sliding Contact Problems With Friction

G. G. Adams
Department of Mechanical Engineering,
Northeastern University,
Boston, MA 02115
Fellow ASME

J. R. Barber
Department of Mechanical Engineering,
University of Michigan,
Ann Arbor, MI 48109
Mem. ASME

M. Ciavarella
Senior Resarcher
CNR-ITC Str. Crocefisso 2/B,
70126 Bari, Italy

J. R. Rice
Division of Engineering and Applied Sciences,
Harvard University,
Cambridge, MA 02138
Fellow ASME

In problems involving the relative sliding to two bodies, the frictional force is taken to oppose the direction of the local relative slip velocity. For a rigid flat punch sliding over a half-plane at any speed, it is shown that the velocities of the half-plane particles near the edges of the punch seem to grow without limit in the same direction as the punch motion. Thus the local relative slip velocity changes sign. This phenomenon leads to a paradox in friction, in the sense that the assumed direction of sliding used for Coulomb friction is opposite that of the resulting slip velocity in the region sufficiently close to each of the edges of the punch. This paradox is not restricted to the case of a rigid punch, as it is due to the deformations in the half-plane over which the pressure is moving. It would therefore occur for any punch shape and elastic constants (including an elastic wedge) for which the applied pressure, moving along the free surface of the half-plane, is singular. The paradox is resolved by using a finite strain analysis of the kinematics for the rigid punch problem and it is expected that finite strain theory would resolve the paradox for a more general contact problem. [DOI: 10.1115/1.1867992]

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the ASME Applied Mechanics Division, September 9, 2002 final revision, October 3, 2003. Editor: R. M. McMeeking.

Formulation

Consider a rigid flat punch, indenting an elastic half-plane as shown in Fig. 1. From the classical equations of plane-strain elasticity (see, e.g., Barber [1], p. 154) the displacements $\bar{u},\bar{v}$ on the surface of the half-plane are given by

$$\frac{d\bar{u}}{d\bar{x}}(\bar{x},0) = -\frac{1-\nu}{\pi\mu}\int_S \frac{\bar{q}(\bar{\xi})d\bar{\xi}}{\bar{x}-\bar{\xi}} - \frac{1-2\nu}{2\mu}\bar{p}(\bar{x}), \tag{1}$$

$$\frac{d\bar{v}}{d\bar{x}}(\bar{x},0) = \frac{1-\nu}{\pi\mu}\int_S \frac{\bar{p}(\bar{\xi})d\bar{\xi}}{\bar{x}-\bar{\xi}} - \frac{1-2\nu}{2\mu}\bar{q}(\bar{x}), \tag{2}$$

where $\bar{p},\bar{q}$ are the pressure and shear, respectively, transmitted at the interface ($\bar{p}$ is positive in compression), ν is the Poisson's ratio, and μ is the shear modulus.

We initially consider a frictionless punch ($\bar{q}=0$) that is moving with a velocity U_0 which is much less than any of the wave speeds of the elastic body. Hence inertia effects can be neglected and the steady solution is given in terms of a moving coordinate system (x,y) where

$$x=\bar{x}-U_0t, \quad y=\bar{y}, \quad u(x,y)=\bar{u}(\bar{x},\bar{y},t), \quad v(x,y)=\bar{v}(\bar{x},\bar{y},t) \tag{3}$$

The velocity of particles in the elastic material is denoted by U, where

$$U \equiv \frac{d\bar{u}}{dt} = \frac{du}{dx}\cdot\frac{dx}{dt} \Rightarrow U = -U_0\frac{du}{dx}(x,0) \tag{4}$$

Under frictionless sliding conditions, Eqs. (1) and (4) yield

$$U = 0 \quad |x| > a,$$

$$= U_0\frac{1-2\nu}{2\mu}p(x) \rightarrow +\infty \quad |x| \rightarrow a^- \tag{5}$$

where a is the half-width of the indenter.

Thus the effect of elastic deformation is to produce a velocity under the punch which is a function of position x and is in average very small if the mean pressure is, as we generally expect, much less than the elastic modulus. However, as the pressure becomes singular at the edges, *the speed U near the corners becomes greater than U_0 and, therefore, the local relative slip velocity is in the opposite direction to that of the punch motion*. Hence material points on the half-plane which enter under the leading edge of the punch are forced to move forward of the punch. Similarly material points near the trailing edge have a velocity greater than that of the punch and so, it would seem, never leave from under the punch.

This behavior itself is paradoxical, but is of particular concern as it persists when sliding is accompanied by Coulomb friction.

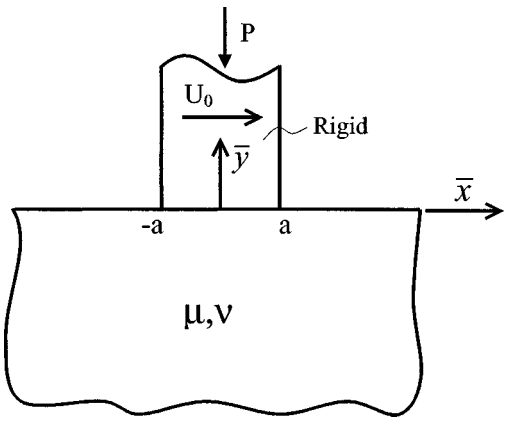


Fig. 1 A flat rigid punch indenting an elastic half-plane

One of the fundamental laws of friction is that the friction force opposes the relative slip velocity. Normally, this direction is inferred a priori, so that if Coulomb's friction law is assumed to hold, we anticipate

$$q(x) = fp(x), \tag{6}$$

where f is the coefficient of friction and Eqs. (1) and (2) reduce to

$$\frac{du}{dx}(x,0) = -\frac{(1-\nu)f}{\pi\mu}\int_{-a}^{a}\frac{p(\xi)d\xi}{x-\xi} - \frac{(1-2\nu)}{2\mu}p(x), \tag{7}$$

$$\frac{dv}{dx}(x,0) = \frac{(1-\nu)}{\pi\mu}\int_{-a}^{a}\frac{p(\xi)d\xi}{x-\xi} - \frac{(1-2\nu)}{2\mu}fp(x). \tag{8}$$

Since the punch is flat, we have

$$\frac{dv}{dx}(x,0) = 0, \quad x \in (-a,a) \tag{9}$$

and hence

$$\int_{-a}^{a}\frac{p(\xi)d\xi}{x-\xi} = \frac{(1-2\nu)\pi}{2(1-\nu)}fp(x), \quad x \in (-a,a). \tag{10}$$

Using this result to eliminate the integral in (7), we obtain

$$\frac{du}{dx}(x,0) = -\frac{(1-2\nu)(1+f^2)}{2\mu}p(x), \quad x \in (-a,a) \tag{11}$$

The contact pressure $p(x)$ must be positive and $\nu \leqslant 0.5$, so we conclude that du/dx is always negative under the punch (except for an incompressible material, in which case it vanishes), giving a positive value of the particle velocity U in Eq. (4) which *becomes unbounded in a region near the punch corners* (*thus* $U > U_0$) *and hence where the local relative sliding motion is opposed to that of the punch motion*.

The paradox would continue to occur in the case of two elastic materials, for those combinations of punch wedge angles and material constants for which the pressure induced is singular (Dundurs and Lee, [2]; Gdoutos and Theocaris, [3]). Note that the elastic displacement in an elastic punch is constant in the frame of reference moving with it, and accordingly would not affect the relative sliding velocity.

Finite Strain Kinematics Analysis

Consider now X,Y as material coordinates in the reference state, and x,y the spatial coordinates of the same points in the deformed state. A deformation state can be represented as $x=x(X,Y,t)$, whereas displacements can be written as $u=x-X$ and $v=y-Y$. Hence, the *stretch ratio* λ (ratio of length in the deformed state to the length in the reference state) of a material filament $(dX,0)$ can be given as

$$\lambda = \sqrt{\left(\frac{dx}{dX}\right)^2 + \left(\frac{dy}{dX}\right)^2} \tag{12}$$

We observe that $\lambda > 0$.

Consider the flat rigid indenter as maintained on the plane $y=y_0$ (<0) moving at speed U_0 in the $+X$ direction. The material to be indented and also any apparatus which supports it along some plane $Y=y_1$ $(>y_0)$ is to be considered translationally invariant relative to the X coordinate. Then, if a steady state solution exists to the problem, the displacement u must have the form

$$u = u(X - U_0 t, Y) \tag{13}$$

and hence the velocity in the $+X$ direction of material points on the surface of the body is

$$\frac{du}{dt} = -U_0\frac{du}{dX} = -U_0\left(\frac{dx}{dX} - 1\right) \tag{14}$$

When there is contact with the flat face of the indentor, $y=y_0=const$ and, therefore, $dy/dX=0$. Hence, the stretch ratio defined by (12) reduces to

$$\lambda = \frac{dx}{dX} \tag{15}$$

and thus the velocity of material points along the contact with the indentor is

$$\frac{du}{dt} = -U_0(\lambda - 1) \tag{16}$$

This in turn gives the slip velocity s as

$$s \equiv U_0 - \frac{du}{dt} = \lambda U_0 > 0 \tag{17}$$

and a negative value for s can never occur since the stretch ratio cannot be negative. The paradox happens in regions where the infinitesimal strain solution predicts compressive strains so large that λ is predicted (impossibly) to be negative. Had we evaluated it, we would have done so by using

$$\varepsilon_{xx} = \frac{du}{dX} = \lambda - 1 \tag{18}$$

along the contact zone under the indentor. That means that the strain ε_{xx} can be no more negative than -1 (since λ has to be positive), whereas we fail this test when we approach the singularity at the corner of the indenter.

Discussion

This paradox would occur also in elastodynamics. In fact below the Rayleigh wave speed (c_R) the solution to any contact problem is the same as the corresponding quasi-static problem with a reduced modulus (which goes to zero at c_R), [4]. The elastodynamic solution for a normal point force moving at constant speed over the surface of an half-plane becomes resonant at the Rayleigh wave speed (c_R) and above that speed a downward force produces an upward displacement.[1] Hence, for elastodynamic problems involving, for example, a rigid punch sliding over an elastodynamic half-plane, the solution behaves as the static indentation of an elastic half-plane of reduced modulus (which approaches zero at the Rayleigh wave speed). However, it is still true that a compressive strain of magnitude greater than unity is needed to produce the paradox (although the required normal pressure is reduced).

[1]This phenomenon leads to a paradoxical behavior of its own, i.e., the Craggs–Roberts paradox [5,6].

Therefore, the paradox considered in this paper is even more likely to occur, as strains can be arbitrarily large with any pressure, provided we are close enough to c_R.

The finite strain kinematics analysis shows that the paradox disappears when the correct kinematics is used. In this case it is a reasonable engineering solution to use the infinitesimal theory with the assumption that slip is always in the original direction of sliding, because the paradox occurs only in very small regions in which the infinitesimal theory is unrealistic.

There is, however, a class of problems where we see some doubt as to the proper formulation using infinitesimal theory. For the moving punch, we propose to use the direction of the punch motion to determine the relative slip velocity, i.e., we ignore the velocity reversal due to the singularity. Now suppose that the punch is stationary and is subjected to an incoming wave. The direction of the particle motion beneath the punch governs the slip direction and the singularity does not produce a slip reversal, due to (4), because $U_0=0$. However, the imposed motion itself may be sufficient to give slip reversal in some regions. Now consider the case in which the punch is given a small velocity. According to (4) any finite (no matter how small) velocity will produce slip, near the moving singularities, in the opposite direction as that due to the punch motion alone. In this case it is unclear as to whether to ignore the effect of the velocity reversal, as we did for the moving punch without the incoming wave, or to include its effect as we suggest when the punch is perfectly stationary.

Acknowledgments

M.C. is pleased to acknowledge the support from CNR-Consiglio Nazionale delle Ricerche (short term fellowship in July 2000), for his visit to Harvard University, permitting also the completion of the present work.

References

[1] Barber, J. R., 1992, *Elasticity*, Kluwer Academic Publishers, Boston.
[2] Dundurs, J., and Lee, M.-S., 1972, "Stress Concentrations at a Sharp Edge in Contact Problems," J. Elast., **2**, pp. 109–112.
[3] Gdoutos, E. E., and Theocaris, P. S., 1975, "Stress Concentrations at the Apex of a Plane Indenter Acting on an Elastic Half Plane," ASME J. Appl. Mech., **52**, pp. 688–692.
[4] Cole, J. D., and Huth, J. H., 1958, "Stresses Produced in a Half-Plane by Moving Loads," ASME J. Appl. Mech., **25**, pp. 433–436.
[5] Craggs, J. W., and Roberts, A. M., 1967, "On the Motion of a Heavy Cylinder Over the Surface of an Elastic Half-Space," ASME J. Appl. Mech., **34**, pp. 207–209.
[6] Georgiadis, H. G., and Barber, J. R., 1993, "On the Super-Rayleigh/Subseismic Elastodynamic Indentation Problem," J. Elast., **31**, pp. 141–161.

Journal of Applied Mechanics

Discussion

Discussion: "Axial Loading of Bonded Rubber Blocks" (Horton, J. M., Tupholme, G. E., and Gover, M. J. C., 2002, ASME J. Appl. Mech., 69, pp. 836–843)

T. J. Pence
Department of Mechanical Engineering, Michigan State University, East Lansing, MI, 48824-1226
e-mail: pence@egr.msu.edu

The paper by Horton, Tupholme, and Gover [1] analyzes the deformation behavior of rubber using the isotropic, infinitesimal strain theory of elasticty for an incompressible material. Accordingly, the Poisson's ratio $\nu=1/2$ and Young's modulus is three times the shear modulus ($E=3\mu$). In this setting, the stresses σ_{ij} determine the (infinitesimal) strains ϵ_{ij}, but the strains only determine the stresses up to a hydrostatic pressure. Hence, the strains determine the shear stresses and also determine the normal stress differences. The strains follow from the displacements u_i in the usual fashion $\epsilon_{ij}=(u_{i,j}+u_{j,i})/2$. In their analysis of a rectangular block with Cartesian coordinates (x,y,z) and associated displacement components (u,v,w) the authors in their analytic development arrive at a displacement field (Eqs. (14), (12), (24))

$$u=-x\frac{dw}{dz},$$

$$v=0,$$

$$w=\frac{3F}{4EA}\left\{z-\frac{2\sinh\frac{\alpha z}{2}\cosh\left[\frac{\alpha}{2}(h-z)\right]}{\alpha\cosh\frac{\alpha h}{2}}\right\},$$

where $\alpha=4\sqrt{3}/b$ and F, E, A, b, h are constants with units of force, stress, area, length, length, respectively. On the basis of the stress equation of equilibrium associated with the x direction, and in conjunction with the stress-strain-displacement relations reviewed above, they also obtain [Eq. (17)]

$$\sigma_{zz}=\frac{E}{3}\left[4\frac{dw}{dz}-\frac{1}{2}\left(\frac{b^2}{4}-x^2\right)\frac{d^3w}{dz^3}\right]-\frac{F}{A}.$$

These fields are central to the ensuing development, in particular to the calculation of percentage errors associated with previous treatments (e.g., Table 1).

It is indeed the case that the above fields follow from a development based upon $u=u(x,z)$, $v=0$, $w=w(z)$, and the selected equilibrium equation. The above fields now permit the determination of all the components of stress (see, e.g., (15)). On this basis one verifies that the stress equations of equilibrium are satisfied with respect to the x and y directions. However, with respect to the z direction, one obtains that

$$\frac{\partial\sigma_{xz}}{\partial x}+\frac{\partial\sigma_{yz}}{\partial y}+\frac{\partial\sigma_{zz}}{\partial z}=-\frac{3\sqrt{3}F}{Ab^3}(b^2-8x^2)\operatorname{sech}\left(\frac{2\sqrt{3}h}{b}\right)\sinh\left(\frac{2\sqrt{3}(h-2z)}{b}\right).$$

In view of the fact that the full set of equilibrium equations are not satisfied, there would seem to be grounds for concern regarding the validity of the improved expressions put forward in this paper. Similar concerns would naturally arise with respect to other treatments that use a similar methodology.

References

[1] Horton, J. M., Tupholme, G. E., and Gover, M. J. C., 2002, "Axial Loading of Bonded Rubber Blocks," ASME J. Appl. Mech., **69**, pp. 836–843.

Closure to "Discussion of 'Axial Loading of Bonded Rubber Blocks' " (2005, ASME J. Appl. Mech., 72, p. 453)

J. M. Horton
Reflecting Roadstuds Ltd., Boothtown, Halifax, HX3 6TR, UK

G. E. Tupholme
School of Computing and Mathematics, University of Bradford, Bradford, BD7 1DP, UK
e-mail: g.e.tupholme@bradford.ac.uk

Our immediate reaction to the comments of Pence was one of amazement that such a supposedly prestigious journal would consider it appropriate to devote valuable space to what, we feel, should be obvious to a reasonably well-informed reader who has delved into the related literature.

In his first paragraph, Pence appears to review a few random elements of the very well-known basics of classical elasticity theory (even to the extent of giving the absolutely fundamental relationships between the displacement and strain components). Several of these were, in fact, willingly deleted from our original manuscript, upon receiving the observation of a reviewer that, "It is not necessary to remind readers of *Journal of Applied Mechanics* … for a linear elastic model."

He then, seemingly unnecessarily, simply reproduces a few of our expressions, before making the observation that the equilibrium equation in the z direction is not satisfied. Actually this would clearly not be expected with the basic objectives and assumptions of the analyses presented. However, he naively appears to regard this as a hugely serious flaw that casts doubt on the worth of the realistic approximate expressions derived for the apparent Young's modulus, offering improvements on those previously available.

The approximations developed by Gent and Lindley [1] and Gent [2] have been widely quoted and used in the engineering industry for assessing axial stiffness. As we hoped was clearly explained in our introduction in Sec. 1, they were derived on the basis of two fundamental assumptions: (i) that planes initially normal to the direction of loading remain normal after loading, and (ii) that the deformed shapes of the free lateral surfaces are parabolic. Subsequently, as we pointed out in Secs. 1 and 5.2, the validity of the assumption (ii) has been questioned by several authors in interpreting their experimental results—with comments including "the assumption of a parabolic profile is erroneous" and that the next step "would be an improved method of estimating the 'bulgeability.' " It was our aim therefore to provide such estimates with this assumption of parabolic profiles removed, while maintaining the more reasonable first assumption.

However, it is a natural consequence of this simplifying assumption (i) that, for an incompressible material, it becomes unrealistic to satisfy exactly the equilibrium equation in the z direction (except on the central plane, $z=h/2$) on an infinitesimal volume. Further, this same assumption leads to the prediction of the existence of a shear stress on the unloaded boundary which, as observed in the paper, cannot actually physically exist.

If Pence were to refer to the already cited references of Gent and Lindley [1] and Gent [2], he would deduce that the theories therein lead to solutions for the stress components that also do not satisfy the equilibrium equation in the direction of loading, nor do the expressions derived in the later considerations of, for example, Constantinuo, Kartoum, and Kelly [3], Chalhoub and Kelly [4], and Tsai and Lee [5].

Additionally, it should perhaps be pointed out that all the above papers, and others, have assumed the rubber block to have a small thickness and have either predicted or assumed parabolic deformed profiles. In contrast, our analysis applies to a block of any thickness, and predicts that, especially for blocks of small shape factor, the profile is noticeably not parabolic. This is reassuringly in agreement with the experimental findings of Mott and Roland [6] and others.

In conclusion, we would suggest that, contrary to Pence's concerns in his final sentence, methodologies similar to that adopted in our paper have proved invaluable and extremely useful in related analyses. Particularly worthy of note are the expressions that Horton, Gover, and Tupholme [7,8] presented for the radial stiffness and tilting stiffness of a rubber bush mounting of finite length. Not only were there no useful estimates available previously, but moreover they yield numerical values that agree well with the available experimental data.

We are grateful to the Editor of the *Journal of Applied Mechanics* for giving us this opportunity to respond. We hope that our comments will enhance the appreciation of the potential importance and value of our results, for those readers who have not worked directly in this area of rubber technology and are therefore less familiar with the relevant literature.

References

[1] Gent, A. N., and Lindley, P. B., 1959, "The Compression of Bonded Rubber Blocks," Proc. Inst. Mech. Eng., **173**, pp. 111–122.
[2] Gent, A. N., 1994, "Compression of Rubber Blocks," Rubber Chem. Technol., **67**, pp. 549–558.
[3] Constantinuo, M. C., Kartoum, A., and Kelly, J. M., 1992, "Analysis of Compression of Hollow Elastomeric Bearing," Eng. Struct., **14**, pp. 103–111.
[4] Chalhoub, M. S., and Kelly, J. M., 1990, "Effect of Bulk Compressibility on the Stiffness of Cylindrical Base Isolation Bearings," Int. J. Solids Struct., **26**, pp. 743–760.
[5] Tsai, H-C., and Lee, C-C., 1998, "Compressive Stiffness of Elastic Layers Bonded between Rigid Plates," Int. J. Solids Struct., **35**, pp. 3053–3069.
[6] Mott, P. H., and Roland, C. M., 1995, "Uniaxial Deformation of Rubber Cylinders," Rubber Chem. Technol., **68**, pp. 739–745.
[7] Horton, J. M., Gover, M. J. C., and Tupholme, G. E., 2000, "Stiffness of Rubber Bush Mountings Subjected to Radial Loading," Rubber Chem. Technol., **73**, pp. 253–264.
[8] Horton, J. M., Gover, M. J. C., and Tupholme, G. E., 2000, "Stiffness of Rubber Bush Mountings Subjected to Tilting Deflection," Rubber Chem. Technol., **73**, pp. 619–633.

